

CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS
DEL INSTITUTO POLITÉCNICO NACIONAL

LABORATORIO NACIONAL DE GENÓMICA PARA LA
BIODIVERSIDAD

Unidad Irapuato

Búsqueda de desequilibrio mitonuclear en poblaciones humanas

Tesis que presenta

I.B.Q Eduardo Martín González Orozco

Para obtener el grado de

Maestro en Ciencias

En la especialidad de

Biología Integrativa

Director de Tesis:

Rafael Montiel Duarte

Asesores:

Andrés Moreno Estrada

Selene Lizbeth Fernández Valverde

El presente trabajo se realizó en el Laboratorio de Interacción Núcleo-Mitocondrial y Paleogenómica en la Unidad de Genómica Avanzada del CINVESTAV Unidad Irapuato. Bajo la dirección del Dr. Rafael Montiel y con la asesoría del Dr. Andrés Moreno Estrada y la Dra. Selene Lizbeth Fernández Valverde. Agradezco al Consejo Nacional de Ciencia y Tecnología por la beca por estudios de maestría N° 59070 otorgada durante el período comprendido de septiembre del 2015 a agosto del 2017

Fueron muchas las personas que contribuyeron con el siguiente trabajo. Primeramente, me gustaría agradecer al Dr. Rafael Montiel por brindarme la oportunidad de desarrollar mis ideas y este proyecto dentro de su laboratorio. Me gustaría agradecer enormemente a Miguel Vallebuena por apoyarme en todo momento, por el tiempo en la discusión de ideas y más importante por introducirme al hermoso mundo de la evolución.

Se agradece de igual manera a los auxiliares Christian Eduardo Martínez Guerrero e Hilda Eréndira Ramos Aboites por su apoyo técnico y discusión científica para la realización del presente trabajo

También quisiera agradecer a los miembros del laboratorio de Interacción núcleo mitocondria y paleogenómica, por la gran aportación de ideas y discusión de estas.

Finalmente, pero no menos importante, agradezco a mi familia y amigos que me ha apoyado en todos mis proyectos y que siempre me han mantenido motivado durante mis logros académicos

Tabla de Contenido

Resumen.....	12
Abstract.....	14
1 Introducción y antecedentes	16
1.1 Interacciones mitonucleares	16
1.1.1 Genoma nuclear humano.....	16
1.1.2 Genoma mitocondrial.....	17
1.1.3 Coevolución mitonuclear.....	18
1.1.3.1 Hipótesis de la “Reina Roja” y del “ADN egoísta”	19
1.1.3.2 Modelo de incompatibilidad híbrida de Dobzhansky-Muller	19
1.1.3.3 Inviabilidad híbrida en levadura	20
1.1.3.4 Inviabilidad híbrida en móridos.....	22
1.1.3.5 Inviabilidad híbrida en copépodos	23
1.1.4 Desequilibrio.....	25
1.1.4.1 Desequilibrio de ligado.....	25
1.1.5 Interacción proteína-proteína	28
1.1.6 Interacción ADN-proteína.....	29
1.1.7 Interacción ARN-proteína	31
1.1.8 Regulación mediada por ARN	31
1.1.9 interacciones ADN-ADN: Inserciones mitocondriales en núcleo	32
1.1.10 Incompatibilidad y desequilibrio mitonuclear en poblaciones humanas	33
2 Hipótesis.....	40
3 Objetivos	40
3.1 Objetivo general:	40
3.2 Objetivos específicos:.....	40
4 Materiales y métodos	41
4.1 Poblaciones e individuos	41
4.1.1 Filtrado de individuos	43
4.2 Filtrado de cromosomas sexuales.....	43
4.3 Filtrado de variantes y subconjunto de poblaciones	44
4.4 Estadístico <i>tau</i> de Goodman-Kruskal	44
4.4.1 Cálculo y prueba de la <i>tau</i> de Goodman-Kruskal.....	46
4.5 Análisis por parejas	46
4.6 Significancia estadística.....	47
4.6.1 Asociaciones perfectas	47
4.6.2 Asociaciones altas.....	50
4.7 Filtrado de resultados de caracterización	52

4.8 Análisis de enriquecimiento de términos GO.....	54
5 Resultados.....	55
5.1 Selección del método estadístico óptimo para detectar desequilibrio mitonuclear.....	55
5.2 Filtrado de la base de datos de 1000 genomas	57
5.2.1 Filtrado por pedigrí.....	57
5.2.2 Filtrado de los SNP	59
5.2.3 Detección de desequilibrio mitonuclear en dos poblaciones usando el análisis por pares de <i>tau</i> de Goodman-Kruskal (un análisis preliminar)	61
5.2.3.1 Resultado de los análisis por parejas de las poblaciones japonesa y china	61
5.2.3.1.1 Asociaciones perfectas.....	63
5.2.3.1.2 Correlaciones altas.....	66
5.3 Detectando desequilibrio mitonuclear en poblaciones individuales	68
5.3.1 Resultados de análisis de desequilibrio mitonuclear en poblaciones individuales	68
5.3.1.1 Asociaciones perfectas	69
5.3.1.2 Asociaciones altas	70
5.3.1.2.2 Resultados significativos y caracterización	75
5.3.2 Localización genómica del desequilibrio mitonuclear	75
5.3.3 Caracterización de regiones candidatas	79
5.4 Detectando desequilibrio mitonuclear en superpoblaciones	86
5.4.1 Resultados de análisis de desequilibrio mitonuclear en superpoblaciones	87
5.4.1.1 Resultados estadísticamente significativos	89
5.4.1.2 Resultados significativos y caracterización.....	93
5.4.2 Localización genómica de desequilibrio mitonuclear en superpoblaciones.....	94
5.4.3 Caracterización de regiones candidatas en el análisis de superpoblaciones.....	96
5.5 Resultados del análisis de enriquecimiento de términos GO en poblaciones y superpoblaciones.....	103
6 Discusión	105
6.1 Poder de resolución de los análisis en búsqueda de desequilibrio mitonuclear.....	105
6.2 Efectos aleatorios en las asociaciones entre SNP nucleares y mitocondriales	105
6.3 Alto número de SNP con desequilibrio mitonuclear en poblaciones africanas y surasiáticas	106
6.4 Regiones con firmas de desequilibrio mitonuclear	107
6.4.1 Regiones con firmas de desequilibrio mitonuclear en poblaciones individuales	107
6.4.2 Regiones con firmas de desequilibrio mitonuclear en superpoblaciones	109
6.5 Comparación de resultados de poblaciones y superpoblaciones.....	111
6.6 Contexto global poblacional de frecuencias alélicas en los SNP bajo desequilibrio mitonuclear	113
6.7 Inserciones mitocondriales en núcleo (NUMT)	121
7 Conclusiones	127

8 Perspectivas.....	128
Tablas suplementarias.....	129
Literatura citada.....	135
Anexos	143

Índice de Figuras

Figura 1. Cariotipo del genoma nuclear humano, representando los 22 pares de cromosomas autosómicos y los cromosomas sexuales X e Y (National Human Genome Research Institute).....	17
Figura 2. Genoma mitocondrial representando los genes que codifican las subunidades de los complejos OXPHOS, rARN y tARN (St. John, 2013).	18
Figura 3. Modelo de Dobzhansky-Muller en un contexto mitonuclear. Se observan mitocondrias en diferentes fondos genéticos nucleares donde la adecuación se ve afectada en la pareja de genomas nucleares y mitocondriales incompatibles (tomado de Sloan et al., 2015).....	20
Figura 4. Incompatibilidad híbrida entre especies diferentes de levadura. En donde se observa que la población roja inicial diverge en dos poblaciones, adquiere mutaciones mitocondriales y ocurre un proceso de coadaptación entre el genoma mitocondrial y nuclear en las dos poblaciones, dando como resultado una incompatibilidad híbrida entre genomas en los híbridos de ambas especies (tomado de Chou & Leu, 2010).	21
Figura 5. Producción de lactato en híbridos múridos y células sin mitocondria (rho) 0. Los defectos de OXPHOS se observaron en los híbridos Rattus (Rn) y Otomys (Oi) como un aumento de 4 veces y 10 veces, respectivamente, en comparación con el control M. m domesticus (Mm)	22
Figura 6. Actividades enzimáticas de OXPHOS en los híbridos. El híbrido de Rattus norvegicus (Rn) mostró deficiencia en los complejos I, III y defecto parcial en IV. El de Otomys irroratus (Oi) mostró un defecto en los complejos I y II (McKenzie et al., 2003)	23
Figura 7. Inviabilidad híbrida en copépodos. Se observa una disminución de la adecuación en híbridos F3 en los que está presente el material genético paterno no coadaptado. La adecuación se recupera en el retrocruzamiento materno en el que se recupera la mayoría del material genético nuclear coadaptado materno. El ADN nuclear se representa por cromosomas lineales, en tanto que el mtADN de forma circular. En rojo para una población y en blanco para otra. Los genomas nucleares y mitocondrial coadaptados se muestran del mismo color (Ellison & Burton, 2008b). ...	24
Figura 8 Complejos de OXPHOS. LA ATP sintasa está indicada como el complejo V. De color rojo las proteínas codificadas en la mitocondria, de azul las codificadas en el núcleo (Rand et al., 2004)...	29
Figura 9 . Región D-loop mitocondrial, factores de transcripción codificados en el núcleo se unen en esta región para regular la transcripción de genes mitocondriales (Genetical Society (Great Britain) & Scarpulla, 1987)	30
Figura 10. Curvas de crecimiento de híbridos con haplogrupos mitocondriales H y J tratados con radiación UV y controles sin tratamiento de radiación UV (Malik et al., 2014).....	34

Figura 11. Resumen esquematizado de los estudios de cíbridos con los haplogrupos H y J donde se describe de que manera afecta a cada haplogrupo la exposición a luz UV(tomado de Malik <i>et al.</i> , 2014).	35
Figura 12. Desequilibrio de ligado mitonuclear analizando la base de datos del HGDP. Cada punto representa la media de los estadísticos de LD utilizados por cromosoma: r^2 estadístico que representa a manera de correlación el desequilibrio de ligado entre los SNP nucleares y los haplogrupos mitocondriales y D' que es el estadístico de desequilibrio de ligado normalizado. (Sloan et al., 2015).	37
Figura 13. Filogenia de ADN nuclear.Los colores representan cada población. Los haplotipos mitocondriales se superponen en el árbol (Rishishwar & Jordan, 2017).....	38
Figura 14. Distribución de muestras de la base de datos de los 1000 genomas (Auton et al., 2015).	41
Figura 15. Esquema representado la comparación por parejas (líneas oblicuas de colores) entre los SNP (barras verticales de colores) nucleares (nDNA – verde) y mitocondriales (mtDNA – mitocondrial).....	46
Figura 16. Ejemplo de covariación perfecta ($\tau = 1$) entre genotipos de un SNP nuclear (Nuc) y de un SNP mitocondrial (Mit) en nueve individuos (Ind).	47
Figura 17. Ejemplo de asociación perfecta entre genotipos mitocondriales (Mit) y nucleares (Nuc) con más de un individuo (Ind) variando.....	49
Figura 18. Ejemplo de determinación de probabilidad de frecuencias genotípicas, para tres casos de ejemplo de variación observada en 3,200,000 SNP.	51
Figura 19 Diagrama de flujo de caracterización y filtrado de resultados de análisis de tau por parejas.....	52
Figura 20. Gráfica de cajas y bigotes representando la distribución de valores de tau mayores a 0.1.	62
Figura 21. Gráfica de barras en donde cada una de ellas indica el número de SNP nucleares asociados de manera perfecta a un SNP mitocondrial en población japonesa.	63
Figura 22. Gráfica de barras que indica el número de SNP mitocondriales con correlaciones perfectas con SNP nucleares en población japonesa, seccionados según el número de individuos que varían en la posición para ese SNP. Los porcentajes indican el porcentaje esperado sobre el total de valores de asociaciones perfectas dado efectos de azar.	64
Figura 23. Gráfica de puntos indicando el número de SNP nucleares con asociación perfecta a un SNP mitocondrial. JPT, población japonesa; CHB, población china.	65
Figura 24 . Gráfica de puntos indicando el número de SNP nucleares con correlación perfecta a un SNP mitocondrial (azul, eje x primario) y número de individuos variando (rojo, eje x secundario). JPT, población japonesa; CHB, población china.	65
Figura 25. Distribución de densidad acumulativa de valores de tau reales y simulados en la población japonesa. El valor máximo de tau obtenido de los datos simulados fue de 0.79	66
Figura 26. Distribución de densidad acumulativa de valores de tau reales y simulados en la población china. El valor máximo de tau obtenido de los datos simulados fue de 0.79	67
Figura 27. Gráficas de cajas y bigotes representando la distribución de valores de tau mayores a 0.1 en cada población. CEU (europeos), FIN (Finlandeses), JPT (Japoneses), CHB (Chinos), BEB (Hablantes de Bengali), PJL (Punjabi), MXL (Mexicanos), PUR (Puertorriqueños), YRI (Yorubas) y GWD (Gambianos)	69

Figura 28. Distribución de densidad cumulativa de datos reales y simulados de valores de tau en población europea (CEU).	71
Figura 29. Distribución de densidad cumulativa de datos reales y simulados de valores de tau en población yoruba (YRI).	72
Figura 30. Distribución de densidad cumulativa de datos reales y simulados de valores de tau en población mexicana (MXL).	73
Figura 31. Distribución de densidad cumulativa de datos reales y simulados de valores de tau en población punjabi (PJI).	74
Figura 32. Número de SNP nucleares con desequilibrio mitonuclear significativo, en regiones de 24 gigabases, del análisis con poblaciones individuales. Los cromosomas con mayores número de SNP en desequilibrio son el 1, 2, 12-15 y 22.	76
Figura 33. Número de SNP mitocondriales con desequilibrio mitonuclear significativo, en ventanas de 500 bases, del análisis con poblaciones individuales.	78
Figura 34. Gráfica de pastel representando porcentajes de los resultados de caracterización de los SNP nucleares únicos bajo desequilibrio (No. de SNPs, %).	80
Figura 35. Gráfica de pastel representando porcentajes de los resultados de caracterización de los SNP mitocondriales únicos bajo desequilibrio.	80
Figura 36. Gráfica de cajas y bigotes de distribución de valores tau mayores a 0.1 en superpoblaciones. AFR (africanos), AMR (americanos mestizos), EAS (asiáticos del este), EUR (europeos), SAS (asiáticos del sur).	88
Figura 37. Distribuciones de función cumulativa de datos reales vs. simulados de tau en la superpoblación europea.	90
Figura 38. Distribuciones de función cumulativa de datos reales vs. simulados de tau en la superpoblación de Asia del este.	91
Figura 39. Distribuciones de función cumulativa de datos reales vs. simulados de tau en la superpoblación africana.	91
Figura 40. Distribuciones de función cumulativa de datos reales vs. simulados de tau en la superpoblación americana mestiza.	92
Figura 41. Distribuciones de función cumulativa de datos reales vs. simulados de tau en la superpoblación surasiática.	92
Figura 42. Densidad de los SNP nucleares con desequilibrio mitonuclear en regiones de 24 gigabases en el análisis con superpoblaciones. En este caso, los cromosomas con mayor densidad de SNP bajo desequilibrio son el 1, 4 y 10, seguidos del 2 y 6.	94
Figura 43. Densidad de los SNP mitocondriales con desequilibrio mitonuclear en ventanas de 500 bases en el análisis con superpoblaciones.	95
Figura 44. Gráfica de pastel representando porcentajes de los resultados de caracterización de SNP nucleares únicos bajo desequilibrio en las superpoblaciones analizadas.	97
Figura 45. Gráfica de pastel representando porcentajes de los resultados de caracterización de SNP mitocondriales únicos bajo desequilibrio en superpoblaciones.	98
Figura 46. Heatmap de las frecuencias alélicas mínimas en distintas superpoblaciones para los SNP que presentaron desequilibrio mitonuclear en la superpoblación americana mestiza.	114
Figura 47. Heatmap de valores de tau de los SNP con desequilibrio mitonuclear en la superpoblación americana mestiza.	115

Figura 48. Heatmap de las frecuencias alélicas mínimas en distintas superpoblaciones para los SNP que presentaron desequilibrio mitonuclear en la superpoblación africana.....	115
Figura 49. Heatmap de valores de tau de los SNP con desequilibrio mitonuclear en la superpoblación africana.....	116
Figura 50. Heatmap de las frecuencias alélicas mínimas en distintas superpoblaciones para los SNP que presentaron desequilibrio mitonuclear en la superpoblación de Asia del este.....	117
Figura 51. Heatmap de valores de tau de los SNP con desequilibrio mitonuclear en la superpoblación asiática del sur.....	117
Figura 52. Heatmap de las frecuencias alélicas mínimas en distintas superpoblaciones para los SNP que presentaron desequilibrio mitonuclear en la superpoblación europea.	118
Figura 53. Heatmap de valores de tau de los SNP con desequilibrio mitonuclear en la superpoblación europea.	118
Figura 54. Heatmap de las frecuencias alélicas mínimas en distintas superpoblaciones para los SNP que presentaron desequilibrio mitonuclear en la superpoblación surasiática.....	119
Figura 55. Heatmap de valores de tau de los SNP con desequilibrio mitonuclear en la superpoblación surasiática.	119
Figura 56. Alineamiento del pseudogen MTATP6P1 y el gen MT-ATP6. Las letras A, B y C indican los SNP con desequilibrio mitonuclear encontrados.....	121
Figura 57. Esquema representando la presencia de dos bases en las posiciones del pseudogen MTATP6P1 y el gen MT-ATP6 con desequilibrio mitonuclear identificado en la superpoblación africana.	122
Figura 58. Profundidad de alineamiento en la región de 266,266 pares de bases del cromosoma 1 correspondiente a la base de datos de 1000 genomas.....	123
Figura 59. Profundidades de los NUMT por cromosoma en el genoma humano.....	124
Figura 60. Profundidades locales a lo largo del NUMT HSA_NumtS_001_b1.....	125
Figura 61. Visualización de la región NUMT HSA_NumtS_001_b1 y los resultados de PhyloP en donde las barras azules representan una alta conservación.	126

Índice de tablas

<i>Tabla 1. Posibles genotipos entre el genoma nuclear y los citotipos mitocondriales y de cloroplasto.</i>	27
<i>Tabla 2. Número total de individuos del panel de los 1000 genomas por población.</i>	42
<i>Tabla 3. Tabla de contingencia entre genotipos nucleares (X) y mitocondriales (Y).</i>	55
<i>Tabla 4. Tabla de contingencia de genotipos nucleares y mitocondriales con asociación perfecta (arriba). Resultados de estadísticos clásicos de LD y tau (abajo).</i>	55
<i>Tabla 5. Tabla de contingencia de genotipos nucleares y mitocondriales con asociación alta (arriba). Resultados de estadísticos clásicos de LD y tau (abajo).</i>	56
<i>Tabla 6. Tabla de contingencia de genotipos nucleares y mitocondriales con correlación media (arriba). Resultados de estadísticos clásicos de LD y tau (abajo).</i>	56
<i>Tabla 7. Número de individuos filtrados no relacionados en poblaciones de la base de datos 1000 genomas.</i>	57
<i>Tabla 8. Número total de SNP seleccionados y número total de individuos seleccionados por superpoblación.</i>	59
<i>Tabla 9. Número de SNP seleccionados y número total de individuos seleccionados por superpoblación.</i>	60
<i>Tabla 10. Número total de parejas de SNP nucleares y mitocondriales analizadas en las poblaciones japonesa y china.</i>	61
<i>Tabla 11. Número total y porcentaje de resultados con tau mayor a 0.1</i>	61
<i>Tabla 12. Número de resultados obtenidos por cada rango de valores de tau.</i>	62
<i>Tabla 13. Resultado de la prueba de K-S y máximo valor de tau simulado en población japonesa y china.</i>	67
<i>Tabla 14. Valores de tau significativos en las poblaciones japonesa y china.</i>	68
<i>Tabla 15. Cromosomas con mayor número de SNP con desequilibrio mitonuclear.</i>	77
<i>Tabla 16. Porcentaje de regiones con DMT caracterizadas.</i>	79
<i>Tabla 17. Principales genes nucleares con el mayor número de SNP únicos con desequilibrio mitonuclear en más de dos poblaciones. Pseudogenes con desequilibrio mitonuclear marcados con doble asterisco.</i>	84
<i>Tabla 18. Principales genes mitocondriales con el mayor número de SNP únicos con desequilibrio mitonuclear.</i>	85
<i>Tabla 19. Número total de combinaciones analizadas por superpoblación.</i>	87
<i>Tabla 20. Total de valores de tau mayores a 0.1 por superpoblación.</i>	87
<i>Tabla 21. Número de valores de tau por rango en cada superpoblación.</i>	89
<i>Tabla 22. Resultados de la prueba K-S contra la distribución simulada, p valor y máximo valor de tau obtenido en las simulaciones.</i>	90
<i>Tabla 23. SNP con desequilibrio mitonuclear estadísticamente significativo en distintas superpoblaciones.</i>	93
<i>Tabla 24. Número y porcentajes de regiones caracterizadas totales en el análisis de superpoblaciones.</i>	96
<i>Tabla 25. Número de SNP nucleares únicos por categoría caracterizados en las superpoblaciones.</i>	97

<i>Tabla 26. Resultados de caracterización de regiones mitocondriales y nucleares en las superpoblaciones analizadas.</i>	<i>98</i>
<i>Tabla 27. Principales genes nucleares con mayor número de SNP únicos con desequilibrio mitonuclear en más de dos superpoblaciones. Pseudogenes correspondientes a NUMT marcados con doble asterisco.....</i>	<i>99</i>
<i>Tabla 28. Genes nucleares con mayor número de SNP únicos con desequilibrio mitonuclear en cada superpoblación. Pseudogenes con desequilibrio mitonuclear marcados con asterisco.</i>	<i>100</i>
<i>Tabla 29. Principales genes mitocondriales con el mayor número de SNP únicos con desequilibrio mitonuclear.....</i>	<i>102</i>

Resumen

El origen de las células eucariontes basado en la teoría endosimbionte, implicó una relación simbiótica entre una α -proteobacteria y un precursor de célula eucarionte, interacción la cual dio lugar a la conformación de células complejas con dos genomas: nuclear y mitocondrial

La correcta interacción de la mitocondria con el núcleo es esencial para la viabilidad de la célula, y estas interacciones se han estudiado en diferentes niveles: proteína-proteína, cuya importancia radica en que sólo uno por ciento de las proteínas dentro de la matriz mitocondrial son codificadas en su ADN, por lo que una interacción correcta entre proteínas nucleares y mitocondriales es necesaria para llevar a cabo procesos biológicos como lo es la producción de energía y esto se puede observar en la conformación de los complejos de la fosforilación oxidativa. Interacción ADN-proteína, la cual se lleva a cabo en la regulación de expresión de genes mitocondriales, al interactuar factores de transcripción con el ADN mitocondrial, en rutas de biogénesis y homeostasis mitocondrial. Interacción ARN-proteína, que se lleva a cabo principalmente en la conformación del mitoribosoma, el cual se compone de ARN ribosomal codificado en la mitocondria y los componentes proteicos codificados en el núcleo, asimismo se ha asociado la regulación indirecta y directa de funciones mitocondriales por parte de micro ARN y ARN largo no codificante. Interacciones ADN-ADN, que se lleva a cabo mediante inserciones de fragmentos de ADN mitocondrial en el genoma nuclear, dando lugar a los NUMT por sus siglas en inglés.

Dadas las interacciones descritas, debe haber una coadaptación entre alelos mitocondriales y nucleares para lograr una alta aptitud. Estas coadaptaciones se pueden observar mediante el desequilibrio mitonuclear (asociación no aleatoria entre alelos mitocondriales y nucleares). El desequilibrio mitonuclear y la coadaptación entre el genoma nuclear y mitocondrial se han caracterizado en varias células de mamíferos, utilizando híbridos (líneas celulares con mtADN exógeno) en diferentes especies de muridos. Asimismo, los estudios en poblaciones de *Tigriopus californicus* mostraron un efecto de incompatibilidad mitonuclear reflejado en la adecuación de los híbridos, conocida como incompatibilidad híbrida.

En los estudios de poblaciones humanas, la búsqueda de este desequilibrio se ha realizado utilizando haplotipos mitocondriales y polimorfismos de un solo nucleótido (SNP). En un primer estudio se tomaron quinientos SNP de la base de datos "*Human Genome Diversity Project*", reportando valores de correlaciones bajos. Otro más correlacionó las distancias genéticas nucleares con haplogrupos mitocondriales concluyendo que en poblaciones con distancia genética nuclear cercana se pueden encontrar haplogrupos mitocondriales de una ubicación geográfica diferente sin una aparente consecuencia en la salud de los individuos del panel analizado.

En este trabajo se detectó el desequilibrio mitonuclear, realizando un análisis de alta resolución en pares de SNP, nucleares y mitocondriales, empleando el coeficiente de correlación *tau* de Goodman-Kruskal, el cual nos indica si existe asociación entre variables nominales. Este análisis se realizó en 10 poblaciones (chinos, japoneses, americanos con ancestría europea, finlandeses, yorubas, gambianos, mexicanos, puertorriqueños, punjabis y hablantes de bengalí) y en 5 superpoblaciones (americanos mestizos, europeos, africanos, asiáticos del este y asiáticos del sur) pertenecientes al panel de 1000 genomas.

Se lograron detectar 1,219 parejas de SNP en poblaciones y alrededor de 1,186 parejas de SNP en superpoblaciones con desequilibrio mitonuclear, los cuales se localizaron en regiones correspondientes a genes nucleares del metabolismo de ácidos grasos, de la apoptosis y de la homeostasis celular; genes con interacción indirecta con la mitocondria. En cuanto al genoma mitocondrial, los SNP se localizaron en genes codificantes para unidades proteicas pertenecientes a los complejos de la fosforilación oxidativa

Finalmente, el presente trabajo muestra evidencias de la existencia de desequilibrio mitonuclear, el cual se hipotetiza es resultado de coevolución entre el genoma nuclear y el mitocondrial, en poblaciones humanas; dicho desequilibrio se presentó de manera sutil y por ende este solo pudo ser detectado mediante un análisis de alta resolución, analizando toda la variación disponible en el panel de 1000 genomas. Asimismo el desequilibrio fue detectado en genes con evidencias de interacción mitocondrial indirecta, por lo que este trabajo brinda candidatos para un trabajo posterior funcional de incompatibilidad mitonuclear.

Abstract

The origin of eukaryotic cells based on the endosymbiont theory, involved a symbiotic relationship between an α -proteobacteria and a precursor of eukaryotic cell, interaction that led to the conformation of complex cells with two genomes: nuclear and mitochondrial

The correct interaction of the mitochondria with the nucleus is essential for the viability of the cell, and these interactions have been studied at different levels: protein-protein, whose importance lies in the fact that only one percent of the proteins within the mitochondrial matrix are encoded in their DNA, so a correct interaction between nuclear and mitochondrial proteins is necessary to carry out biological processes such as energy production. This can be observed in the conformation of oxidative phosphorylation complexes. DNA-protein interaction, which is carried out in the regulation of mitochondrial gene expression, by the interaction of transcription factors with mitochondrial DNA in the pathways of mitogenesis and mitochondrial homeostasis. RNA-protein interaction, which is mainly carried out in the conformation of the mitoribosome, which is composed of mitochondrial ribosomal ARN and protein components encoded in the nucleus, as well as indirect and direct regulation of mitochondrial functions by part of micro RNA and long non-coding RNA. DNA-DNA interactions, which is carried out by insertions of mitochondrial DNA fragments in the nuclear genome.

Given the interactions described, there must be a coadaptation between mitochondrial and nuclear alleles to achieve high fitness. These coadaptations can be observed through the mitonuclear disequilibria (non-random association between mitochondrial and nuclear alleles). Mitonuclear disequilibria and their relationship with coadaptation between both genomes have been characterized in several mammalian cells, using cybrids (cell lines with exogenous mtDNA) in different murine species. Likewise, the studies in populations of *Tigriopus californicus* showed an effect of mitonuclear incompatibility reflected in the adaptation of the hybrids, known as hybrid incompatibility.

In studies of human populations, the search for these disequilibria has been made using mitochondrial haplotypes and single nucleotide polymorphisms (SNP). In a first study, five hundred SNPs were taken from the "Human Genome Diversity Project" database, reporting low correlation values. Another work correlated the nuclear genetic distances with mitochondrial haplogroups, concluding that in populations with close nuclear genetic distance mitochondrial haplogroups can be found from a different geographic location without an apparent consequence in the health of the individuals.

In this work, the mitonuclear disequilibria was detected, performing a high-resolution analysis in nuclear and mitochondrial SNP pairs using the Goodman-Kruskal tau

correlation coefficient, which indicates whether there is an association between nominal variables. This analysis was carried out in 10 populations (Chinese, Japanese, Americans with European ancestry, Finns, Yorubas, Gambians, Mexicans, Puerto Ricans, Punjabis and Bengali speakers) and 5 super populations (Ad mixed Americans, European, African, East Asian and South Asian). belonging to the 1000 genomes panel.

We were able to detect 1,219 pairs of SNPs in populations and about 1,186 pairs of SNPs in superpopulations with mitonuclear disequilibria, which were located in regions corresponding to nuclear genes of fatty acid metabolism, apoptosis and cellular homeostasis; genes with indirect interaction with the mitochondria. Regarding the mitochondrial genome, the SNPs were located in genes coding for protein units belonging to the oxidative phosphorylation complexes

Finally, the present work shows evidence of the existence of mitonuclear disequilibria, which is hypothesized is the result of coevolution between the nuclear and mitochondrial genomes, in human populations; This disequilibria was presented in a subtle way and therefore this could only be detected by a high resolution analysis, that involves all the variation available in the panel of 1000 genomes. Likewise, the disequilibria was detected in genes with evidence of indirect mitochondrial interaction, providing candidates for a later functional work of mitonuclear incompatibility.

1 Introducción y antecedentes

1.1 Interacciones mitonucleares

Las interacciones entre el genoma nuclear y el mitocondrial han sido factor clave para la viabilidad de las células eucariotas complejas; interacciones que han sido moldeadas por procesos coevolutivos, desde la aparición de las primeras células eucariotas, según la teoría endosimbionte.

La sinergia lograda entre las interacciones de ambos genomas se puede entender como el resultado de un proceso coadaptativo y puede ser observada a diferentes niveles: ADN-ADN, proteína-ADN, proteína-ARN y proteína-proteína.

Previo a un planteamiento hipotético sobre la detección de firmas de interacción mitonuclear en poblaciones humanas, es imperativo hacer mención de algunas características de ambos genomas.

1.1.1 Genoma nuclear humano

El genoma nuclear humano está conformado por 3.2×10^8 nucleótidos, divididos en 23 pares de cromosomas, de los cuales 22 corresponden a autosómicos y el restante a cromosomas sexuales (XX en mujeres y XY en hombres). Aproximadamente, han sido identificados poco más de 40 000 genes, de los cuales se estima que 2500 son genes específicos de ARN no codificante (Kent *et al.*, 2002; Zerbino *et al.*, 2018)

Se ha identificado que un 62% del genoma humano comprende regiones intergénicas (Brown, 2002). Dentro del genoma nuclear existen dos categorías de secuencias repetidas: repetidos intercalados (los cuales se distribuyen de manera aleatoria en el genoma) y en *tándem* (los cuales se distribuyen uno después de otro), siendo los más abundantes la categoría de intercalados, y específicamente los *Short Interspersed Nuclear Elements (SINEs)*, de los cuales se han identificado alrededor de 1,558,000 elementos (Brown, 2002).

Así mismo, la regulación de genes a nivel transcripcional es llevada a cabo por diferentes elementos: promotores, aisladores, silenciadores y potenciadores. Dichos factores de transcripción interactúan de manera directa con el ADN para llevar a cabo la regulación; se han identificado apropiadamente mas de 1,600 proteínas factores de transcripción (Lambert *et al.*, 2018)

La información genética nuclear se hereda de forma biparental, con recombinación de material genético entre parentales (Karaman, 2003). La tasa de mutación estimada en el genoma nuclear humano es de 1.1×10^{-9} por sitio por generación (Roach *et al.*, 2010).

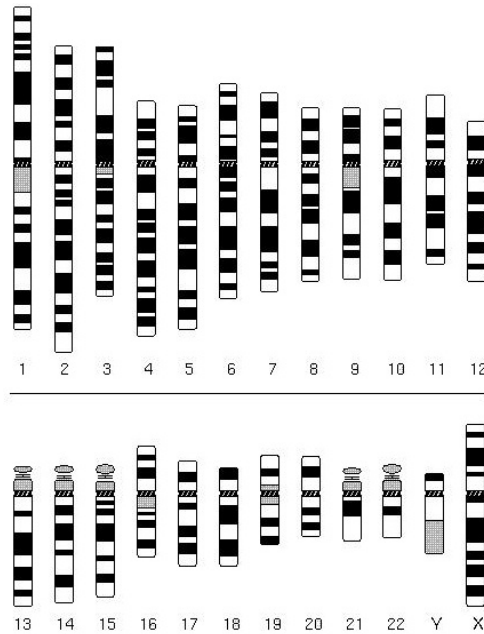


Figura 1. Cariotipo del genoma nuclear humano, representando los 22 pares de cromosomas autosómicos y los cromosomas sexuales X e Y (National Human Genome Research Institute)

1.1.2 Genoma mitocondrial

Un genoma adicional que se localiza dentro de células eucariotas, es el mitocondrial, que en humanos posee las siguientes características: Es circular y de doble cadena, comprende alrededor de 16,500 pares de bases (pb) y 37 genes, de los cuales 13 son genes que codifican proteínas de los complejos de la fosforilación oxidativa (OXPHOS, por sus siglas en inglés), 22 codifican ARN de transferencia (tARN) y 2 ARN ribosomal (rARN). No obstante, cabe hacer notar que el 99% de las proteínas de la matriz mitocondrial están codificadas en el genoma nuclear, las cuales son importadas (Meisinger, Sickmann, & Pfanner, 2008). En base a la composición de G + T, las cadenas de ADN mitocondrial (mtADN) pueden clasificarse como: "pesada" (H) y "ligera" (L). La mayoría de la información mitocondrial se encuentra en la cadena pesada. La cadena ligera codifica solo ocho tARN y un único polipéptido. El D-loop es el sitio de control para la expresión del mtADN, presenta una estructura de tres cadenas y es sitio de unión a factores de transcripción (los cuales están codificados en el núcleo) (Taanman, 1999).

La tasa de mutación estimada más recientemente para el genoma mitocondrial humano es de 0.017×10^{-6} sustituciones por sitio por año, excluyendo a la región control, la cual tiene una tasa mutacional más alta correspondiente a $0.075 - 0.16 \times 10^{-6}$ sustituciones por sitio por año (Pakendorf & Stoneking, 2005).

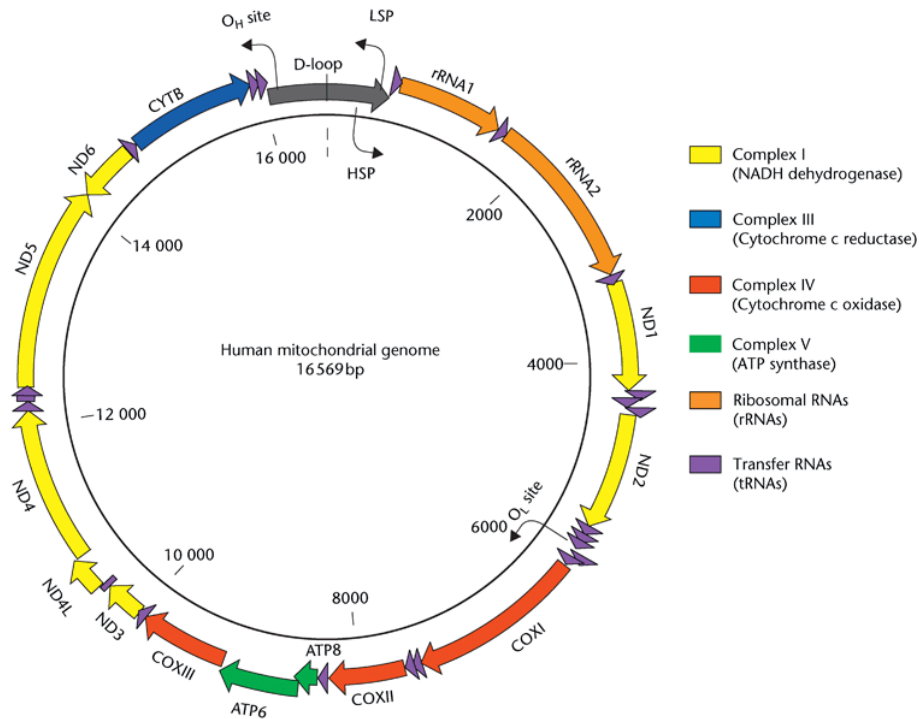


Figura 2. Genoma mitocondrial representando los genes que codifican las subunidades de los complejos OXPHOS, rARN y tARN (St. John, 2013).

1.1.3 Coevolución mitonuclear

El origen del genoma mitocondrial, de acuerdo con la teoría endosimbiótica, ocurrió hace aproximadamente dos mil millones de años, cuando una alfa-proteobacteria fue atrapada por un precursor de células eucarióticas. A medida que el precursor evolucionó y adquirió nuevas funciones, la mayoría del genoma de la alfa-proteobacteria se transfirió al genoma nuclear del precursor, y el genoma restante de la alfa-proteobacteria evolucionó en la mitocondria. La interacción entre ambos genomas dio como resultado una coevolución que contribuyó a que la ruta de la fosforilación oxidativa se llevara a cabo permitiendo la respiración aeróbica en células eucariotas (Friedman & Nunnari, 2014).

Como se mencionó con anterioridad, desde el surgimiento de las células eucariotas, el genoma mitocondrial y nuclear han coexistido e interactuado. Estas interacciones, como se describieron en las secciones anteriores, pueden darse en varios niveles, y tener como resultado con el paso de generaciones una coevolución, manteniendo la coadaptación de ambos genomas.

1.1.3.1 Hipótesis de la “Reina Roja” y del “ADN egoísta”

A través de la necesidad de coadaptación entre los genomas nuclear y mitocondrial, ambos genomas se mantendrían coevolucionando de manera constante, siendo un proceso dinámico que se puede describir mediante la hipótesis de la Reina Roja.

La hipótesis de la Reina Roja ha sido planteada en modelos de huésped-patógeno en los que un elemento evoluciona constantemente para infectar la mayor cantidad de genotipos posibles (en el caso de los patógenos) y otro para resistir la infección (en el caso del huésped), lo que lleva a una “carrera armamentista” entre ambos organismos (Clay & Kover, 1996). Si extrapolamos esta idea al genoma mitocondrial y nuclear, cada genoma se está adaptando de manera continua a los cambios presentados por el otro genoma, dando como resultado un proceso coadaptativo con interacciones eficientes.

Otro fenómeno presente en la mitocondria, es la existencia de ADN mitocondrial “egoísta”, entendiendo estos elementos genéticos "egoístas" como ADN con ventaja en su heredabilidad y sin ventaja de adecuación para el organismo (Phillips *et al.*, 2015). En el caso de las mitocondrias la herencia asexual aumenta las posibilidades de heredar mutaciones deletéreas (Chou & Leu, 2015).

Se ha descrito la existencia ADN mitocondrial egoísta en *Caenorhabditis briggsae* (Clark *et al.*, 2012). Este fenómeno propicia la herencia de alelos mitocondriales que podrían causar incompatibilidades intergenómicas mitonucleares. Esta incompatibilidad se ha descrito en plantas, en las que un alelo materno relacionado con elementos quiméricos en la mitocondria causa infertilidad y, en algunos casos, la fertilidad es recuperada por acción de genes nucleares (Case & Willis, 2008; Chou & Leu, 2015; Hanson & Bentolila, 2004).

1.1.3.2 Modelo de incompatibilidad híbrida de Dobzhansky-Muller

El modelo de Bateson-Dobzhansky-Muller postula la inviabilidad de los híbridos como consecuencia de la incompatibilidad de alelos parentales en la progenie. En el contexto de interacciones mitonucleares este modelo toma forma como la incompatibilidad de alelos nucleares y mitocondriales, afectando directamente a la adecuación del organismo (Orr, 1996).

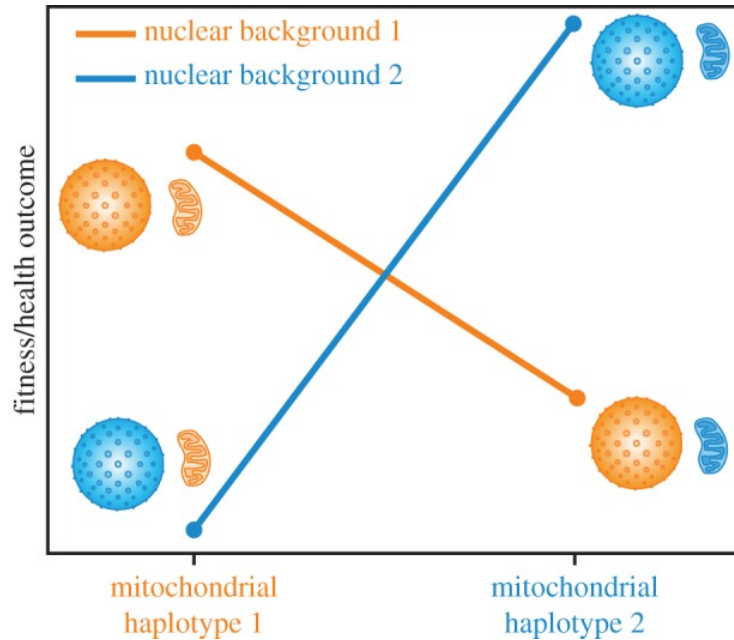


Figura 3. Modelo de Dobzhansky-Muller en un contexto mitonuclear. Se observan mitocondrias en diferentes fondos genéticos nucleares donde la adecuación se ve afectada en la pareja de genomas nucleares y mitocondriales incompatibles (tomado de Sloan et al., 2015).

Un escenario en el que este fenómeno puede ocurrir es el de una población ancestral, de la cual derivan dos poblaciones que posteriormente presentan mutaciones independientes en alguno de sus dos genomas, las cuales se fijan y se coadaptan al otro genoma. Cuando ambas poblaciones con sus genomas coadaptados generan un híbrido, la progenie se ve afectada por la incompatibilidad entre genomas (ver ejemplo de incompatibilidad híbrida en levaduras, Figura 7).

1.1.3.3 Inviabilidad híbrida en levadura

La inviabilidad del híbrido bajo un contexto de interacción mitonuclear se ha observado en levaduras. Los experimentos realizados, en los que se observó este fenómeno, consistieron en construir líneas celulares de *S. cerevisiae* y *S. paradoxus* con ADN mitocondrial exógeno, conocidos como cíbridos. En este estudio, las mitocondrias se intercambiaron entre ambas especies de levadura. Se descubrieron incompatibilidades entre el gen *sc-MRS1* en *S. cerevisiae* y su ortólogo *sp-MRS1* en *S. paradoxus*. En el que *sc-MRS1-t* fue incapaz de empalmar un intrón en *COX-*

1, gen codificante de la subunidad I de citocromo oxidasa, que es esencial para la fosforilación oxidativa en las mitocondrias, lo que provoca una deficiencia en esta vía, consecuencia de la incompatibilidad entre *S. cerevisiae* MRS -1 y *S. paradoxus* COX-1 (Chou & Leu, 2010).

Otro caso de estudio de inviabilidad híbrida mitonuclear en levadura, se realizó con híbridos de *S. cerevisiae* y *S. bayanus*, en los que el gen AEP2 (factor de transcripción) de *S. bayanus* no interactuó con la región 5'-UTR del gen ONL11, parte del complejo ATP sintasa, de *S. cerevisiae* (Chou & Leu, 2010).

En ambos experimentos se utilizaron especies diferentes de levadura, en la que se intercambiaron sus genomas mitocondriales, observando una incompatibilidad en los híbridos. Si se considera esta incompatibilidad híbrida como resultado de una coevolución en poblaciones separadas entre los genomas mitocondriales y nucleares, podríamos considerar el modelo de Dobzhansky-Muller como una hipótesis fuerte en el caso de levaduras (Figura 8).

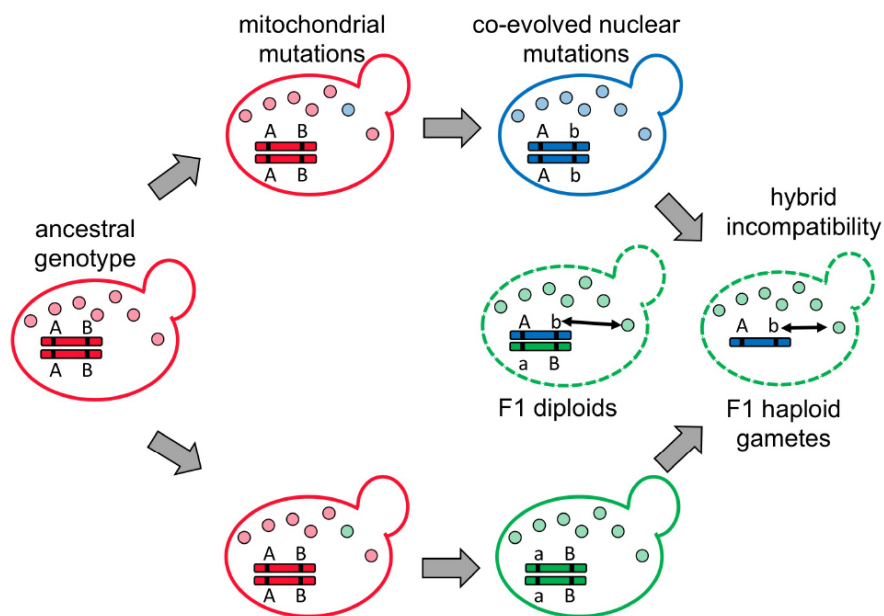


Figura 4. Incompatibilidad híbrida entre especies diferentes de levadura. En donde se observa que la población roja inicial diverge en dos poblaciones, adquiere mutaciones mitocondriales y ocurre un proceso de coadaptación entre el genoma mitocondrial y nuclear en las dos poblaciones, dando como resultado una incompatibilidad híbrida entre genomas en los híbridos de ambas especies (tomado de Chou & Leu, 2010).

1.1.3.4 Inviabilidad híbrida en múridos

Se han realizado experimentos para probar la inviabilidad híbrida mitoconuclear en múridos. McKenzi *et al.* (2003) crearon cíbridos de múridos de diferentes especies: *Mus caroli*, *Mus pahari*, *Rattus norvegicus* y *Otomys irroratu* con el fondo nuclear de *Mus musculus domesticus* mediante la inserción de mitocondrias exógenas en células sin ADN mitocondrial. Dentro del objetivo del experimento se planteó observar los efectos sobre la cadena respiratoria. Para ello, se midió la producción de lactato y la actividad enzimática de OXPHOS.

Los resultados mostraron que los cíbridos con especies más cercanas filogenéticamente al fondo nuclear no tuvieron un efecto significativo sobre la eficiencia de la cadena respiratoria. Pero en el caso de *Rattus norvegicus* y *Otomys irroratu*, que eran especies más distantes, se observó disfunción en OXPHOS y aumento de la producción de lactato, indicando un defecto en la fosforilación oxidativa (Figuras 9 y 10). Los complejos disfuncionales de la cadena respiratoria en *Rattus norvegicus* fueron el I y el III, con una actividad enzimática del 44% y 37% respectivamente en relación con la actividad del control. En *Otomys irroratu* los complejos afectados fueron el III, IV y I con una actividad enzimática del 2%, 44% y 72%, respectivamente, en relación con la actividad del control (McKenzie, Chiotis, Pinkert, & Trounce, 2003).

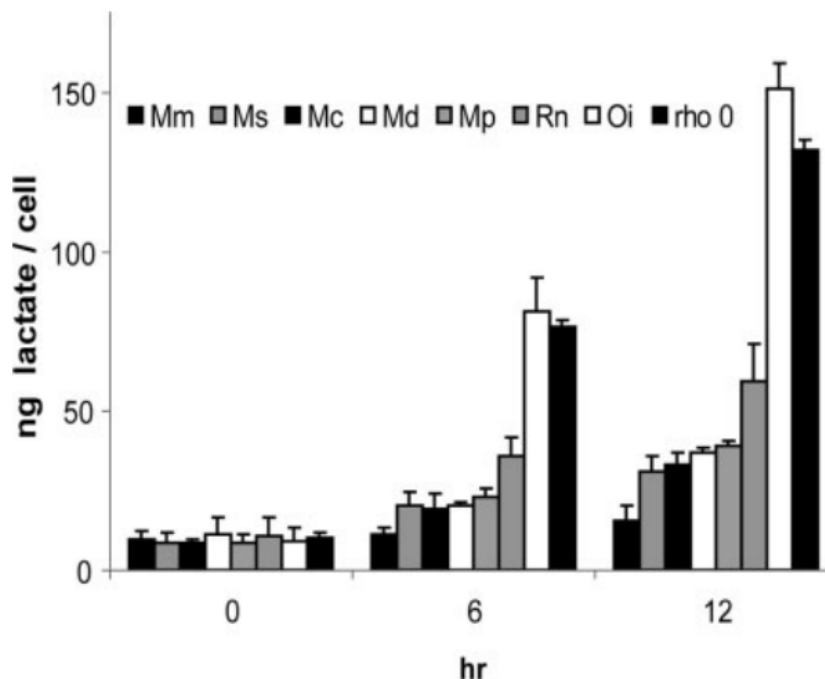


Figura 5. Producción de lactato en cíbridos múridos y células sin mitocondria (rho) 0. Los defectos de OXPHOS se observaron en los cíbridos *Rattus* (Rn) y *Otomys* (Oi) como un aumento de 4 veces y 10 veces, respectivamente, en comparación con el control *M. m domesticus* (Mm)

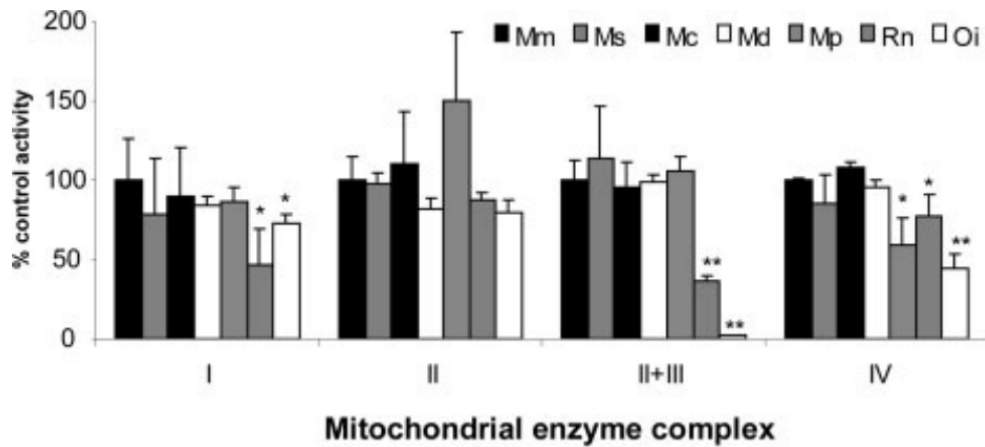


Figura 6. Actividades enzimáticas de OXPHOS en los cíbridos. El cíbrido de *Rattus norvegicus* (Rn) mostró deficiencia en los complejos I, II+III y defecto parcial en IV. El de *Otomys irroratus* (Oi) mostró un defecto en los complejos I y II+III (McKenzie et al., 2003)

En su trabajo, los autores hacen mención de la disfunción del complejo III, e hipotetizan que es debido a la interacción incompatible del citocromo b, ya que el citocromo b es la única subunidad codificada en el núcleo, con las subunidades nucleares vecinas citocromo c1 y "Rieske" una proteína hierro-azufre. Con respecto al complejo I, describieron que las subunidades mitocondriales en el fondo nuclear fueron capaces de ensamblar y retener la actividad respiratoria.

1.1.3.5 Inviabilidad híbrida en copéodos

Los estudios previos descritos han sido hechos con cíbridos de organismos de diferentes especies (aunque de la misma familia). En contraste, los estudios de la incompatibilidad híbrida en copéodos se han realizado en organismos de la misma especie. Se optó por estos organismos debido a que poseen una gran divergencia en el ADN mitocondrial entre poblaciones, pero son capaces de producir híbridos. Ellison y Burton, en 2007, generaron híbridos de *Tigriopus californicus* de tres poblaciones diferentes. Midieron la fecundidad, la supervivencia y la producción de ATP. Posteriormente observaron una reducción en la adecuación, así como una disminución en la tasa de producción de ATP en las crías F3 en comparación con los padres. En la descendencia F1 se observó un aumento en la adecuación debido al fenómeno de vigor híbrido en el que los copéodos tienen al menos una copia haploide del ADN nuclear materno coadaptado.

Finalmente, se realizó un retrocruzamiento entre la descendencia F3 y las líneas parentales para observar si había recuperación en la producción de ATP y la

adecuación. En el caso de los retrocruzamientos paternos no hubo recuperación, pero en los retrocruzamientos maternos se produjo la recuperación del fenotipo a medida que el material genético nuclear coadaptado interactuó con el ADN mitocondrial materno y se recuperó la adecuación (Figura 11) (Ellison & Burton, 2008b).

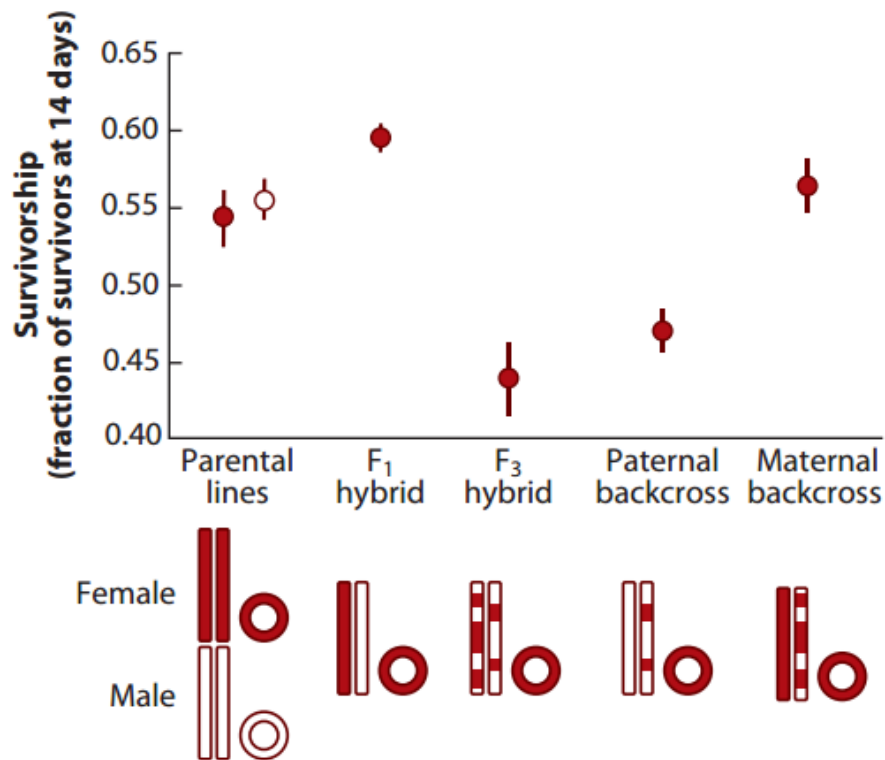


Figura 7. Inviabilidad híbrida en copépodos. Se observa una disminución de la adecuación en híbridos F₃ en los que está presente el material genético paterno no coadaptado. La adecuación se recupera en el retrocruzamiento materno en el que se recupera la mayoría del material genético nuclear coadaptado materno. El ADN nuclear se representa por cromosomas lineales, en tanto que el mtADN de forma circular. En rojo para una población y en blanco para otra. Los genomas nucleares y mitocondrial coadaptados se muestran del mismo color (Ellison & Burton, 2008b).

Se hace mención que los efectos percibidos fueron sobre la producción de ATP, concluyendo así que las incompatibilidades entre ambos genomas tienen lugar en el sistema de transporte de electrones (Burton, Pereira, & Barreto, 2013).

Específicamente, se identificó que la incompatibilidad está mediada por una ARN polimerasa, codificada en el núcleo (mtRPOL), encargada de la transcripción del genoma mitocondrial y responsable de la afinidad y unión de factores de transcripción a la región promotora en la mitocondria. Siendo que en los copépodos híbridos existen incompatibilidades entre la mtRPOL y el ADN mitocondrial, repercutiendo en

los niveles de expresión de genes involucrados en la fosforilación oxidativa codificados en la mitocondria (Ellison & Burton, 2008a).

1.1.4 Desequilibrio

El término desequilibrio en genética es comúnmente utilizado para referirse a la asociación no aleatoria entre dos alelos, genotipos o marcadores genéticos. Dependiendo del contexto genético, esta definición podría comprender el desequilibrio de ligamiento o de ligado (LD, por sus siglas en inglés) y el desequilibrio citonuclear.

1.1.4.1 Desequilibrio de ligado

El desequilibrio de ligado puede definirse como la asociación no aleatoria de dos alelos que están ligados físicamente y se heredan juntos debido al hecho de que pertenecen al mismo haplotipo (Slatkin, 2008).

El desequilibrio de ligado puede ser representado con la siguiente formula:

$$D_{ij} = P_{AiBj} - P_{Ai}P_{Bj} [1]$$

Donde D_{ij} representa el coeficiente de desequilibrio de ligamiento, y se estima restando la multiplicación de cada frecuencia alélica ($P_{Ai}P_{Bj}$) de la frecuencia haplotípica (P_{AiBj}). Cuando D no es igual a cero, se establece un desequilibrio de ligado. Al contrario, cuando D es igual a cero, se establece el equilibrio, lo que significa que no hay asociación física entre los alelos A y B, y ambos alelos se heredan de manera independiente (Slatkin, 2008).

1) Estimación de LD normalizado (D')

Zhao y col. (2007) desarrollaron una fórmula normalizada que se basa en la propuesta por Lewontin (Lewontin, 1964), en la que se toman las frecuencias alélicas de un loci específico y se normalizan los valores obtenidos en base al mayor valor calculado de D_{ij} (Zhao, Nettleton, & Dekkers, 2007).

$$D' = \sum_{i=1}^k \sum_{j=1}^m (pA_i)(pB_j) \left| \frac{D_{ij}}{D_{ij}^{max}} \right| \quad [2]$$

En donde k y m representan el número de alélos en los locus A y B. pA_i y pB_j son las frecuencias alélicas en cada loci. Ambas tasas de desequilibrio de ligado D_{ij} y D_{ij}^{max} se calculan usando la formula [1].

La fórmula [3] estima un valor de correlación entre A_i y B_j basándose en las frecuencias alélicas

$$r^2 = \sum_{i=1}^k \sum_{j=1}^m (pA_i)(pB_j)r_{ij}^2 \quad [3]$$

Donde:

$$r_{ij}^2 = \frac{D_{ij}^2}{(pA_i)(1 - (pA_i))(pB_j)(1 - (pB_j))} \quad [4]$$

El desequilibrio citonuclear fue definido por Schnabel y Amussen (1989) como la asociación no azarosa entre alelos nucleares y alelos del material genético citoplasmático. Esta definición fue establecida mediante el estudio de asociación de alelos entre los genomas nuclear, mitocondrial y cloroplástico en plantas. En dicho estudio, se describió que ciertos casos de esterilidad masculina en plantas se deben a la interacción de uno o más genes nucleares con un locus mitocondrial (Schnabel & Asmussen, 1989).

El estadístico usado se basó en las frecuencias alélicas en loci de cada genoma. “*Un locus autosomal con dos alelos A y a, un locus haploide mitocondrial con dos tipos (M y m) y un locus de cloroplasto con dos tipos (C y c)*” (Schnabel & Asmussen, 1989).

Con base en esto se definieron los posibles genotipos formados por los citotipos y los genotipos nucleares (Tabla 1).

Tabla 1. Posibles genotipos entre el genoma nuclear y los citotipos mitocondriales y de cloroplasto.

Citotipo	AA	Aa	Aa	Total
M/C	u_{11}	v_{11}	w_{11}	x_{11}
M/c	u_{12}	v_{12}	w_{12}	x_{12}
m/C	u_{21}	v_{21}	w_{21}	x_{21}
m/c	u_{22}	v_{22}	w_{22}	x_{22}
Total	u	v	w	1.0

Asimismo, se propuso la siguiente fórmula para estimar el desequilibrio entre tres locus:

$$D_{AA/MC} = freq(AA/M/C) - freq(AA)freq(M/C) \quad [5]$$

Esta asociación no azarosa se ha atribuido tanto a interacciones epistáticas (Fields, McCauley, McAssey, & Taylor, 2014) como a fuerzas evolutivas, que incluyen la migración, la hibridización y los efectos de adecuación (Latta, Linhart, & Mitton, 2001; Schnabel & Asmussen, 1989).

Si se toma la definición de desequilibrio citonuclear, en un contexto de interacción específica entre núcleo y mitocondria, el desequilibrio mitonuclear (DMT) se podría definir como “la asociación no azarosa entre alelos nucleares y mitocondriales”, en donde las interacciones epistáticas serían la causa principal.

La búsqueda de desequilibrio mitonuclear en poblaciones humanas se ha llevado a cabo mediante el uso de los estimadores D' y r^2 , a pesar de que los alelos mitocondriales y nucleares no estén ligados de manera física (Sloan, Fields, & Havird, 2015).

1.1.5 Interacción proteína-proteína

Las interacciones entre núcleo y mitocondria se pueden llevar a cabo a nivel de proteínas. Estas interacciones están involucradas en su mayoría en la producción de energía para el funcionamiento de la célula (Nicholls, 2002).

Como fue mencionado, el 99% de las proteínas de la matriz mitocondrial están codificadas en el núcleo e importadas a la mitocondria, siendo que en la mitocondria únicamente se encuentran codificadas proteínas de los complejos del proceso metabólico de la fosforilación oxidativa. La ruta metabólica de OXPHOS tiene como función principal la síntesis de ATP, para proporcionar energía a la célula, por medio de esta molécula. Para lograrlo, esta ruta se compone de cuatro complejos en los cuales existe una transferencia de electrones que a su vez da lugar al bombeo de protones al espacio intermembranal, formando así un gradiente de protones que es utilizado por la ATP sintasa, para la síntesis de ATP. Asimismo, para lograr el transporte de electrones, la ruta utiliza las siguientes coenzimas: ubiquinona en la membrana lipídica y citocromo c en el espacio intermembranal (J.-Q. Chen, Cammarata, Baines, & Yager, 2009).

Los complejos de fosforilación oxidativa (OXPHOS) están constituidos por proteínas codificadas en ambos genomas, con la excepción del complejo II, en el que todas sus subunidades están codificadas en el genoma nuclear. Estos complejos se encuentran en la extensión tubular de la membrana interna, conocida como la cresta. Dado que el correcto acoplamiento de los complejos es esencial para el funcionamiento de la ruta de OXPHOS, debe existir una coordinación de cambios evolutivos (coevolución) de aminoácidos entre las proteínas mitocondriales (Rand, Haney, & Fry, 2004), siendo entonces un sistema de suma importancia tanto para el estudio de las interacciones mitonucleares como de procesos coevolutivos.

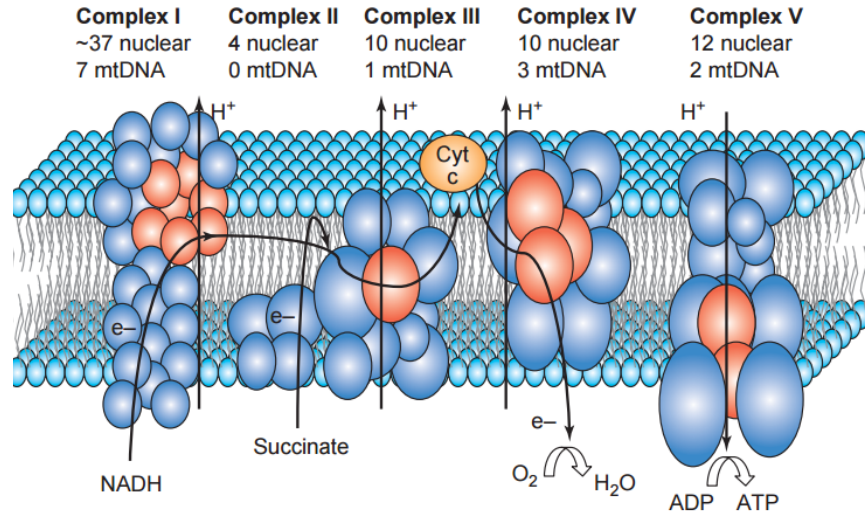


Figura 8 Complejos de OXPHOS. LA ATP sintasa está indicada como el complejo V. De color rojo las proteínas codificadas en la mitocondria, de azul las codificadas en el núcleo (Rand et al., 2004)

1.1.6 Interacción ADN-proteína

La interacción mitonuclear puede darse también a nivel de proteína y ADN, más específicamente entre ADN mitocondrial y proteínas codificadas en el núcleo. Dado que el genoma mitocondrial no codifica toda la maquinaria necesaria para la regulación transcripcional, diversos factores de transcripción son codificados en el núcleo e importados a la mitocondria, dando lugar a una interacción ADN-proteína.

La regulación transcripcional mitocondrial tiene lugar en una región conocida como D-loop, de aproximadamente 1kb, en la cual inician los procesos de transcripción y replicación de la mitocondria, por lo que es conocida también como región de control. Es ahí donde los factores de transcripción los cuales se encuentran codificados en el núcleo e importados a la mitocondria, se unen al ADN mitocondrial para llevar a cabo la regulación de los genes mitocondriales (Yamamoto, 2001).

Hay dos sitios principales de inicio de la transcripción en la mitocondria, IT_{H1} ubicado en la cadena H e IT_L , ubicado en la cadena L (Taanman, 1999). Los factores de transcripción codificados en el núcleo pueden actuar de dos formas: directa e indirectamente. La regulación indirecta mitocondrial consiste en la regulación transcripcional de genes nucleares de relevancia para la biogénesis y funcionamiento de la mitocondria, como es el caso de los factores de transcripción NRF1 y NRF2 que regulan los genes que metabolizan antioxidantes y xenobióticos (Ohtsuji et al., 2008). La regulación mitocondrial directa se lleva a cabo mediante la regulación transcripcional de genes mitocondriales. Algunos factores de

transcripción que regulan directamente en la mitocondria son TFAM y TFB2M (Leigh-Brown, Enriquez, & Odom, 2010).

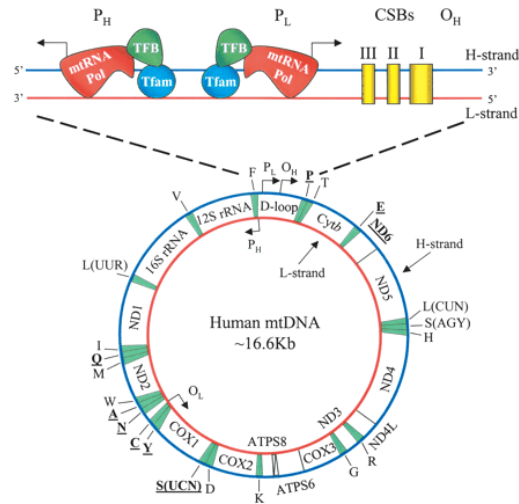


Figura 9 . Región D-loop mitocondrial, factores de transcripción codificados en el núcleo se unen en esta región para regular la transcripción de genes mitocondriales (Genetical Society (Great Britain) & Scarpulla, 1987)

Se han identificado interacciones entre enzimas codificadas en el núcleo con funciones asociadas a la regulación de procesos homeostáticos del ADN mitocondrial. Este es el caso del gen *ACO1P* identificado en levadura, el cual corresponde a una aconitasa, cuya función canónica es la catálisis de la esteroisomerización de citrato a isocitrato en el ciclo de Krebs, tiene además una función independiente relacionada con el mantenimiento de la estabilidad mitocondrial, y se propone que actúa en el empaquetamiento del ADN mitocondrial (X. J. Chen, Wang, Kaufman, & Butow, 2005).

De igual manera, se han identificado otros genes con actividad de regulación de biogénesis y mantenimiento mitocondrial: ATAD3 y PHB (prohibitina) se involucran de manera cooperativa en el mantenimiento del ADN mitocondrial, así como su replicación y traducción (He *et al.*, 2012). La actividad B-actina y la cadena pesada de miosina IIA se asociaron con el número de copias mitocondriales (Reyes *et al.*, 2011). REV3, una ADN polimerasa que se localiza en la mitocondria se ha identificado participa en la homeostasis mitocondrial (Singh *et al.*, 2015). Lig3 que codifica para una ligasa, cuya función descrita es la reparación de ADN y que de manera conjunta con el gen *Xrcc1* se encarga de la reparación del ADN nuclear, también está presente en la mitocondria y de igual manera se encarga de la reparación de ADN mitocondrial, aunque sin requerir del gen *Xrcc1* (Gao *et al.*, 2011).

La proteasa ClpX regula la segregación mitocondrial, fungiendo como chaperona de la proteína TFAM, que regula de manera directa la segregación uniforme del ADN mitocondrial en la células (Kasashima, Sumitani, & Endo, 2012). Asimismo, otra proteasa, la proteasa Lon, se ha involucrado en regulación del mantenimiento de ADN mitocondrial, al regular la tasa de TFAM:mtADN en la mitocondria, siendo de gran importancia para la homeostasis mitocondrial (Matsushima, Goto, & Kaguni, 2010).

Las interacciones descritas hasta el momento ejemplifican la complejidad de la interacción núcleo-mitocondrial.

1.1.7 Interacción ARN-proteína

La interacción entre ARN y proteínas de núcleo y mitocondria se da principalmente en la conformación del ribosoma mitocondrial (mitoribosoma) y la maquinaria de traducción. El mitoribosoma se compone de ARN ribosomal el cual está codificado en la mitocondria y los componentes protéicos codificados en el núcleo, 80 aproximadamente (Amunts, Brown, Toots, Scheres, & Ramakrishnan, 2015), específicamente dos moléculas de ARN ribosomal se encuentran codificadas en la mitocondria: 12S rARN y 16S rARN (Smirnov, Entelis, Martin, & Tarassov, 2011).

A su vez, otras moléculas de ARN codificado en mitocondria interactúan con el mitoribosoma, estas moléculas son los ARN de transferencia, 22 de los cuales están codificados en el ADN mitocondrial y participan en la traducción de proteínas mitocondriales (Amunts *et al.*, 2015). Otra interacción está dada entre las enzimas aminoacil t-ARN sintetasas que están codificadas en el núcleo y son importadas a la mitocondria, y los tARN codificados en el ADN mitocondrial, que interactúan para llevar a cabo la formación de los aminoacil-tARN (los tARN cargados con su respectivo aminoácido) (Konovalova & Tynismaa, 2013).

1.1.8 Regulación mediada por ARN

Otra interacción descrita entre el núcleo y mitocondria es la regulación de procesos biológicos mediados de manera indirecta por ARN.

Dhir *et al.* (2018) describen un fenómeno en el que la acumulación de ARN de doble cadena mitocondrial desencadena reacciones inmunes. Asimismo, se hace mención de los mecanismos para la regulación de ADN mitocondrial, específicamente en la regulación de acumulación de ARN de doble cadena, en el que la ARN helicasa mitocondrial (*SUV3*) y la polinucleótido fosforilasa (*PNPasa*) juegan un papel importante, ya que cuando estos elementos son reprimidos, se genera una acumulación de ARN de doble cadena. Dicha acumulación es censada

por MDA5, una helicasa de ARN de doble cadena, y la proteína antiviral de señalización MAVS, activando así una cascada de señalización mediada por la respuesta de interferón tipo I, teniendo como consecuencia la activación de la respuesta inmune (Dhir *et al.*, 2018).

Otro proceso que ha sido investigado es la regulación por medio de los microARN (miARN). La función descrita de los microRNA ha sido la inhibición de genes blanco que son reconocidos por estos ARN pequeños no codificantes de una sola cadena. Se han identificado que tienen actividad regulatoria dentro de la mitocondria, de los cuales algunos son sintetizados en la misma mitocondria, pero, en su mayoría, son importados del núcleo. Se ha descrito que estos miARN mitocondriales actúan sobre las rutas de fosforilación oxidativa, cadena transportadora de electrones y metabolismo de lípidos (Duarte, Palmeira, & Rolo, 2014).

De igual manera, los ARN largos no codificantes (lncARN), tienen actividad regulatoria reportada dentro de procesos biológicos mitocondriales. Como es el caso del lncARN *CUT60*, el cual funge como regulador indirecto en el mantenimiento del ADN mitocondrial. Este lncARN regula la represión funcional del gen *ATP16*, que cuando es reprimido resulta en un fenotipo en el cual se pierde el ADN mitocondrial. El mecanismo descrito de *CUT60* radica en la correcta terminación transcripcional de este ARN, de no lograrse, se produce un ARN quimérico (“read-through transcript”) en la región promotora del gen *ATP16* reprimiendo dicho gen. Los autores argumentan que el fenotipo tan drástico observado es debido a que el gen *ATP16* es una subunidad de la ATP sintetasa mitocondrial (du Mee, Ivanov, Parker, Buratowski, & Marquardt, 2018).

1.1.9 interacciones ADN-ADN: Inserciones mitocondriales en núcleo

La interacción entre ambos genomas puede darse mediante inserciones de fragmentos de ADN mitocondrial en el genoma nuclear. Estas inserciones nucleares de ADN mitocondrial, o NUMT por sus siglas en inglés, es un fenómeno que ocurre en células eucariotas. En humanos, se han identificado por lo menos 612 inserciones (Hazkani-Covo, Zeller, & Martin, 2010). Se ha determinado un total de 53 NUMT específicos para humanos utilizando la base de datos USCS (Kent *et al.*, 2002). También se ha observado estructura poblacional entre nueve NUMT específicos de humanos mediante la construcción de un árbol de *neighbor-joining* utilizando las distancias genéticas de los NUMT seleccionados de 13 poblaciones (parte del panel de diversidad de la base de datos de 1000 genomas). La topología resultante del árbol concordó con la distribución geográfica previamente reportada de poblaciones en el panel de los 1000 genomas (Lang *et al.*, 2012). En otro estudio realizado por Ricchetti y col., (2004) estimaron una tasa de inserción de 5.4 inserciones por millón de años en el genoma humano, basada en las comparaciones genómicas de inserciones con chimpancé (Ricchetti, Tekaia, & Dujon, 2004).

Se han propuesto varios mecanismos sobre cómo estos fragmentos mitocondriales se insertan en el genoma nuclear: lisis del compartimento mitocondrial, fusiones de membranas mitocondriales y nucleares, y encapsulación del ADN mitocondrial dentro del núcleo, siendo la vía más aceptada la degradación de mitocondrias anormales mediada por una vacuola, en el que el ADN mitocondrial se inserta por un mecanismo de rupturas de ADN de doble cadena (Hazkani-Covo *et al.*, 2010).

Por otra parte, los NUMT se han asociado a enfermedades humanas, pues algunas inserciones se dan dentro del marco de lectura del gen, causando mutaciones en el sitio de empalme *splice site mutations*. Hay dos casos reportados de NUMT en genes asociados a enfermedades: uno en el factor VII plasmático (causando una deficiencia en este factor) y el otro la mucopolipina 1 (MCOLN1, que causa mucopolipidosis IV). También se han reportado inserciones en los exones teniendo como consecuencia un codón de terminación prematuro, como son los casos del NUMT en el gen de GLI3 causando el síndrome de Pallister-Hall, y el NUMT en el de USH1C que causa el síndrome de Usher (Hazkani-Covo *et al.*, 2010). Asimismo los NUMT se han identificado como una fuente de falsos positivos en los estudios de heteroplasmia ya que regiones nucleares con este tipo de inserciones se amplifican y se confunden con ADN mitocondrial (Parr *et al.*, 2006).

1.1.10 Incompatibilidad y desequilibrio mitonuclear en poblaciones humanas

En humanos, el estudio de las interacciones entre los genomas mitocondrial y nuclear, así como la búsqueda de incompatibilidades, se ha llevado a cabo tanto de manera experimental como *in silico*.

Se han realizado experimentos con híbridos, enfocados en los haplogrupos mitocondriales H y J. Siendo H el de mayor frecuencia en poblaciones de Eurasia Occidental y J en poblaciones de Oriente próximo y norte de África. El estudio experimental fue motivado por el hecho de haber identificado susceptibilidad a desarrollar degradación macular asociada a edad (AMD, por sus siglas en inglés), enfermedad multifactorial, con diversos factores genéticos y ambientales, los cuales involucran: dieta, fumar, exposición a luz, radiación ionizante y por luz ultravioleta. en poblaciones con el haplogrupo mitocondrial H. Los híbridos fueron realizados a partir de la fusión de líneas celulares *ARPE-19 Rho0* (células sin ADN mitocondrial) y plaquetas con los haplogrupos H y J. Se identificó expresión diferencial entre híbridos con haplogrupos H y J en genes involucrados en rutas de: Sistema complemento (*CD55*, *CD59*, *C1QC*, *C1S*, *C4B*, *C4BPB* y *CFHR4*); Pro-inflamatorio (*TGFB2* y *IL-6*) y Apoptosis (*BBC-3*, *BCL2L13* y *RARA*) (Kenney *et al.*, 2014).

Posteriormente, se hicieron experimentos sobre los mencionados híbridos y su respuesta a un tratamiento de luz UV. Se midieron perfiles de expresión en genes identificados por Kenney y col. (2014), encontrando también expresión diferencial

entre ambos cíbridos tratados. Asimismo, se evaluó el crecimiento celular de cíbridos tratados y no tratados, observando una tasa de crecimiento menor en cíbridos tratados con radiación UV para los dos haplogrupos H y J. No obstante, los cíbridos con haplogrupo J tuvieron las tasas más altas de crecimiento celular en ambas condiciones, fenómeno que los autores reportaron con anterioridad hipotetizando que se debe a que los cíbridos de H producen altos niveles de ATP y utilizan predominantemente OXPHOS, mientras que los de J utilizan la glicólisis (Malik *et al.*, 2014).

Growth Curve H vs. J Cybrids Treated with 10mJ UV Radiation

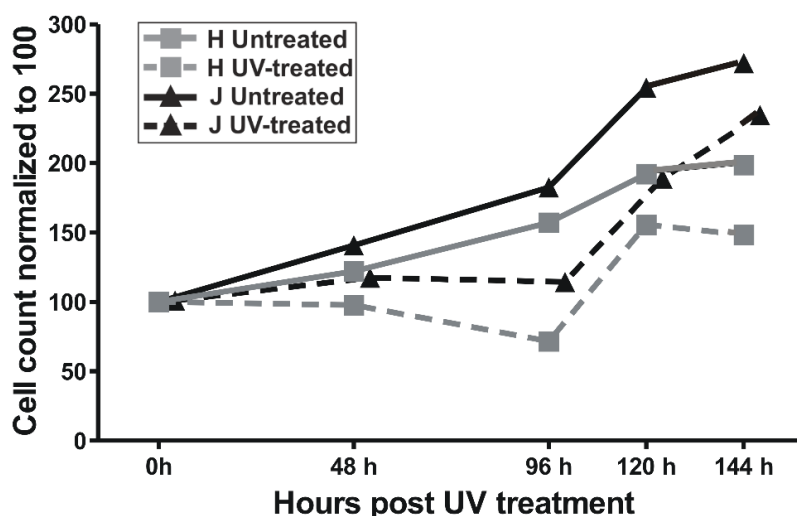


Figura 10. Curvas de crecimiento de cíbridos con haplogrupos mitocondriales H y J tratados con radiación UV y controles sin tratamiento de radiación UV (Malik *et al.*, 2014).

El estudio realizado se resume en la Figura 11, que esquematiza la expresión diferencial entre cíbridos, y cómo el factor ambiental de luz UV tiene consecuencias diferentes en cada cíbrido, según el haplogrupo mitocondrial, a nivel de expresión de genes de crecimiento celular, inflamación y apoptosis.

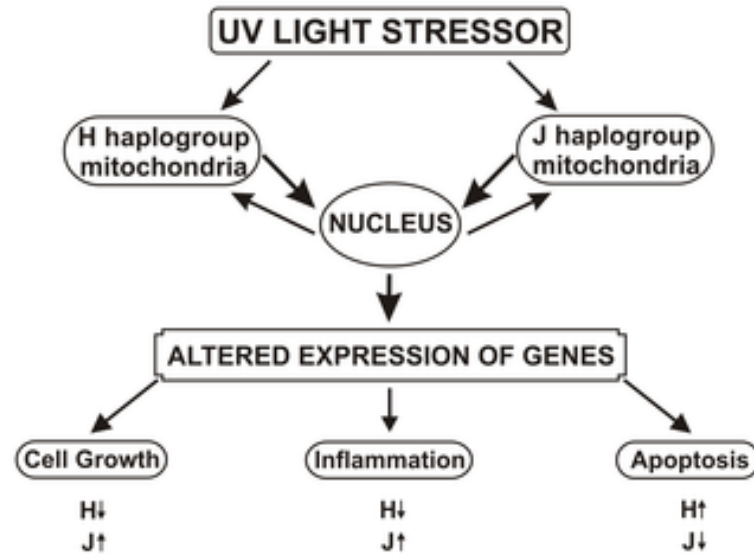


Figura 11. Resumen esquematizado de los estudios de cíbridos con los haplogrupos H y J donde se describe de que manera afecta a cada haplogrupo la exposición a luz UV (tomado de Malik *et al.*, 2014).

El estudio realizado por Malik y col. (2014), fue un primer intento de tratar de asociar un fenotipo, producto de la interacción de núcleo y mitocondria, con mutaciones propias de haplogrupos mitocondriales de manera experimental; sin embargo, es necesario tomar en cuenta el fondo genético nuclear, un aspecto que se ha observado en estudios en donde se han identificado mutaciones en el ADN mitocondrial ligadas a cáncer.

Por ejemplo, Petro y col. (2005) identificaron 38 mutaciones con efectos patogénicos relacionadas con cáncer de próstata, utilizando células epiteliales de cáncer de próstata, donde algunas correspondieron a haplogrupos mitocondriales: H15, D5 y M13. Asimismo se identificaron mutaciones fuertemente asociadas en el gen que codifica para citocromo oxidado subunidad I (*COI*), en el que previamente se han descrito mutaciones con otros tipos de cáncer. Cabe mencionar que en este estudio se utilizó la técnica de generación de cíbridos con las mutaciones previamente identificadas para la evaluación de su efecto (Petros *et al.*, 2005).

Sin embargo, posteriormente, Salas y col. (2005) hicieron una crítica en base a que las mutaciones identificadas asociadas a cáncer de próstata corresponden a los haplogrupos mitocondriales H15, común en Europeos, y D5 y M13, comunes en Asiáticos del este, pero que no existen reportes en la población japonesa masculina de un incremento en casos de cáncer de próstata, siendo que en esta población ambos haplogrupos D5 y M13 se encuentran en elevadas frecuencias (Salas *et al.*, 2005).

Este caso dio lugar a la formulación de una hipótesis basada en la interacción necesaria que debe existir entre el núcleo y mitocondria para el desarrollo de

tumores. Santos y col. (2008) hacen mención que en el estudio de mutaciones mitocondriales asociadas a tumorigénesis es necesario tomar en cuenta el fondo genético nuclear, debido a las interacciones que se dan entre ambos genomas; así mismo, proponen que estas interacciones podrían explicar la discrepancia observada en el estudio de Petros y col. (2005) con relación al cáncer de próstata (Santos *et al.*, 2008).

Una aproximación al estudio de las interacciones mitonucleares se ha basado en la genómica poblacional y datos de secuenciación masiva en experimentos *in silico*, en los que se ha tenido como objetivo encontrar un desequilibrio entre alelos de los genomas nucleares y mitocondriales y detectar una posible incompatibilidad mitonuclear en poblaciones humanas.

El primer estudio en el que se intentó encontrar desequilibrio de ligado mitonuclear en poblaciones humanas fue llevado a cabo por Sloan y col. (2015). Se utilizaron 583,300 SNP del Proyecto de Diversidad del Genoma Humano (HGDP, por sus siglas en inglés) que comprende 1,043 individuos de 53 poblaciones y 327 156 SNP localizados en genes núcleomitocondriales, tales como subunidades de OXPHOS, tARN sintetasas y ARN / ADN polimerasas.

Asimismo, se asignaron haplogrupos mitocondriales utilizando la herramienta Haplogrep v. 2.0 (Weissensteiner *et al.*, 2016). Posteriormente se realizó un análisis de desequilibrio de ligado entre los SNP nucleares y los haplotipos mitocondriales. Las herramientas estadísticas utilizadas fueron estadísticos clásicos de desequilibrio de ligado como D' y r^2 (Sloan *et al.*, 2015).

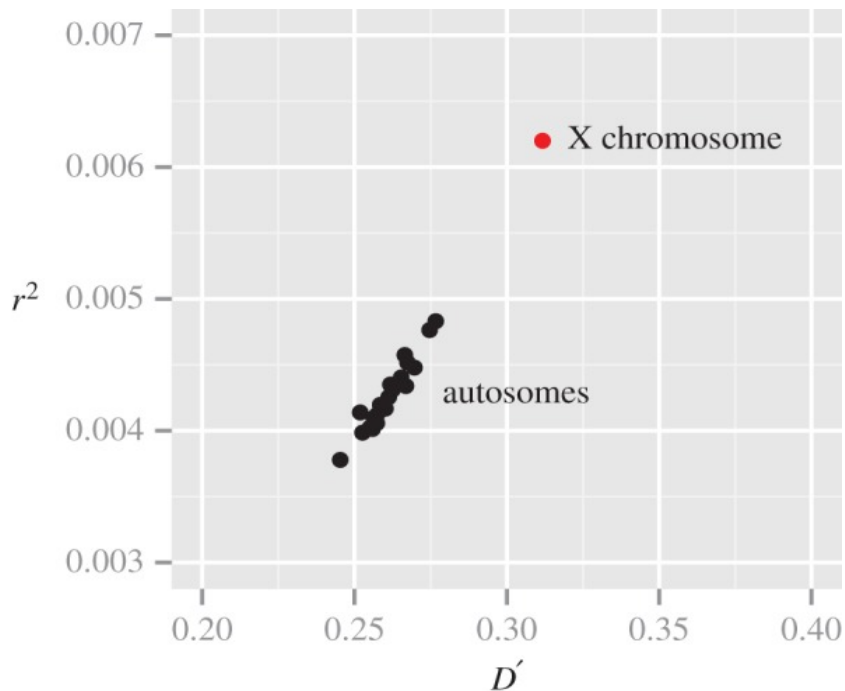


Figura 12. Desequilibrio de ligado mitonuclear analizando la base de datos del HGDP. Cada punto representa la media de los estadísticos de LD utilizados por cromosoma: r^2 estadístico que representa a manera de correlación el desequilibrio de ligado entre los SNP nucleares y los haplogrupos mitocondriales y D' que es el estadístico de desequilibrio de ligado normalizado. (Sloan et al., 2015).

Los autores concluyen que existen niveles bajos pero significativos de LD en los cromosomas autosómicos. En el cromosoma X, la posible explicación de un valor de LD elevado, en comparación con los cromosomas autosómicos, es la diferencia en el tamaño efectivo de la población, el cual correspondería a la mitad del tamaño efectivo de población de los cromosomas autosómicos.

Los autores advierten que sus resultados podrían ser la consecuencia de la naturaleza conservadora de su estadístico. Asimismo, expusieron que el uso de un arreglo de genotipado de SNP podría sesgar los resultados hacia polimorfismos de alta frecuencia, pues los de baja frecuencia no suelen estar representados en estos arreglos y por lo tanto no serían considerados.

El segundo artículo, realizado por Rishishwar y Jordan (2007), fue motivado por las posibles consecuencias de la práctica del reemplazo mitocondrial (MR), que consiste en combinar el ADN nuclear, de una mujer con un trastorno mitocondrial, con una mitocondria no afectada de un donante de óvulos. Los autores mencionan su interés en encontrar posibles implicaciones y existencia de incompatibilidades mitonucleares. Los autores utilizaron la base de datos del Proyecto 1000 Genomas para obtener variantes nucleares y mitocondriales; también usaron variantes de ADN mitocondrial del Proyecto de Diversidad Genética Humana. La idea principal

era contrastar la divergencia del ADN mitocondrial con la del ADN nuclear dentro de una población para encontrar posibles incompatibilidades. Se construyó una filogenia usando el ADN nuclear, con el software PLINK (Purcell et al., 2007) en la que se alinearon los haplogrupos mitocondriales, previamente determinados con HaploGrep2.

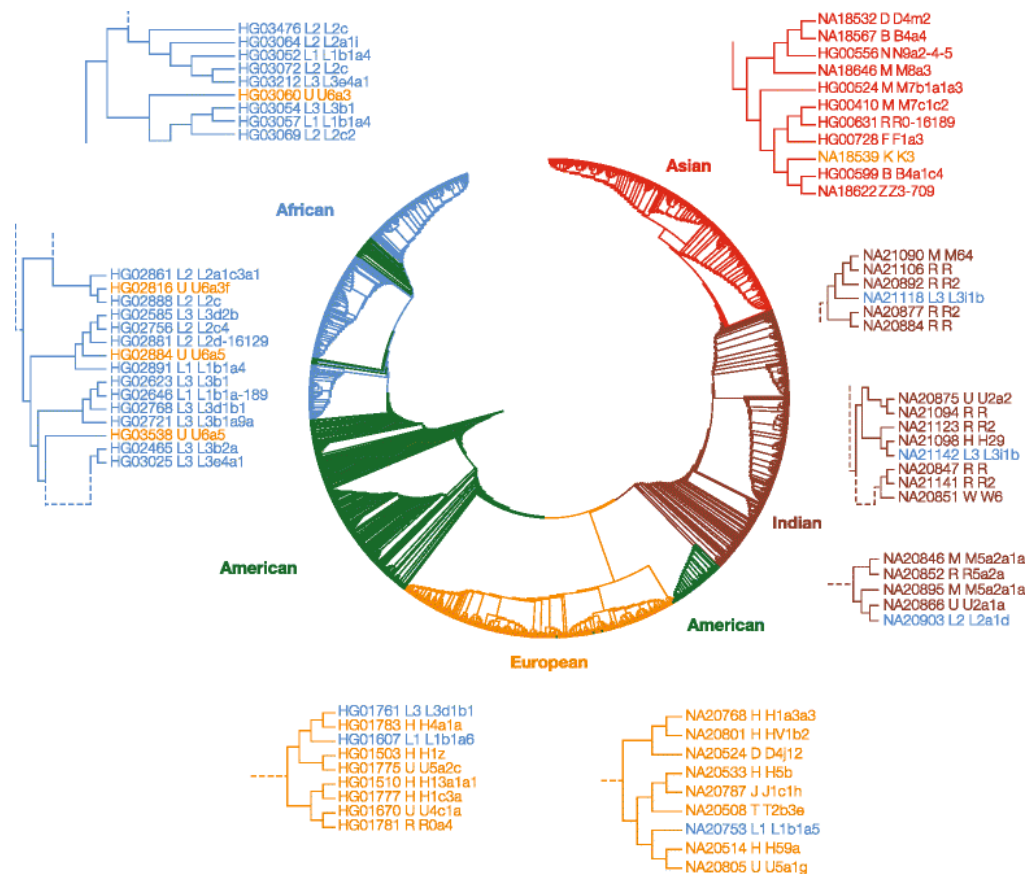


Figura 13. Filogenia de ADN nuclear. Los colores representan cada población. Los haplotipos mitocondriales se superponen en el árbol (Rishishwar & Jordan, 2017).

En la Figura 15 se muestra que la topología del árbol construido a partir de variación nuclear fue la esperada, en la cual se observan *superpoblaciones* (Poblaciones del panel de 1000 genomas agrupados en bloque continentales). Asimismo, se observaron individuos que presentaron discordancia entre sus haplotipos mitocondriales y su distancia filogenética, siendo estos individuos muy cercanos en la filogenia con haplotipos mitocondriales distantes. Los autores concluyen que en la naturaleza coexisten genomas divergentes nucleares y mitocondriales, y como todos los donantes tanto del Proyecto de 1000 Genomas como del *Human Diversity*

Project eran sanos, argumentan que la posible falta de correspondencia entre el ADN nuclear y el mitocondrial no es perjudicial (Rishishwar & Jordan, 2017).

Como se describió anteriormente, ambos trabajos concluyen que la existencia de incompatibilidades entre el genoma nuclear y mitocondrial no presenta riesgo alguno. En la publicación de Sloan y col. (2015), lograron detectar desequilibrio a niveles bajos, lo que esgrimen para concluir la inexistencia de una incompatibilidad mitonuclear con algún riesgo significativo. Rishishwar y Jordan (2017), concluyen lo mismo, con el argumento de que en la naturaleza existen individuos sanos a pesar de ser discordantes entre su distancia nuclear y su haplogrupo mitocondrial dentro de una población. Todo ello a pesar de que el efecto de una incompatibilidad mitonuclear podría no ser letal, sino presentar efectos menos severos, como fue detectado en los experimentos con cíbridos de Kenney y col. (2014) y Malik y col. (2014).

Dado los antecedentes mencionados, se podría concluir que el fenómeno de incompatibilidad mitonuclear podría ser un fenómeno cuyos efectos son sutiles, es decir, ligeramente deletéreos, por lo que para detectar desequilibrio mitonuclear es necesario un análisis con alta resolución en el cual se utilice toda la variación posible en el genoma nuclear y mitocondrial. Por otra parte, los efectos muy deletéreos o letales, no serían identificados en análisis poblacionales, pues como ya apuntaban Rishishwar y Jordan (2017), estos casos serían eliminados en etapas tempranas del desarrollo. Asimismo regiones candidatas a incompatibilidades mitonucleares en humanos se esperaría pudieran ser detectadas en un análisis de alta resolución. Estas posibles incompatibilidades se han observado entre homínidos arcáicos y humanos modernos. Sharbrough y colaboradores (2017) plantean una coevolución en homínidos entre el genoma nuclear y mitocondrial, restringiendo la introgresión en humanos modernos de loci nucleares codificantes para proteínas involucradas en interacciones nucleares y mitocondriales; lo cual corresponde con la hipótesis de incompatibilidad entre el ADN mitocondrial moderno y el genoma nuclear de los homínidos arcáicos de los grupos Neandertal y Denisova (Sharbrough, Havird, Noe, Warren, & Sloan, 2017).

2 Hipótesis

Existe desequilibrio mitonuclear significativo en poblaciones humanas

3 Objetivos

3.1 Objetivo general:

Detectar y caracterizar SNP con desequilibrio mitonuclear significativo en poblaciones humanas

3.2 Objetivos específicos:

- 1) Desarrollar método cuantitativo para detección de desequilibrio mitonuclear.
- 2) Detectar desequilibrio significativo mitonuclear en poblaciones y superpoblaciones humanas
- 3) Caracterizar bioinformáticamente los SNP con desequilibrio mitonuclear en poblaciones y superpoblaciones humanas

4 Materiales y métodos

4.1 Poblaciones e individuos

Las poblaciones utilizadas para el análisis se obtuvieron de la fase 3 del Proyecto de 1000 Genomas (Auton *et al.*, 2015). Los datos consisten en 3 502 individuos secuenciados con una profundidad media de 7.2X en la secuenciación del genoma completo y 65.7X en la secuenciación dirigida a exoma, refiriéndose al término de profundidad (N°X) al total de secuencias promedio que cubren el genoma de referencia. Las muestras se tomaron de 26 poblaciones agrupadas en 5 superpoblaciones: africana (AFR), americana mestiza (AMR), asiática oriental (EAS), europea (EUR) y surasiática (SAS).

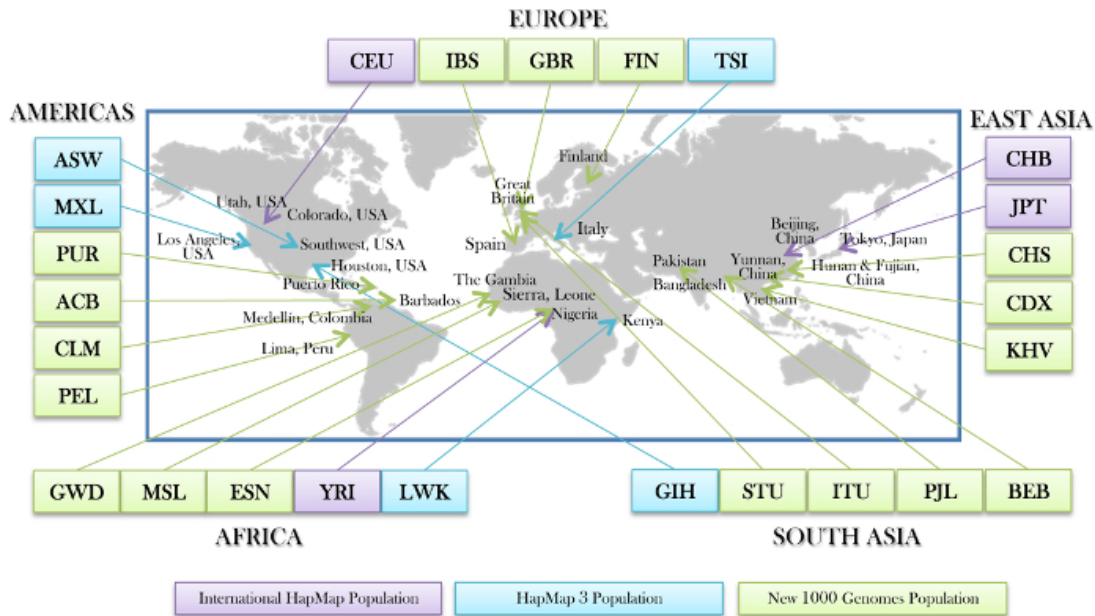


Figura 14. Distribución de muestras de la base de datos de los 1000 genomas (Auton et al., 2015).

Tabla 2. Número total de individuos del panel de los 1000 genomas por población.

Población	Número de individuos	Superpoblación
Chinos Han en Beijing, China (CHB)	108	Asia del Este (EAS)
Japoneses en Tokyo, Japón (JPT)	105	Asia del Este (EAS)
Chinos Han del sur (CHS)	171	Asia del Este (EAS)
Chinos Dai en Xishuangbanna, China (CDX)	109	Asia del Este (EAS)
Ciudad Kinh in Ho Chi Minh, Vietnam (KHV)	124	Asia del Este (EAS)
Residentes de Utah con ascendencia del Norte y Oeste de Europa (CEU)	184	Europeos (EUR)
Toscanos en Italia (TSI)	112	Europeos (EUR)
Finlandeses en Finlandia (FIN)	105	Europeos (EUR)
Británicos en Inglaterra y Escocia (GBR)	107	Europeos (EUR)
Población Ibérica en España (IBS)	162	Europeos (EUR)
Yorubas en Ibadan, Nigeria (YRI)	186	Africanos (AFR)
Luhya en Webuye, Kenya (LWK)	116	Africanos (AFR)
Gambianos en divisiones del oeste de Gambia (GWD)	180	Africanos (AFR)
Mende en Sierra Leona (MSL)	128	Africanos (AFR)
Esan en Nigeria (ESN)	173	Africanos (AFR)
Americanos con ascendencia africana en E.U.A (ASW)	112	Africanos (AFR)
Africanos del caribe en Barbados (ACB)	96	Africanos (AFR)
Residentes de los Ángeles, con ascendencia mexicana (MXL)	107	Americanos mestizos (AMR)
Puertorriqueños de Puerto Rico (PUR)	150	Americanos mestizos (AMR)
Colombianos de Medellín, Colombia (CLM)	148	Americanos mestizos (AMR)
Peruanos de Lima, Perú (PEL)	130	Americanos mestizos (AMR)
Indios Gujarati de Houston, Texas (GIH)	113	Asiáticos del sur (SAS)

Punjabi de Lahore, Pakistán (PJL)	158	Asiáticos del sur (SAS)
Bengalíes de Bangladesh (BEB)	144	Asiáticos del sur (SAS)
Tamil Sri Lankan de Reino Unido (STU)	128	Asiáticos del sur (SAS)
Indios Telugu de Reino Unido (ITU)	118	Asiáticos del sur (SAS)

4.1.1 Filtrado de individuos

Como el trabajo se centra en las interacciones entre el genoma nuclear y el mitocondrial, es importante filtrar a los individuos por su relación familiar, ya que esto podría generar un sesgo en los resultados, pues tendrían una probabilidad mayor de coheredar alelos de ambos genomas, inflando los valores de asociación.

Para el filtrado utilizamos el archivo de pedigrí proporcionado por el proyecto 1000 genomas

(ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20130606_sample_info/20130606_sample_info.xlsx), en el que se menciona el sexo, los *ID* parentales y las relaciones de cada muestra. Seleccionamos personas sin relación con ninguna otra muestra de la base de datos.

También seleccionamos individuos con variantes disponibles, tanto nuclear como mitocondrial, ya que algunos individuos en el panel no tenían secuencias disponibles para ambos genomas.

4.2 Filtrado de cromosomas sexuales

Como fue reportado en el trabajo de Sloan y col. (2015), en el que se midió desequilibrio de ligado con 300 SNP de haplotipos nucleares y mitocondriales, el cromosoma sexual X presentaba valores más altos de D' y r^2 en comparación con los cromosomas autosómicos, argumentando que una posible causa podría ser el tamaño efectivo de la población distinto para este cromosoma. Asimismo, el número de muestras femeninas y masculinas es desigual, y sus ploidías para los cromosomas sexuales de igual manera eran desiguales, por lo que se determinó que los cromosomas sexuales podrían sesgar el análisis de desequilibrio y por lo tanto fueron excluidos del presente análisis.

4.3 Filtrado de variantes y subconjunto de poblaciones

Las variantes de la base de datos de 1000 genomas alineados contra el genoma de referencia Hg19 (GRCh p.13) se proporcionan como un archivo de formato de llamada de variante (VCF) (Danecek *et al.*, 2011)

(<ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/>), en el que se describen todas las variantes para cada individuo, que son: polimorfismo de un solo nucleótido (SNP, por sus siglas en inglés), variantes estructurales como inserciones y deleciones (INDELS) y variantes de número de copia (CNV, por sus siglas en inglés). En el presente análisis solo los SNP fueron seleccionados, descartando las CNV e INDELS. El software utilizado para el filtrado de variantes fue *bcftools* (Li *et al.*, 2009), con los siguientes parámetros:

```
bcftools view -v SNP --max-allele 2 <file.vcf.gz> > <filter_file.vcf>
```

El parámetro 2 de la opción "max-allele" se aplicó para seleccionar solo sitios bialélicos, descartando variantes multialélicas.

Adicionalmente, fueron seleccionados un número de alelos diferentes a la referencia, mediante dos opciones diferentes:

- a) **-min-ac 1**: SNP con una frecuencia mínima de 1
- b) **-min-ac 3**: SNP con una frecuencia mínima de 3

Para el subconjunto de cada población, fueron usados los ID proporcionados de cada muestra:

(ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20130606_sample_info/20130606_sample_info.xlsx).

Se utilizó *bcftools* para seleccionar el subconjunto de población, con las siguientes opciones:

```
bcftools view --samples-file <sample.ids> <file.vcf.gz> > <pop.file.vcf>
```

4.4 Estadístico *tau* de Goodman-Kruskal

Este estadístico describe la asociación entre variables nominales, en la que X predice Y, el cual pretende demostrar una dependencia o independencia entre ambas variables. El estadístico se basa en la estimación del error de probabilidad

de predecir la categoría Y sin el conocimiento de la categoría X ($P(B)$) y con el conocimiento dado de la categoría X ($P(A)$) (Goodman & Kruskal, 1972),

$$\tau_{yx} = \frac{P(B) - P(A)}{P(B)} \quad [6]$$

La estimación de los errores de probabilidad en las predicciones usa las frecuencias de las categorías en la variable Y ($a_{i.}$) y X ($a_{.j}$) y los pares de categorías XY (a_{ij}).

La estimación de error de probabilidades puede ser calculada en base a las frecuencias de cada categoría, como lo propone Somers (1962), utilizando las siguientes formulas:

La probabilidad de error en predecir la variable Y, sin conocimiento de la variable X está dada por:

$$P(B) = \sum_j \left(\frac{a_{i.}}{n} \right) \left(1 - \frac{a_{i.}}{n} \right) \quad [7]$$

Donde n es el número de categorías.

La probabilidad de error en predecir la variable Y, con conocimiento de la variable X está dada por:

$$P(A) = \sum_j \left[\sum_i \left(\frac{a_{ij}}{a_{.j}} \right) \left(\frac{a_{.j} - a_{ij}}{a_{.j}} \right) \right] \left(\frac{a_{.j}}{n} \right) \quad [8]$$

Para el análisis de desequilibrio mitonuclear se aplicó el estadístico *tau* de Goodman-Kruskal, para detectar la asociación entre los alelos mitocondriales (Y) y nucleares (X) siendo para las categorías Y (1,0) como alelos haploides y para X las categorías (1/1, 0/1, 0/0) como alelos diploides, descartando variantes multialélicas.

4.4.1 Cálculo y prueba de la *tau* de Goodman-Kruskal

Para la prueba del estadístico *tau* de Goodman-Kruskal se usó la librería de R *DescTools* (Aho *et al.*, 2018) en la que se encuentra la función: *GoodmanKruskalTau()* la cual nos permite estimar *tau* ingresando el conjunto de categorías para cada variable. Asimismo se utilizó la librería *genetics* (Warnes, Gorjanc, Leisch, & Man, 2012) para estimar desequilibrio de ligado con los estadísticos clásicos, específicamente con la función *LD()*, para posteriormente comparar resultados entre estimadores.

Las pruebas fueron realizadas con 30 genotipos nucleares y mitocondriales (tomando los mitocondriales como homocigotos para la función *LD()*). El código puede ser consultado en github bajo el nombre de *test_Gk.r* en:

https://github.com/eduardo1992/tesis_EduardoGonzalez/blob/master/test_GK.r

4.5 Análisis por parejas

Para identificar desequilibrio mitonuclear se realizó un análisis por parejas entre SNP mitocondriales y nucleares. El *script GK_corr.R* fue diseñado para realizar el análisis por parejas utilizando la *tau* de Goodman-Kruskal sobre información genómica. Los *scripts* utilizan como entrada los archivos donde se reportan las variantes nucleotídicas (VCF) del núcleo y la mitocondria.

El análisis toma cada SNP mitocondrial y nuclear, para posteriormente estimar el desequilibrio entre cada pareja de SNP, aplicando la *tau* de Goodman Kruskal, obteniendo así un análisis de alta resolución de los genomas completos nuclear y mitocondrial (Figura 15).

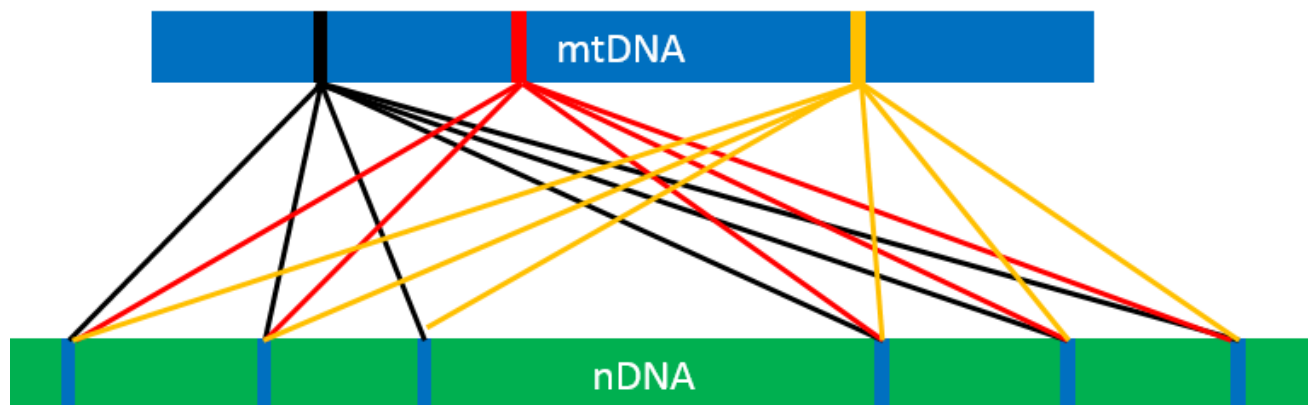


Figura 15. Esquema representado la comparación por parejas (líneas oblicuas de colores) entre los SNP (barras verticales de colores) nucleares (nDNA – verde) y mitocondriales (mtDNA – mitocondrial).

El análisis por parejas en datos genómicos fue realizado en la supercomputadora Abacus-I, parte del Laboratorio de Matemática Aplicada y Cómputo de Alto Rendimiento del Departamento de Matemática del Cinvestav, localizado en el estado de México.

4.6 Significancia estadística

Para establecer significancia estadística en los resultados obtenidos del análisis de desequilibrio mitonuclear por parejas, fueron definidos dos casos: asociaciones perfectas y asociaciones altas, como se describe a continuación.

4.6.1 Asociaciones perfectas

Los valores de asociaciones perfectas de τ ($\tau = 1$) se obtienen cuando las variables tienen una asociación perfecta entre sus categorías. En este caso, cuando el análisis arroja un valor de 1, los alelos mitocondriales se asocian de manera perfecta con genotipos nucleares en un grupo de individuos.

Mit	Nuc	
T	T/T	Ind 1
T	T/T	.
T	T/T	.
T	T/T	.
T	T/T	.
T	T/T	.
T	T/T	.
T	T/T	.
C	T/G	Ind 9

Figura 16. Ejemplo de covariación perfecta ($\tau = 1$) entre genotipos de un SNP nuclear (Nuc) y de un SNP mitocondrial (Mit) en nueve individuos (Ind).

La probabilidad de que un solo individuo i de un total de individuos n posea una variante por azar en el genoma nuclear, en donde dicha variante es diferente al resto de los individuos analizados puede aproximarse por $(\frac{1}{\text{Número de individuos } (n)})$. La probabilidad de que ese individuo sea también el único que posea una variante dada en su mtADN por azar sería $(\frac{1}{\text{Número de individuos } (n)}) \times (\frac{1}{\text{Número de individuos } (n)}) = (\frac{1}{\text{Número de individuos } (n)})^2$, ya que son eventos azarosos independientes.

Finalmente, en base a las probabilidades previamente descritas podemos definir el número esperado por azar de asociaciones perfectas con solo un individuo variando, como:

$$E = \frac{1}{N} * T \quad [9]$$

Donde:

E = Número de SNP esperados bajo desequilibrio debido a efectos azarosos.

N = Número de individuos en la población.

T = Número de casos de un individuo variando en la población.

En el caso de más de un individuo variando en la población, se debe considerar la categoría de genotipo de los SNP mitocondriales y nucleares para establecer la probabilidad de obtener asociaciones perfectas debido a los efectos aleatorios. Por lo que cada categoría de genotipo nuclear y mitocondrial deben estar asociadas, como se observa en la Figura 17.

Mit	Nuc	
T	T/T	Ind 1
T	T/T	.
T	T/T	.
C	T/G	.
T	T/T	.
C	T/G	.
T	T/T	.
T	T/T	.
C	T/G	Ind 9

Figura 17. Ejemplo de asociación perfecta entre genotipos mitocondriales (Mit) y nucleares (Nuc) con más de un individuo (Ind) variando.

Si sumamos la probabilidad de los genotipos nucleares (0/0, 1/1, 0/1) a la fórmula [9], podemos obtener la estimación del número de parejas de SNP mitonucleares en desequilibrio perfecto debido a efectos aleatorios en una población:

$$E = \left(\frac{1}{G * N} \right)^v * T \quad [10]$$

Donde:

E = Número de SNP esperados bajo desequilibrio debido a efectos azarosos.

N = Número de individuos en la población

v = Número de individuos variando

T = Número de casos de v individuos variando en toda la población

Como podemos observar en la fórmula [10], si aumenta el número de individuos que varían, la cantidad de correlaciones perfectas esperadas debidas a efectos aleatorios disminuye.

Para encontrar el número de individuos variando, se calcularon las frecuencias de genotipo para cada SNP con un *script* llamado *minimum_genotype_frequency.pl*, el cual toma como entrada los archivos VCF de la población de interés y genera un archivo MGF (frecuencia mínima de genotipo) que contiene la siguiente información por cada SNP: Frecuencia de genotipos (0/0, 0/1, 1/1), posiciones genómicas y el genotipo de frecuencia mínima reportado.

```
perl minimum_genotype_frequency.pl <file.vcf> > file.mgf
```

Finalmente, se utilizó el total de resultados esperados por efectos de azar de cada caso de asociación perfecta para establecer un corte de confianza, dependiendo del comportamiento de los datos (ver por ejemplo Figura 25 en Resultados).

4.6.2 Asociaciones altas

Para detectar valores altos y significativos de *tau*, se realizó una simulación. La cual consistió en generar variaciones aleatorias en genotipos nucleares y mitocondriales con un número seleccionado de individuos y sobre esta variación generada en ambos genomas estimar *tau* para obtener una distribución simulada.

Con la finalidad de realizar una simulación lo más cercana a la variación observada en las poblaciones a simular, se obtuvieron las probabilidades de observar genotipos con alelos alternos a la referencia utilizando los archivos VCF, en los cuales se muestra todos los SNP del genoma nuclear y mitocondrial. La frecuencia observada de cada genotipo con alelo alterno fue obtenida para cada población, es decir se contabilizó la frecuencia observada de SNP con genotipos heterocigotos (0|1) y homocigotos (1|1) del alelo alterno en cada población (el genotipo 0|0 solo contiene el alelo de la referencia). Posteriormente, dichas frecuencias de alelos alternos a la referencia observadas para cada genotipo se dividieron entre el total de SNP analizados en la población de interés, obteniendo así la probabilidad de cada genotipo dada una población, Estas probabilidades estimadas fueron utilizadas para simular la variación al azar de ambos genomas:

$$\text{Probabilidad de frec. genotípica de alelo alterno} = \frac{\text{N}^\circ \text{ observaciones frec. genotipo alelo alterno}}{\text{N}^\circ \text{ de polimorfismos totales}}$$

Número total de sitios polimórficos = 3,200,000

Alelos de la referencia: Nuclear(T)
Alelos alternos: Nuclear (C)

Mitocondrial (C)
Mitocondrial (G)

(Núcleo- Mitocondria)	N --- M	N --- M	N --- M
Individuo1-->	T/T---C	T/C---C	T/T---G
Individuo2-->	T/T---C	T/C---G	C/C---G
Individuo3-->	C/C---G	C/C---G	C/C---G
Individuo4-->	C/C---G	C/C---G	C/C---G
Individuo5-->	T/T---C	C/C---C	C/C---C
Individuo6-->	C/C---G	C/C---G	C/C---G
Individuo7-->	C/C---G	C/C---G	C/C---G
<i>Núcleo</i>	0.42		0.14
Frecuencia 0 0:	0	0	0
Frecuencia 0 1:	0.57	0.28	0.85
Frecuencia 1 1:	-----	0.71	-----
<i>Mitocondria</i>	0.42		0.14
Frecuencia 0:	0.57	0.28	0.85
Frecuencia 1:		0.71	
	Caso 1	Caso 2	Caso 3
	(200,000 Observaciones)	(1,000,000 Observaciones)	(2,000,000 Observaciones)
Probabilidad= $\frac{\text{Observaciones}}{\text{Total}}$	Probabilidad = 0.0625	Probabilidad= 0.3125	Probabilidad= 0.625

Figura 18. Ejemplo de determinación de probabilidad de frecuencias genotípicas, para tres casos de ejemplo de variación observada en 3,200,000 SNP.

En el ejemplo de la Figura 18, se observan tres casos de frecuencias genotípicas del genoma nuclear y mitocondrial en una población hipotética con 3 200 000 sitios polimórficos; en la parte de abajo se muestra las observaciones hipotéticas de cada caso de frecuencia en la población, las cuales al ser divididas entre el número de polimorfismos totales obtenemos la probabilidad de observar las frecuencias de genotipo de alelo alternativo de cada uno de los casos.

A manera de ejemplo tomemos el caso uno, el cual vemos que corresponde a una frecuencia de genotipos nucleares: 0|0 = 0.42, 0|1=0 y 1|1=0.57; y de alelos mitocondriales: 0=0.42 y 1=0.57. Como se mencionó anteriormente, se toma en cuenta la frecuencia de los genotipos del alelo alternativo, por lo que observamos que las frecuencias de genotipos de alelo alternativo 0|1 (0), 1|1 (0.57) y 1 (0.42) se observaron en nuestro caso hipotético uno en 200 000 SNP, por lo que al dividir este número de observaciones entre el total de SNP en la población (3,200,000 SNP), obtenemos la probabilidad de obtener la frecuencia de 0.57 de genotipo nuclear 1|1 y 0.42 para el genotipo mitocondrial 1 es de 200 000 en 3,200,000 lo cual da una probabilidad correspondiente a 0.0625. Se omite en el caso uno el genotipo nuclear 0|1 ya que su frecuencia fue de 0.

Cabe destacar que las probabilidades fueron estimadas para núcleo y mitocondria de manera independiente, pero con la finalidad de facilitar la explicación se toman las mismas observaciones hipotéticas para ambos genomas.

Las probabilidades obtenidas para cada caso de genotipos nucleares y mitocondriales se utilizaron para alimentar el *script Random_SimulatorV3.withProbab.R* que utiliza los archivos MGF (los cuales contienen las frecuencias genotípicas reales) de la población de interés, el código se puede consultar en github con la siguiente liga:

https://github.com/eduardo1992/tesis_EduardoGonzalez/blob/master/Random_SimulatorV3.withProbab.R

Las distribuciones de los valores *de tau* simulado y real se compararon con la prueba de Kolmogórov-Smirnov (K-S) para identificar desigualdad entre las distribuciones real y simulada. Esta prueba es parte del paquete R *statistics* (“R: A Language and Environment for Statistical Computing,” 2008) y la funciones utilizadas fueron *ks.test()* para la prueba K-S y *cdfr()* para graficar la distribución cumulativa de los valores reales y simulados. Finalmente, como punto de corte se tomaron los valores máximos obtenidos en las simulaciones (ver por ejemplo Figura 26 en Resultados).

4.7 Filtrado de resultados de caracterización

El proceso de caracterización de los resultados obtenidos por el análisis de parejas es el siguiente:

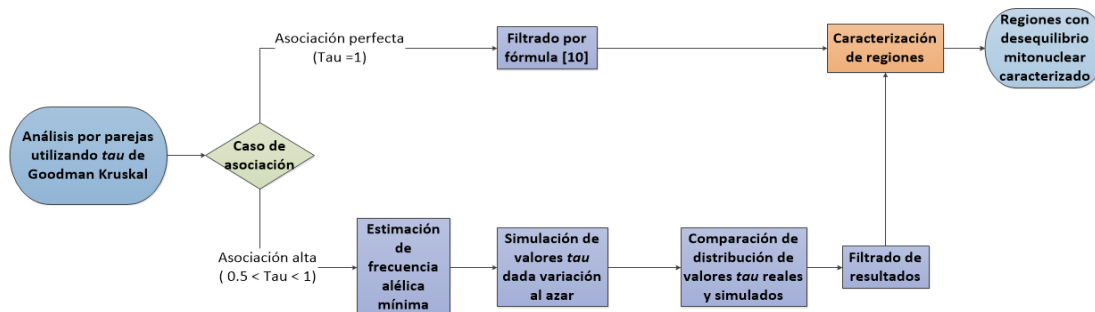


Figura 19 Diagrama de flujo de caracterización y filtrado de resultados de análisis de tau por parejas

Los resultados obtenidos del *script GK_corr.R* se filtraron con el valor de corte, basado en simulaciones o en la fórmula [10] dependiendo del caso de asociación (perfecta o alta). Los resultados se reportaron en el archivo .filM que consiste en un archivo de tres filas dividido por comas que contiene la posición nuclear, la posición mitocondrial y el valor de *tau*. Ejemplo:

CP.1.103,0.6,MT.20

CP siglas para cromosoma y posición indicados por los siguientes dos números, cromosoma 1 posición 103; 0.6 indica el valor *de tau* entre dicha posición nuclear con la siguiente posición reportada, que es la mitocondrial (MT) y que en este caso es la posición 20. Por lo que este ejemplo reporta que la posición 103 del cromosoma 1 tiene un valor de asociación de 0.6 con la posición 20 de la mitocondria.

Los valores de asociación fueron filtrados por la segunda columna del archivo de salida con la siguiente línea de comando:

```
gawk -f “,” ‘$2 > value’ file.filtM > file.filter.fitM
```

Para caracterizar las regiones candidatas bajo desequilibrio mitonuclear, el *output* fue convertido a formato BED (Brown, 2002), el cual consiste en un archivo de tres columnas divididas por marca de tabulador, indicando cromosoma, posición inicial y posición final. La conversión fue realizada con el script: *film_2_bed.pl*.

El archivo BED de los candidatos fue caracterizado con la herramienta de *bedtools* (Quinlan & Hall, 2010) *intersect* para identificar regiones correspondientes entre dos archivos BED; en este caso, entre los candidatos y la base de datos en formato BED, la cual se obtuvo de ENSEMBL (Zerbino *et al.*, 2018). Las diferentes categorías de la base de datos fueron descargadas utilizando la herramienta en línea *biomart* (Smedley *et al.*, 2015), cuyas coordenadas se encuentran bajo el genoma de referencia Hg19 (GRCH37 p.13). Esta base de datos contiene las siguientes categorías:

- Genes (Ensembl Genes 91)
 - Exons
- Elementos regulatorios (Ensembl Regulation 92)
 - Motivos de unión
 - Otros elementos regulatorios
 - Sitio de inicio de transcripción
 - *Enhancers*
 - Características reguladoras
 - Sitio de unión a CTCF
 - Sitio de cromatina abierta
 - Promotor
 - Región adyacente a promotor

4.8 Análisis de enriquecimiento de términos GO

El análisis de enriquecimiento de términos de ontología de genes (GO, por sus siglas en inglés) se realizó utilizando el software ShinyGO en modalidad online (Ge & Jung, 2018), utilizando todas las bases de ontología disponibles (proceso biológico, componente celular y función molecular). Con una tasa de falsos descubrimientos (FDR, por sus siglas en inglés) de 0.05.

El link al *web server* de la herramienta de ShinyGO es el siguiente:

<http://bioinformatics.sdstate.edu/go/>

5 Resultados

5.1 Selección del método estadístico óptimo para detectar desequilibrio mitonuclear

Para evaluar el poder estadístico de la τ de Goodman-Kruskal para detectar la asociación no aleatoria entre alelos, se ha comparado con los estadísticos clásicos de LD, D' y r^2 , utilizados por el estudio de Sloan *et al.* (2015).

La siguiente tabla muestra la frecuencia de cada categoría en el contexto de una prueba de desequilibrio entre un alelo mitocondrial (Y) y un alelo nuclear (X), cabe mencionar que para los estadísticos de D' y r^2 se tomaron como homocigotos los genotipos mitocondriales.

Tabla 3. Tabla de contingencia entre genotipos nucleares (X) y mitocondriales (Y).

		X_1	X_2	X_3		
		0 0	0 1	1 1		
Y_1	0	a_{11}	a_{12}	a_{13}	$a_{.1}$	
Y_2	1	a_{21}	a_{22}	a_{23}	$a_{.2}$	
		$a_{.1}$	$a_{.2}$	$a_{.3}$		

1) Asociación perfecta

La siguiente tabla muestra los resultados de los distintos estadísticos en asociaciones perfectas:

Tabla 4. Tabla de contingencia de genotipos nucleares y mitocondriales con asociación perfecta (arriba). Resultados de estadísticos clásicos de LD y tau (abajo).

		Genotipos nucleares		
		0 0	0 1	1 1
Genotipos mitocondriales	0	17	0	0
	1	0	3	0

Estadístico	Valor
r^2	0.458764
D'	0.9990836
τ	1

2) Asociación alta

En el siguiente caso, el alelo 0 mitocondrial posee una asociación con el genotipo nuclear 0|0. Pero el alelo mitocondrial 1 se asocia con 0|1 y 0|0. Como el alelo 0 se encuentra en mayor frecuencia, la asociación es alta:

Tabla 5. Tabla de contingencia de genotipos nucleares y mitocondriales con asociación alta (arriba). Resultados de estadísticos clásicos de LD y tau (abajo).

		Genotipos nucleares		
		0 0	0 1	1 1
Genotipos mitocondriales	0	13	0	0
	1	3	4	0

Estadístico	Valor
r^2	0.20508
D'	0.9990836
τ	0.646428

3) Asociación media

La mitad de los alelos mitocondriales se asocian con genotipos nucleares:

Tabla 6. Tabla de contingencia de genotipos nucleares y mitocondriales con correlación media (arriba). Resultados de estadísticos clásicos de LD y tau (abajo).

		Genotipos nucleares		
		0 0	0 1	1 1
Genotipos mitocondriales	0	11	0	0
	1	3	4	3

Estadístico	Valor
r^2	0.3052783
D'	0.9995462
τ	0.5238095

Como podemos observar en las Tablas 4, 5 y 6, r^2 y D' no lograron distinguir entre diferentes niveles de correlación entre genotipos mitocondriales y nucleares, mientras que τ sí logró identificar cada uno de los casos de asociación de los ejemplos anteriores, los cuales son una representación de los escenarios esperados más representativos que se observaron en el análisis de detección de desequilibrio mitonuclear. Los estadísticos de desequilibrio de ligado clásicos se centran en detectar una asociación entre alelos en genotipos de una misma ploidía. En los ejemplos anteriores r^2 y D' no pudieron detectar las correlaciones directas entre los genotipos haploides y diploides.

En base a lo descrito, se concluye que la τ de Goodman-Kruskal es robusto a la ploidía en el análisis y solo correlaciona la covariación en categorías. Al tomar cada genotipo como una variable nominal, τ logra identificar las correlaciones entre genotipos nucleares y mitocondriales con gran sensibilidad.

5.2 Filtrado de la base de datos de 1000 genomas

5.2.1 Filtrado por pedigrí

Uno de los posibles sesgos al detectar desequilibrio mitonuclear utilizando la base de datos de 1000 genomas, son los individuos relacionados, ya que dos pares de SNP mitocondriales y nucleares podrían detectarse como asociados en la población debido a una relación genética más estrecha entre individuos.

Seleccionamos 2,490 muestras no relacionadas, de 19 poblaciones, cuyos individuos poseían secuencias de genoma mitocondrial y nuclear disponible.

Tabla 7. Número de individuos filtrados no relacionados en poblaciones de la base de datos 1000 genomas.

Población	Número de individuos	Superpoblación
Chinos Han en Beijing, China (CHB)	103	Asia del Este (EAS)
Japoneses en Tokyo, Japón (JPT)	104	Asia del Este (EAS)
Chinos Han del sur (CHS)	105	Asia del Este (EAS)
Chinos Dai en Xishuangbanna, China (CDX)	92	Asia del Este (EAS)
Ciudad Kinh in Ho Chi Minh, Vietnam (KHV)	99	Asia del Este (EAS)

Residentes de Utah con ancestría del Norte y Oeste de Europa (CEU)	95	Europeos (EUR)
Toscanos en Italia (TSI)	107	Europeos (EUR)
Finlandeses en Finlandia (FIN)	99	Europeos (EUR)
Británicos en Inglaterra y Escocia (GBR)	90	Europeos (EUR)
Población Ibérica en España (IBS)	107	Europeos (EUR)
Yorubas en Ibadan, Nigeria (YRI)	106	Africanos (AFR)
Luhya en Webuye, Kenya (LWK)	97	Africanos (AFR)
Gambianos en divisiones del oeste de Gambia (GWD)	113	Africanos (AFR)
Mende en Sierra Leona (MSL)	84	Africanos (AFR)
Esan en Nigeria (ESN)	99	Africanos (AFR)
Americanos con ancestría africana en E.U.A (ASW)	61	Africanos (AFR)
Africanos del caribe en Barbados (ACB)	96	Africanos (AFR)
Residentes de los Ángeles, con ancestría mexicana (MXL)	64	Americanos mestizos (AMR)
Puertorriqueños de Puerto Rico (PUR)	104	Americanos mestizos (AMR)
Colombianos de Medellín, Colombia (CLM)	94	Americanos mestizos (AMR)
Peruanos de Lima, Perú (PEL)	85	Americanos mestizos (AMR)
Indios Gujarati de Houston, Texas (GIH)	101	Asiáticos del sur (SAS)
Punjabi de Lahore, Pakistán (PJL)	96	Asiáticos del sur (SAS)
Bengalíes de Bangladesh (BEB)	86	Asiáticos del sur (SAS)
Tamil Sri Lankan de Reino Unido (STU)	101	Asiáticos del sur (SAS)
Indios Telugu de Reino Unido (ITU)	102	Asiáticos del sur (SAS)

5.2.2 Filtrado de los SNP

En el presente trabajo se realizaron los análisis de *tau* por parejas para detectar desequilibrio mitonuclear, agrupando a los individuos del panel de dos maneras: La primera fue seleccionando poblaciones individuales, dos de cada superpoblación, y el segundo fue agrupando a los individuos por superpoblaciones. El primer análisis se realizó en 10 poblaciones, descritas en la tabla 8. El segundo fue un análisis utilizando a todos los individuos de cada una de las 5 superpoblaciones del panel: asiáticos del este, sudasiáticos, europeos, africanos y americanos.

Asimismo, se realizó un análisis preliminar con las poblaciones japonesa y china, en el que se establecieron los parámetros en base al comportamiento de los datos, dado que nunca se había utilizado la *tau* de Goodman-Kruskal para análisis de desequilibrio.

Filtrado de los SNP en análisis de poblaciones individuales y análisis preliminar

La siguiente tabla describe las poblaciones y los SNP filtrados, cada color representa la superpoblación a la que pertenece cada población. En el caso de las poblaciones japonesa y china, que se utilizaron para el análisis preliminar, se seleccionó el parámetro correspondiente para escoger los SNP con alelos diferentes a la referencia con una frecuencia mínima de 1 (*--min-ac 1*). En el resto de las 10 poblaciones utilizadas para el análisis de poblaciones individuales, se utilizó el parámetro de una frecuencia mínima de 3, en alelos diferentes a la referencia, dados los efectos aleatorios. Ver sección de resultados 6.2.3 "Detección de desequilibrio mitonuclear en dos poblaciones usando el análisis por pares de *tau* de Goodman-Kruskal (un análisis preliminar)".

Tabla 8. Número total de SNP seleccionados y número total de individuos seleccionados por superpoblación.

Población	Número de SNP nucleares seleccionados	Número de SNP mitocondriales seleccionados	Número de individuos	Superpoblación
Chinos Han en Beijing, China (CHB)	12,122,281	557	103	Asia del Este (EAS)
Japoneses en Tokyo, Japón (JPT)	11,273,533	457	104	Asia del Este (EAS)
Residentes de Utah con ascendencia del Norte y Oeste de Europa (CEU)	10,845,130	382	99	Europeos (EUR)

Finlandeses en Finlandia (FIN)	8,486,189	157	99	Europeos (EUR)
Yorubas en Ibadan, Nigeria (YRI)	18,433,659	528	99	Africanos (AFR)
Gambianos en divisiones del oeste de Gambia (GWD)	18,516,679	194	113	Africanos (AFR)
Residentes de los Ángeles, con ancestría mexicana (MXL)	13,123,114	280	64	Americanos mestizos (AMR)
Puertorriqueños de Puerto Rico (PUR)	10,564,728	161	104	Americanos mestizos (AMR)
Punjabi de Lahore, Pakistan (PJI)	8,805,425	124	96	Asiáticos del sur (SAS)
Hablantes de Bengali de Bangladesh (BEB)	8,486,041	80	86	Asiáticos del sur (SAS)

Se analizaron una media de 12,965,678 SNP nucleares y una media de 97 SNP mitocondriales. Las poblaciones africanas son las que mayor variación presentan en el panel.

Filtrado de los SNP en análisis de superpoblaciones

Para el análisis de superpoblaciones, se consideraron todos los individuos del panel completo agrupados por superpoblaciones. El parámetro utilizado implicó una frecuencia mínima de 3 en alelos diferentes a la referencia (*--min-ac 3*).

Tabla 9. Número de SNP seleccionados y número total de individuos seleccionados por superpoblación.

Superpoblación	Número de SNP nucleares seleccionados	Número de individuos	Número de SNP mitocondriales seleccionados
Asia del este (EAS)	10,716,335	503	457
Europeos (EUR)	11,772,802	498	291
Africanos (AFR)	22,235,077	550	616
Americanos mestizos (AMR)	13,681,156	347	309
Asiáticos del sur (SAS)	12,806,213	486	483

5.2.3 Detección de desequilibrio mitonuclear en dos poblaciones usando el análisis por pares de *tau* de Goodman-Kruskal (un análisis preliminar)

Debido a la naturaleza masiva del análisis por pares entre los SNP mitocondriales y nucleares, se realizó una prueba preliminar utilizando el *script Gk_corr.R* en las poblaciones japonesa y china, con el objetivo de observar el comportamiento del estadístico al detectar asociaciones entre los SNP, para así definir los parámetros de los siguientes análisis en base a su significancia estadística.

La siguiente tabla muestra el número de combinaciones realizadas, las cuales se estiman al multiplicar el número de los SNP nucleares por el número de los SNP mitocondriales.

Tabla 10. Número total de parejas de SNP nucleares y mitocondriales analizadas en las poblaciones japonesa y china.

Población	Número total de parejas de SNP nucleares y mitocondriales
Chinos Han en Beijing, China (CHB)	6,752,110,517
Japoneses en Tokyo, Japón (JPT)	5,152,004,581

5.2.3.1 Resultado de los análisis por parejas de las poblaciones japonesa y china

Un bajo porcentaje de comparaciones totales tuvieron un *tau* mayor a 0.1, lo que indica que una gran mayoría de los resultados estaban por debajo de 0.1, mostrando señales bajas de desequilibrio, como se observa en la Tabla 11. En promedio, un 0.23% de las comparaciones totales fueron mayores a 0.1.

Tabla 11. Número total y porcentaje de resultados con tau mayor a 0.1

Población	Número total de parejas de SNP nucleares y mitocondriales	Número y porcentaje sobre el total de combinaciones con <i>tau</i> mayor a 0.1
Chinos Han en Beijing, China (CHB)	6,752,110,517	83,882,832 (1.24%)
Japoneses en Tokyo, Japón (JPT)	5,152,004,581	63,195,214 (1.22%)

La distribución de los valores de τ totales por encima de 0.1 se muestra en la Tabla 12, así como su representación visual en una gráfica de cajas y bigotes (Figura 20). Los datos graficados representan el 1.24% de los valores totales para chinos y 1.22% para japoneses. En la Tabla 12 se observa que existe un mayor número de asociaciones perfectas en la población china que en la japonesa, lo cual esta ligado al efecto del azar dado el número de alelos raros, lo cual se discutirá más adelante.

Tabla 12. Número de resultados obtenidos por cada rango de valores de tau.

Población	Valores de τ De 0.1-0.5	Valores de τ de 0.5-0.7	Valores de τ de 0.7-0.9	Valores de τ =1
Chinos Han en Beijing, China (CHB)	70,801,299	31,711	666	13,049,156
Japoneses en Tokyo, Japón (JPT)	56,735,868	33,867	612	6,425,479

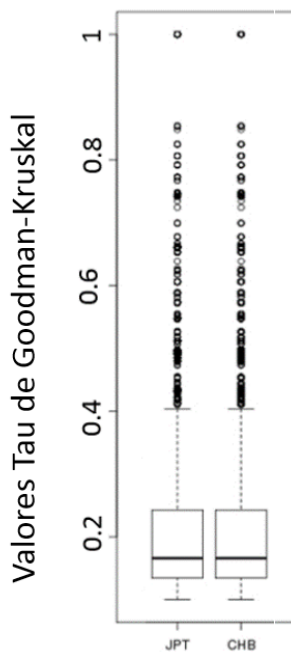


Figura 20. Gráfica de cajas y bigotes representando la distribución de valores de tau mayores a 0.1.

5.2.3.1.1 Asociaciones perfectas

Los valores de correlación perfectos obtenidos fueron el 15% de los resultados con valores de τ mayores a 0.1, siendo un número mayor al esperado, ya que los valores por debajo de 1 fueron menores, lo que indica una naturaleza discreta de la distribución de los valores de τ .

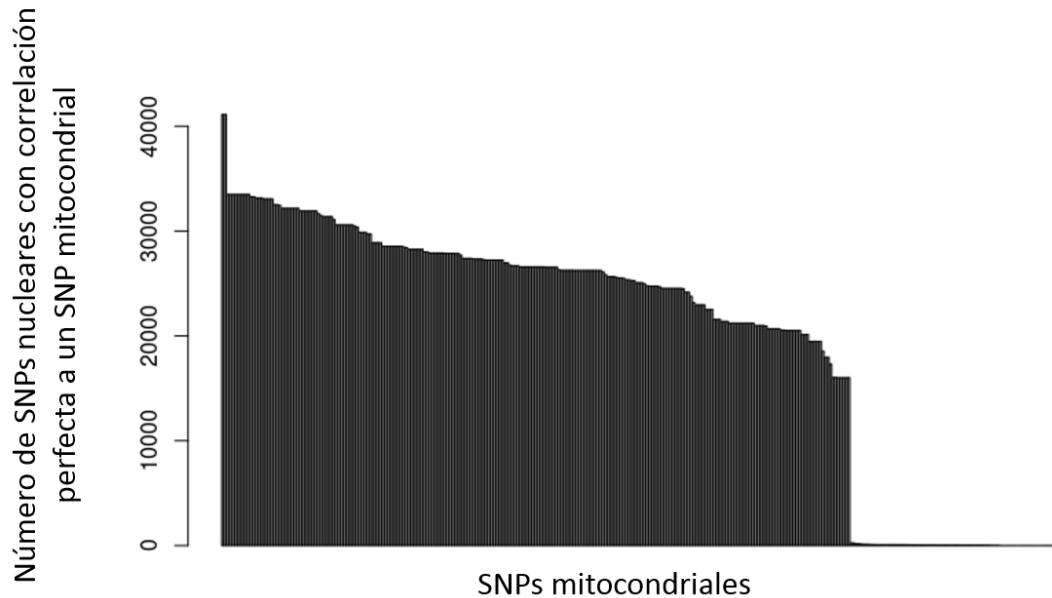


Figura 21. Gráfica de barras en donde cada una de ellas indica el número de SNP nucleares asociados de manera perfecta a un SNP mitocondrial en población japonesa.

En la Figura 21 observamos una gráfica de barras indicando el número de correlaciones perfectas de los SNP nucleares por cada SNP mitocondrial en la población japonesa. Puede observarse un gran bloque de los SNP mitocondriales con un alto número de correlaciones perfectas, en un rango de 18 000 a más de 40, 000.

Este comportamiento en los datos nos hizo cuestionarnos cuál sería la magnitud de los efectos aleatorios sobre el número de correlaciones perfectas identificadas en cada SNP mitocondrial. Al aplicar la fórmula 10, la cual estima el total de correlaciones perfectas debidas al azar, se observó una correlación directa entre el número de individuos que varían para determinada posición y el número esperado de valores obtenidos debido a los efectos aleatorios.

Como resultado de la aplicación de la fórmula 10 sobre el número de correlaciones perfectas se logró seccionar la Figura 21 en tres partes, cada una de estas representando el número de individuos variando en el genoma nuclear y

mitocondrial; y las asociaciones perfectas esperadas debido a efectos aleatorios (Figura 22).

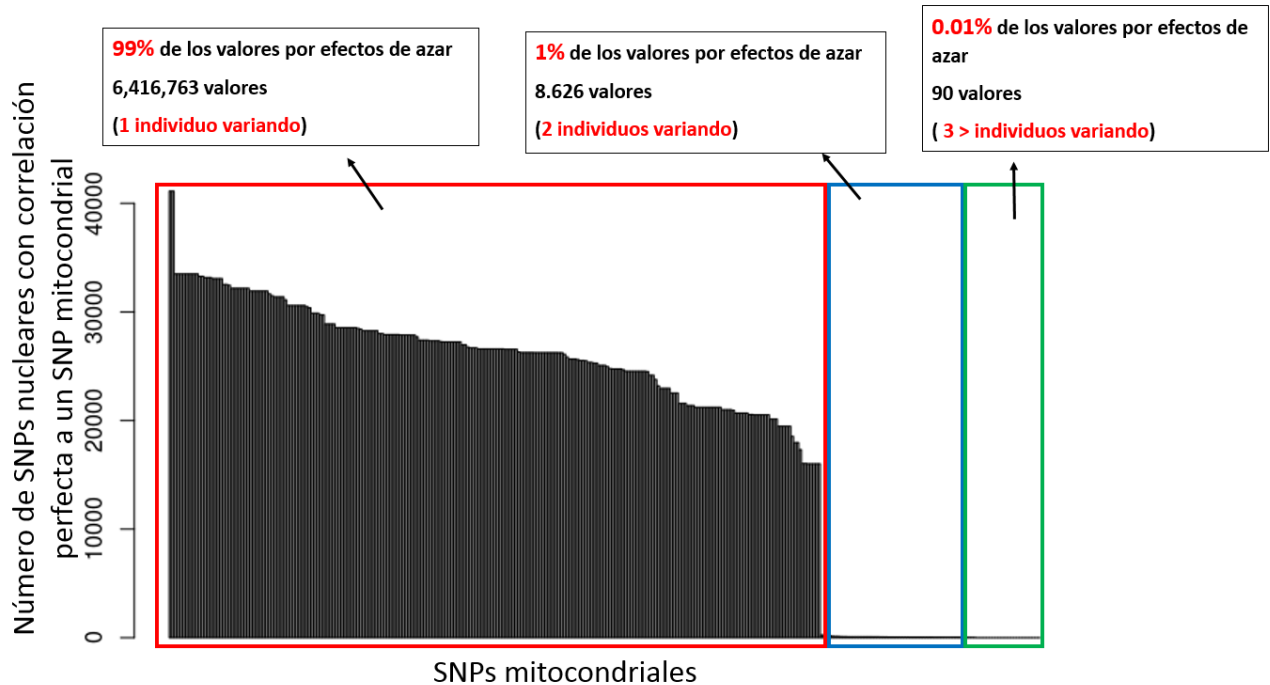


Figura 22. Gráfica de barras que indica el número de SNP mitocondriales con correlaciones perfectas con SNP nucleares en población japonesa, seccionados según el número de individuos que varían en la posición para ese SNP. Los porcentajes indican el porcentaje esperado sobre el total de valores de asociaciones perfectas dado efectos de azar.

Como se observa en la Figura 22, los SNP con $\tau = 1$ y con solo un individuo variando tienen una alta probabilidad de ser un resultado de aleatoriedad. En el bloque con marco rojo, el 90% de los valores no serían estadísticamente significativos. En el marco azul, observamos que los valores disminuyen a 8000, lo que indica que el 1% de los valores se deben a efectos aleatorios. En el marco verde vemos nuevamente una disminución en los valores a 90, indicando que se espera menos del 0.01 % de los 90 valores debido a los efectos aleatorios.

Al aplicar la fórmula [10] determinamos que los resultados confiables se obtuvieron a partir de SNP con tres o más individuos variando en la población. El mismo análisis se repitió en la población china, que también estuvo sujeta a los mismos parámetros de filtrado.

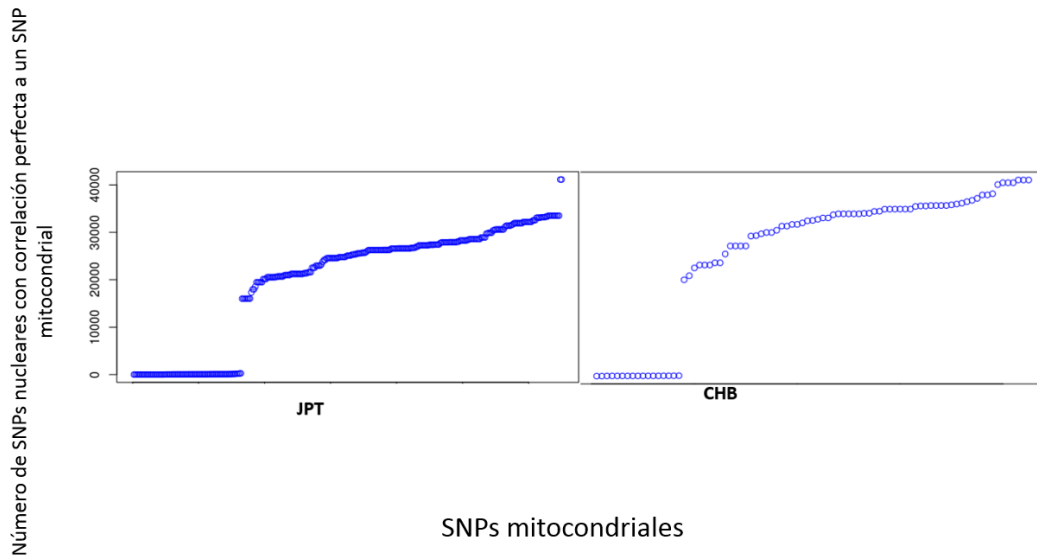


Figura 23. Gráfica de puntos indicando el número de SNP nucleares con asociación perfecta a un SNP mitocondrial. JPT, población japonesa; CHB, población china.

El comportamiento de las correlaciones perfectas en ambas poblaciones fue similar. Agregando el número de individuos a estas gráficas podemos observar la correlación directa entre el número de correlaciones perfectas y el número de individuos que varían (Figura 24).

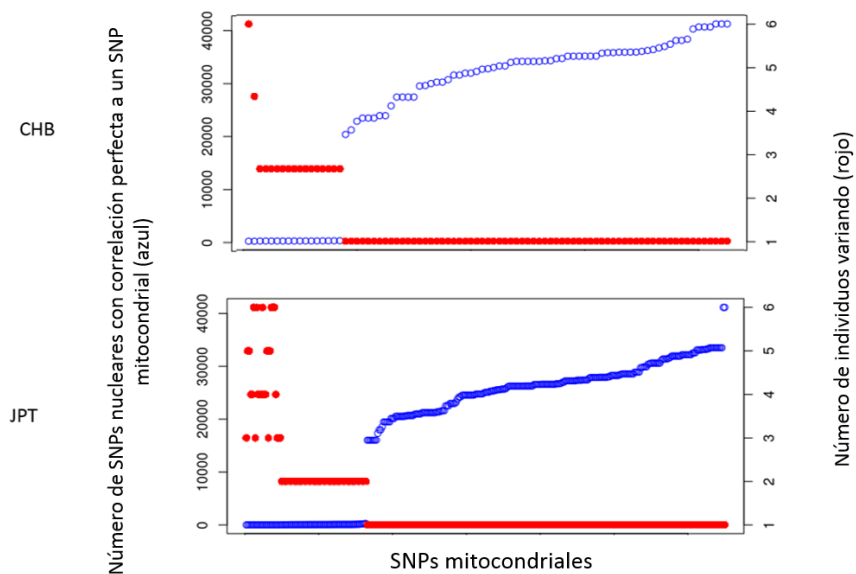


Figura 24 . Gráfica de puntos indicando el número de SNP nucleares con correlación perfecta a un SNP mitocondrial (azul, eje x primario) y número de individuos variando (rojo, eje x secundario). JPT, población japonesa; CHB, población china.

En ambos gráficos podemos apreciar la relación inversamente proporcional entre el número de individuos y el número de correlaciones perfectas identificadas. Concluyendo que para obtener resultados significativos con correlaciones perfectas es necesario ignorar aquellos SNP con tan solo un individuo variando, por lo que para los análisis siguientes de poblaciones individuales y superpoblaciones se decidió utilizar el parámetro de selección de SNP con mínimo de frecuencia de 3 alelos diferentes a la referencia (`--min-ac 3`).

5.2.3.1.2 Correlaciones altas

Las simulaciones preliminares se realizaron en japoneses y chinos para establecer qué posible punto de corte para valores altos y significativos de *tau* se podría obtener.

Para esto, la distribución de los valores de *tau* obtenidos a partir de variación aleatoria (simulación) se comparó con la distribución real de *tau* para cada población, encontrando que en ambos casos las distribuciones teórica y real son diferentes entre sí. Finalmente, el valor máximo de *tau* obtenido en las simulaciones aleatorias, que siempre fue menor al valor máximo real, se seleccionó como el valor de corte de fiabilidad para evitar casos que hayan podido surgir aleatoriamente.

Los siguientes gráficos muestran la distribución de densidad acumulada de los datos reales y simulados para los valores de *tau* en las poblaciones japonesa y china.

Población japonesa

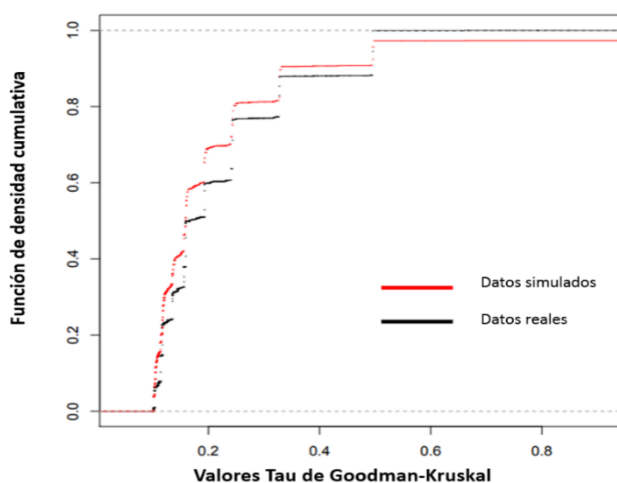


Figura 25. Distribución de densidad acumulada de valores de tau reales y simulados en la población japonesa. El valor máximo de tau obtenido de los datos simulados fue de 0.79

Población china

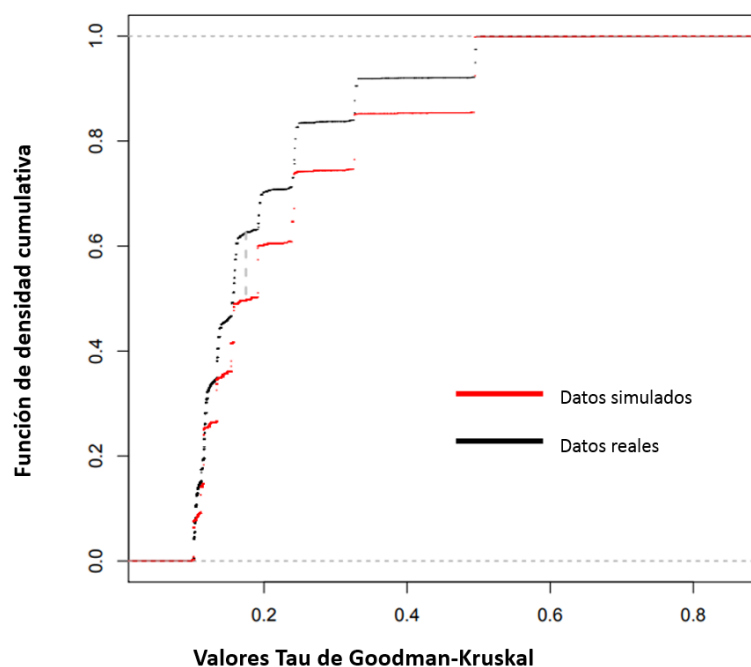


Figura 26. Distribución de densidad acumulativa de valores de tau reales y simulados en la población china. El valor máximo de tau obtenido de los datos simulados fue de 0.79

Se aplicó una prueba de K-S para probar si ambas distribuciones eran estadísticamente diferentes. El valor máximo obtenido por la simulación se seleccionó como el valor de corte para el filtrado.

Tabla 13. Resultado de la prueba de K-S y máximo valor de tau simulado en población japonesa y china.

Población	Resultado test de K-S (D)	Valor de P	Valor máximo de <i>tau</i>
Chinos Han en Beijing, China (CHB)	D=0.11	$0.109 \cdot 10^{-16}$	0.79
Japoneses en Tokyo, Japón (JPT)	D =0.12	$0.109 \cdot 10^{-16}$	0.79

Al aplicar los filtros de correlación perfecta y de correlación alta, anteriormente descritos, obtuvimos en total 111 pares de SNP nucleares y mitocondriales en desequilibrio en población japonesa y 51 en población china (Tabla 14).

Tabla 14. Valores de tau significativos en las poblaciones japonesa y china.

Población	$\tau = 1$	$0.79 < \tau < 1$
Chinos Han en Beijing, China (CHB)	37	14
Japoneses en Tokyo, Japón (JPT)	90	21

Pudimos detectar el desequilibrio mitonuclear en dos poblaciones, lo que ayudó a establecer los parámetros para el análisis de poblaciones individuales y de superpoblaciones. Concluimos que el número mínimo de individuos variando para obtener correlaciones perfectas significativas es de 3; asimismo, los valores de corte de confianza cuando $\tau < 1$, se dan en un valor de 0.79.

5.3 Detectando desequilibrio mitonuclear en poblaciones individuales

Los parámetros para detectar correlaciones estadísticamente significativas se establecieron en el análisis preliminar, por lo que se prosiguió a realizar el análisis en poblaciones individuales con el objetivo de detectar el desequilibrio mitonuclear específico de cada población y relacionarlo con un posible fenómeno de adaptación local.

Se analizaron dos poblaciones de cada superpoblación y se anexaron los resultados significativos obtenidos en el análisis preliminar de japoneses y chinos. En el análisis de poblaciones individuales se realizaron $1 * 10^9$ de combinaciones entre parejas de SNP nucleares y mitocondriales como se observa en la Tabla suplementaria 1

5.3.1 Resultados de análisis de desequilibrio mitonuclear en poblaciones individuales

En el análisis preliminar, los valores de τ mayores a 0.1 representaron menos del 2% del total de comparaciones realizadas. En el caso de todas estas poblaciones, se obtuvieron porcentajes similares, con una media de 0.94% para valores de τ mayores a 0.1. como se observa en la Tabla Suplementaria 2

Las siguientes gráficas de cajas y bigotes muestran la distribución de los valores de τ mayores a 0.1 en las poblaciones analizadas, incluidos japoneses y chinos.

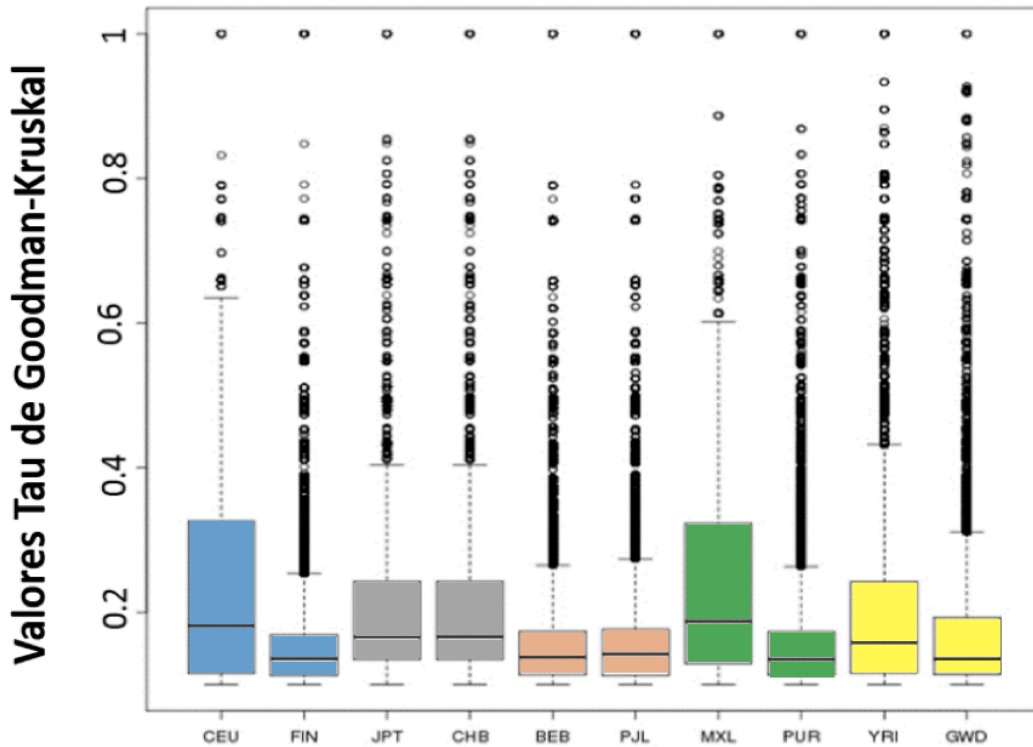


Figura 27. Gráficas de cajas y bigotes representando la distribución de valores de tau mayores a 0.1 en cada población. CEU (europeos), FIN (Finlandeses), JPT (Japoneses), CHB (Chinos), BEB (Hablantes de Bengali), PJJ (Punjabi), MXL (Mexicanos), PUR (Puertorriqueños), YRI (Yorubas) y GWD (Gambianos)

En la Figura 27 se observa que las distribuciones de valores de *tau* en todas las poblaciones son similares, ya que la mediana en todas las poblaciones se ubica en valores de entre 0.1 y 0.2; adicionalmente, todas las poblaciones mostraron valores altos ($tau > 0.6$) y valores de correlación perfectos ($tau = 1$). En la Tabla Suplementaria 3 se muestra el número de valores de *tau* por cada rango en donde se confirma lo graficado en la Figura 27, donde se logran detectar correlaciones perfectas y altas.

5.3.1.1 Asociaciones perfectas

Como se describe en la fórmula [10], con más individuos variando en cada SNP, los efectos de la aleatoriedad disminuyen. Con base en el análisis preliminar se decidió utilizar aquellos SNP con al menos tres individuos variando y aplicando la fórmula [10] se estimó el número esperado de parejas de SNP nucleares y mitocondriales en desequilibrio perfecto debido a la aleatoriedad con tres individuos variando.

Como se observa en la Tabla suplementaria 4, se esperan menos de una pareja de SNP nucleares y mitocondriales por azar utilizando los SNP con tres individuos variando, permitiendo detectar una asociación significativa y perfecta entre los SNP mitocondriales y nucleares libres de efectos aleatorios.

Finalmente obtenemos el número de SNP nucleares y mitocondriales con asociación perfecta en SNP con más de tres individuos variando no atribuibles al azar, en donde las poblaciones africanas y surasiáticas presentaron el mayor número de SNP asociados perfectos., ver Tabla suplementaria 5

5.3.1.2 Asociaciones altas

Debido a la similitud entre los resultados de la simulación en población japonesa y china, no se consideró necesario hacer simulaciones para todas las poblaciones individuales sino únicamente para las siguientes poblaciones representativas de cada bloque de superpoblación: europeos, yorubas, mexicanos y punjabi.

Se generaron un total de 50 millones de SNP con variación simulada para construir la distribución de valores de *tau* simulados.

Los análisis para cada población y la comparación con la distribución real contra la simulada dieron como resultado p-valores significativos con valor $2.2 \cdot 10^{-16}$ y valores máximos de 0.77 a 0.79, ver Tabla suplementaria 6

Europeos (CEU)

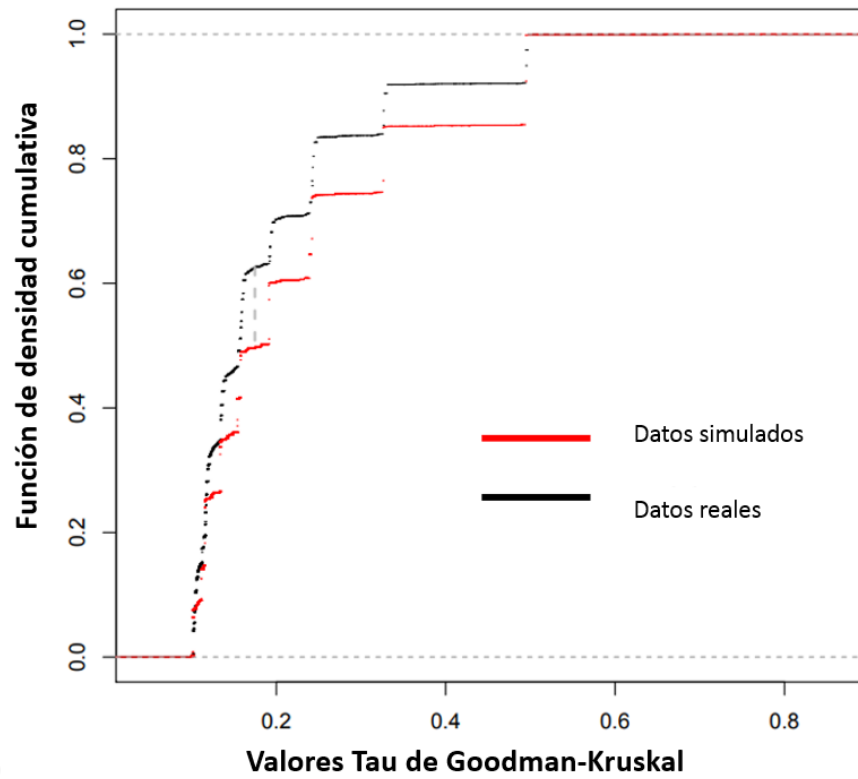


Figura 28. Distribución de densidad acumulativa de datos reales y simulados de valores de tau en población europea (CEU).

Yorubas (YRI)

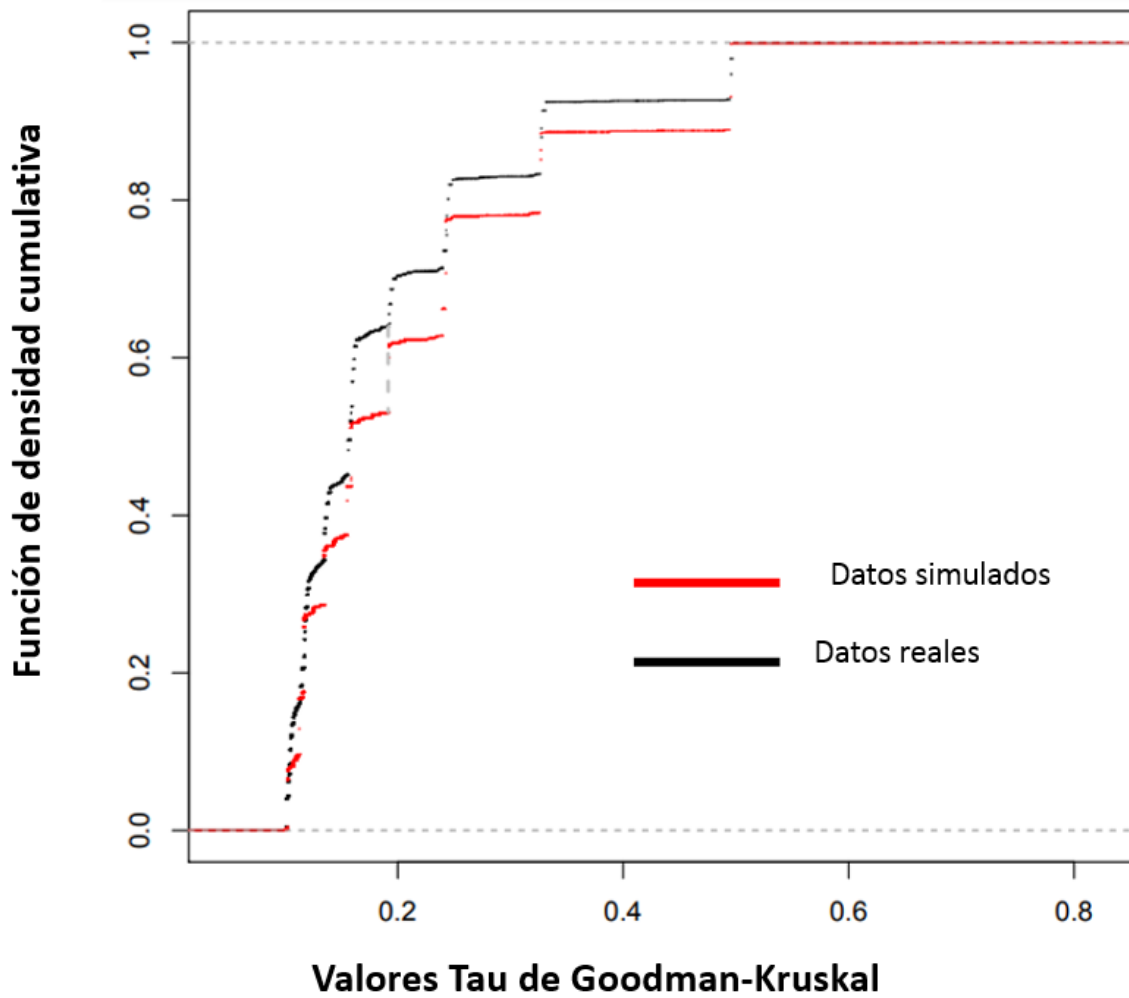


Figura 29. Distribución de densidad acumulativa de datos reales y simulados de valores de tau en población yoruba (YRI).

Mexicanos (MXL)

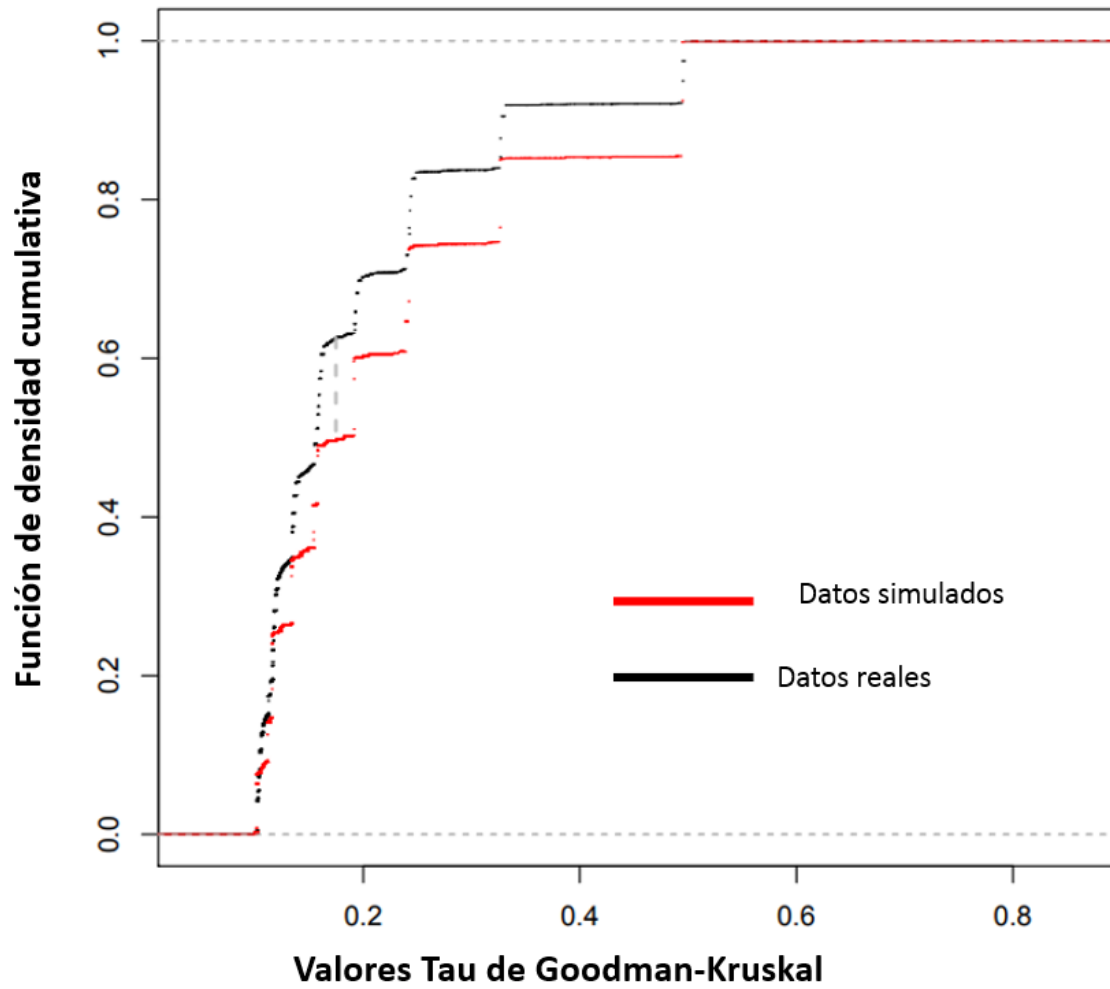


Figura 30. Distribución de densidad acumulativa de datos reales y simulados de valores de tau en población mexicana (MXL).

Punjabis (PJL)

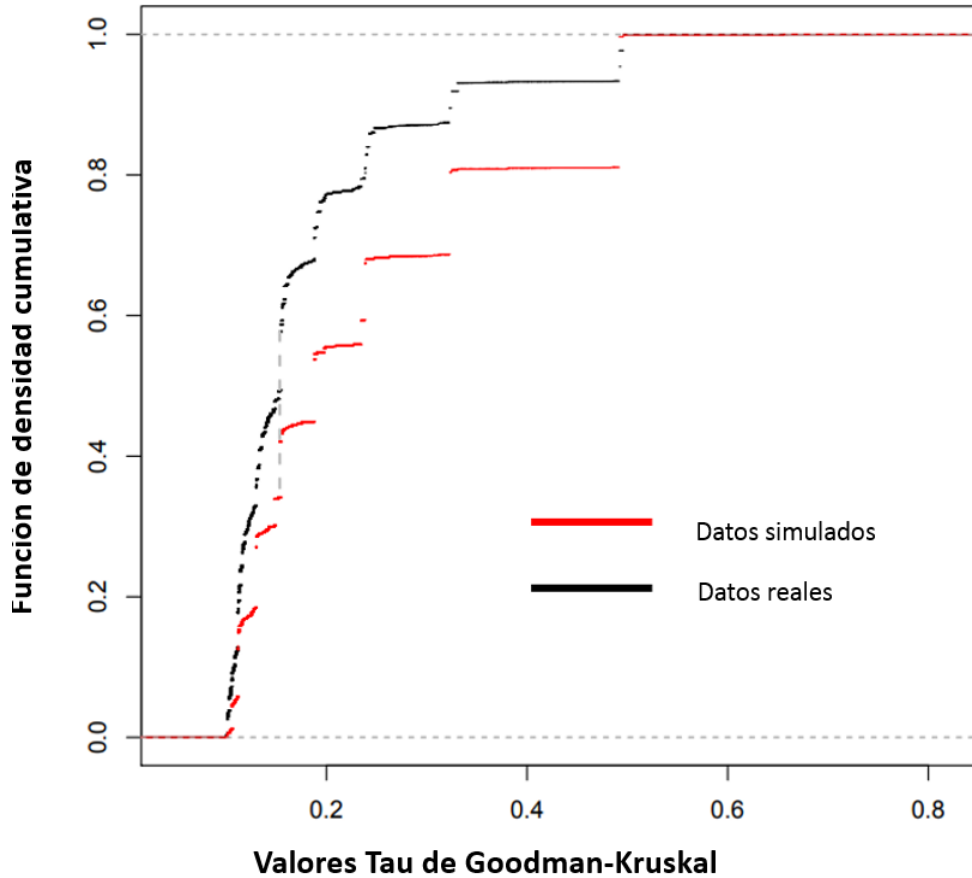


Figura 31. Distribución de densidad acumulativa de datos reales y simulados de valores de tau en población punjabi (PJL).

Los valores de corte utilizados fueron los valores máximos obtenidos por simulación de cada población. Las pruebas de K-S fueron significativas en todas las poblaciones, lo que significa que la distribución real es diferente de la esperada bajo procesos aleatorios.

5.3.1.2 2 Resultados significativos y caracterización

Aplicando los parámetros para establecer significancia estadística (simulaciones y fórmula [10]) a los resultados de *tau*, se obtuvieron alrededor de 1,261 parejas de SNP con desequilibrio mitonuclear en poblaciones individuales, donde las poblaciones de África obtuvieron el mayor número de parejas con un rango de 104 a 341 SNP, seguido por las poblaciones de Asia del sur con 130 a 267 SNP, las poblaciones de Asia del este con 51 a 111 SNP, las poblaciones Americanas con 22 a 113 y finalmente la población de Europa con el menor número con un rango de 1 a 73 SNP.

5.3.2 Localización genómica del desequilibrio mitonuclear

Se estimó la cantidad de SNP bajo desequilibrio mitonuclear a lo largo del genoma nuclear y mitocondrial, en el caso del genoma nuclear la estimación fue realizada en ventanas de 24 Gigabases y mitocondrial en ventanas de 500 bases

En la Figura 32 podemos observar la cantidad de SNP nucleares bajo desequilibrio mitonuclear en ventanas de 24 Giga bases

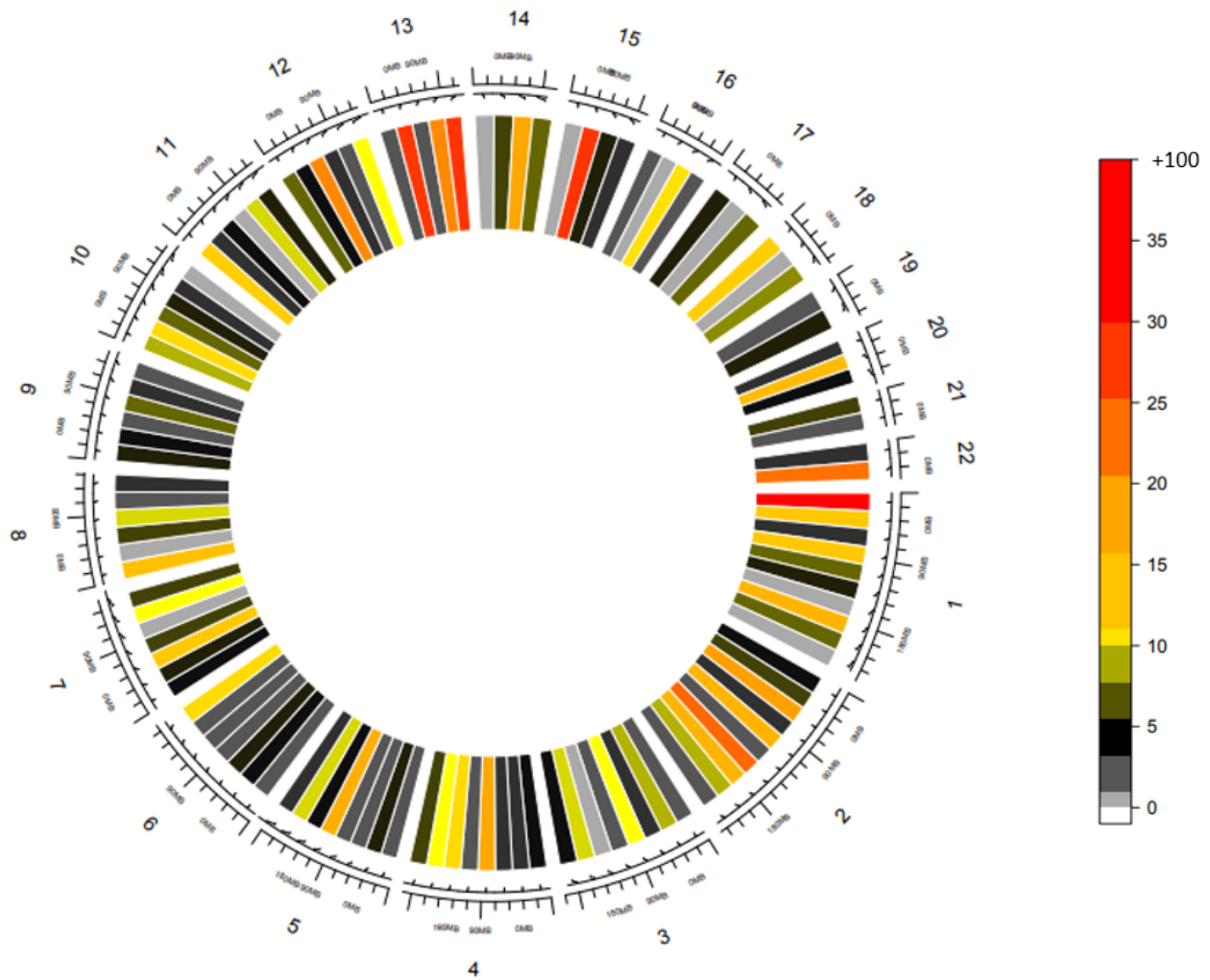


Figura 32. Número de SNP nucleares con disequilibrio mitonuclear significativo, en regiones de 24 gigabases, del análisis con poblaciones individuales. Los cromosomas con mayores número de SNP en disequilibrio son el 1, 2, 12-15 y 22.

Se aprecia que en la mayoría de las regiones (de 24 Gigabases) se encuentra un rango de 1-10 SNP con disequilibrio mitonuclear, distribuidos de manera general a lo largo del genoma. Es decir, el efecto de disequilibrio se encuentra presente en todo el genoma, aunque de forma no muy intensa, pues son pocos SNP en ventanas muy grandes del genoma.

El mayor número de SNP nucleares en ventanas de 24 Gigabases se ubicó en una región del cromosoma 1, con un número aproximado de 346. Esta región comprende las primeras 24 Gigabases de dicho cromosoma. Así mismo, en su mayoría, estos SNP se encuentran localizados en un segmento de 5 Megabases, lo cual fue sorprendente dado que la mayoría de los SNP nucleares con disequilibrio detectados se encuentran muy separados entre sí. Como se explicará posteriormente, esta región corresponde a inserciones mitocondriales en núcleo

(NUMT). Los cromosomas con el siguiente mayor número de SNP con desequilibrio mitonuclear fueron los cromosomas 1, 2, 12, 13, 14, 15 y 22

Tabla 15. Cromosomas con mayor número de SNP con desequilibrio mitonuclear.

<i>Cromosoma</i>	<i>Promedio de SNP con desequilibrio mitonuclear</i>
1	41
2	10
12	8
13	17
14	8
15	17
22	14

Así mismo, se graficó el número de SNP mitocondriales a lo largo del genoma en ventanas de 500 bases, como se ve en la Figura 33:

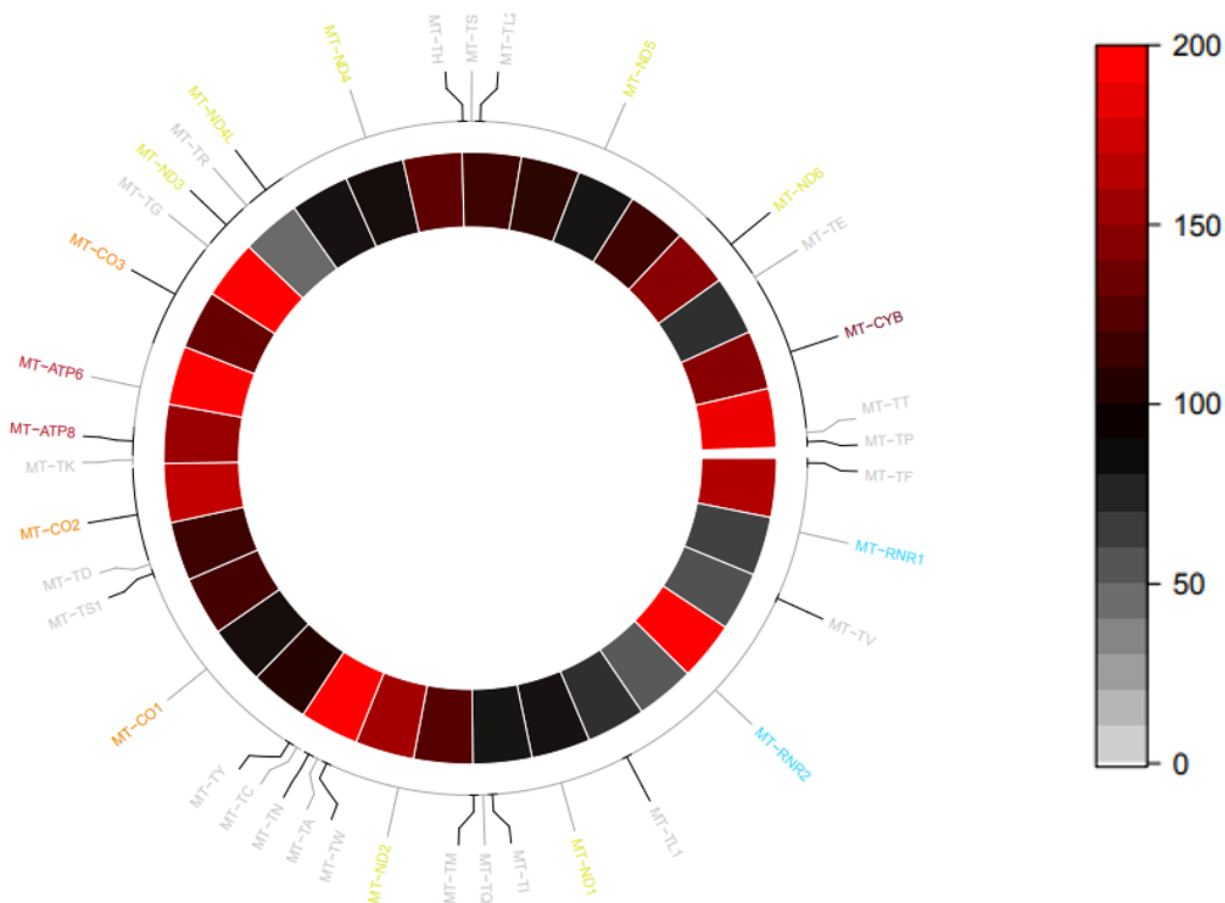


Figura 33. Número de SNP mitocondriales con desequilibrio mitonuclear significativo, en ventanas de 500 bases, del análisis con poblaciones individuales.

En la Figura 33 se aprecia que todas las regiones de 500 bases del genoma mitocondrial poseen al menos 1 SNP bajo desequilibrio mitonuclear. Las regiones ubicadas en los genes de las dos subunidades de ATP sintasa (*MT-ATP6* y *MT-ATP8*), de dos subunidades de citocromo c oxidasa (*MT-CO2* y *MT-CO3*), del 16S rARN (*MT-RNR2*), del citocromo b (*MT-CYB*), de dos subunidades de NADH ubiquinona oxidoreductasa (*MT-ND1* y *MT-ND2*) y de 5 de los tARN (*MT-TY*, *MT-TC*, *MT-TN*, *MT-TA*, *MT-TW*), fueron en las que se localizaron el mayor número de ventanas con más de 200 SNP mitocondriales bajo desequilibrio.

5.3.3 Caracterización de regiones candidatas

Las regiones con desequilibrio nucleomitocondrial de todas las poblaciones se caracterizaron usando la base de datos Ensembl. En donde se identificaron regiones las cuales tuvieran alguna función descrita. Se lograron caracterizar los siguientes porcentajes por población:

Tabla 16. Porcentaje de regiones con DMT caracterizadas.

Población	Número total de regiones con DMT	Número de regiones caracterizadas	Porcentaje de regiones caracterizadas (%)
Chinos Han en Beijing, China (CHB)	51	29	57
Japoneses en Tokyo, Japón (JPT)	111	66	60
Residentes de Utah con ancestría del Norte y Oeste de Europa (CEU)	73	11	15
Finlandeses en Finlandia (FIN)	1	1	100
Yorubas en Ibadan, Nigeria (YRI)	341	231	67
Gambianos en divisiones del oeste de Gambia (GWD)	102	87	85
Residentes de los Ángeles, con ancestría mexicana (MXL)	113	73	64
Puertorriqueños de Puerto Rico (PUR)	22	20	91
Punjabi de Lahore, Pakistan (PJL)	267	161	60
Hablantes de Bengalí de Bangladesh (BEB)	138	71	51

En la Tabla 16, se indica el número de SNP nucleares únicos caracterizados, dichos SNP únicos corresponden a SNP con coordenadas no repetidas dentro del genoma nuclear con desequilibrio mitonuclear, en donde se observa que un poco menos de un tercio de estos SNP corresponden a regiones de inserciones de ADN mitocondrial, asimismo la mitad de los SNP correspondientes a genes se localizan en regiones exónicas, y finalmente un 5% de total de los SNP a regiones regulatorias.

Porcentaje de categorías caracterizadas de **SNP** únicos nucleares con **desequilibrio mitonuclear**

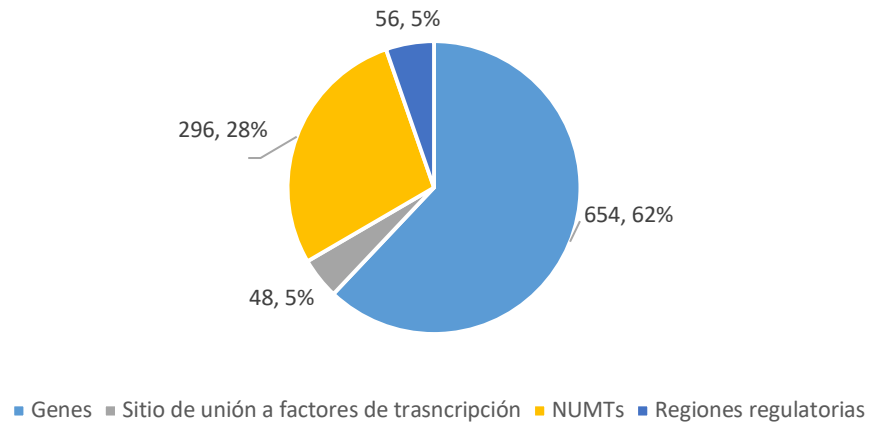


Figura 34. Gráfica de pastel representando porcentajes de los resultados de caracterización de los SNP nucleares únicos bajo **desequilibrio** (No. de SNPs, %).

Número de **SNP** mitocondriales únicos bajo **desequilibrio** caracterizados

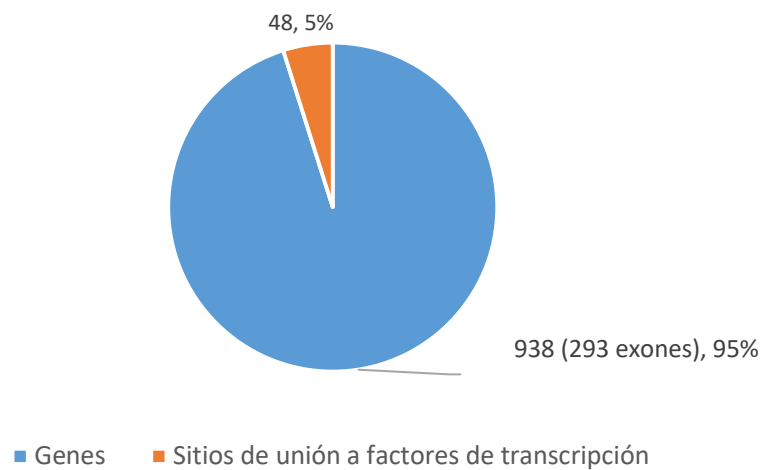


Figura 35. Gráfica de pastel representando porcentajes de los resultados de caracterización de los SNP mitocondriales únicos bajo **desequilibrio**.

Tabla 17. Resultados de la caracterización de regiones mitocondriales y nucleares.

Población	Genes nucleares	Regiones regulatorias nucleares	NUMT	Genes mitocondriales	Regiones regulatorias mitocondriales
Chinos Han en Beijing, China (CHB)	44	4	15	245	37
Japoneses en Tokyo, Japón (JPT)	94	11	36	48	7
Residentes de Utah con ascendencia del Norte y Oeste de Europa (CEU)	18	0	7	9	1
Finlandeses en Finlandia (FIN)	2	0	1	1	0
Yorubas en Ibadan, Nigeria (YRI)	334	20	152	85	14
Gambianos en divisiones del oeste de Gambia (GWD)	157	2	83	46	6
Residentes de los Ángeles, con ascendencia mexicana (MXL)	97	13	24	11	1
Puertorriqueños de Puerto Rico (PUR)	36	8	16	20	3
Punjabi de Lahore, Pakistan (PJL)	171	32	0	39	9
Hablantes de Bengalí de Bangladesh (BEB)	75	15	7	23	6

Podemos observar en la Figura 34 que un 28% de los pares de SNP únicos totales corresponden a las inserciones mitocondriales en el núcleo, 5% a sitios de unión de factores de transcripción y regiones reguladoras, y el 62% a genes. En el caso de los SNP mitocondriales(Figura 35), el 78% fueron genes y el 22% sitios de unión de factores de transcripción. La gran densidad de los SNP con desequilibrio mitonuclear en los NUMT podría suponer un fenómeno biológico de fuerte interacción de estas inserciones con el genoma mitocondrial; sin embargo, esto

podría ser efecto también de algún posible error dentro de los datos analizados, cuestión que será discutida más adelante. La lista completa de genes con desequilibrio mitonuclear puede verse en el Anexo 1.

La siguiente tabla muestra los genes nucleares con mayor número de SNP únicos con desequilibrio mitonuclear de cada población, asimismo, se indican aquellos pseudogenes que se localizan dentro de inserciones mitocondriales (NUMT) los cuales se han marcado con asteriscos

Tabla 18. Genes nucleares con mayor número de SNP únicos con desequilibrio mitonuclear en cada población.
Pseudogenes con desequilibrio mitonuclear marcados con doble asterisco.

Nombre del gen nuclear e ID de Ensembl	Población (número de SNP con DMT)
<i>TRPC6</i> (ENSG00000137672): Transient receptor potential cation channel subfamily C member 6	JPT (7)
** <i>MTCO2P12</i> (ENSG00000229344): Insertion from mitochondrially encoded cytochrome c oxidase II (pseudogene 12)	JPT (6)
** <i>MTND1P23</i> (ENSG00000225972): Insertion from mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 1 (pseudogene 23)	CHB (5)
ENSG00000264634: RNA gene	CHB (4)
** <i>MTATP8P1</i> (ENSG00000240409):Insertion from mitochondrially encoded ATP synthase 8 (pseudogene 1)	YRI (15)
<i>NPAS2</i> (ENSG00000170485): Neuronal PAS domain protein 2)	YRI(15)
<i>RALGPS2</i> (ENSG00000116191): Ral GEF with PH domain and SH3 binding motif 2	YRI (7)
<i>SPC25</i> (ENSG00000152253): SPC25, NDC80 kinetochore complex component	YRI (6)

<i>FTO</i> (ENSG00000140718): FTO, alpha-ketoglutarate dependent dioxygenase	YRI (6)
<i>ABCB11</i> (ENSG00000073734):ATP binding cassette subfamily B member 11	YRI (6)
<i>AVEN</i> (ENSG00000169857): Apoptosis and caspase activation inhibitor	MXL(22)
<i>ADGRL2</i> (ENSG00000117114): Adhesion G protein-coupled receptor L2	MXL(13)
<i>SEPT9</i> (ENSG00000184640):Septin 9	BEB (5)
<i>SFI1</i> (ENSG00000198089): SFI1 Centrin binding protein	BEB (4)
<i>COL9A1</i> (ENSG00000112280): Collagen type IX alpha 1 chain	BEB (4)
<i>CDK8</i> (ENSG00000132964): Cyclin dependent kinase 8	PJL (18)
<i>CPM</i> (ENSG00000135678): carboxypeptidase M	PJL (18)
<i>DPH6</i> (ENSG00000134146): diphthamine biosynthesis	PJL (17)
<i>DPH6-AS1</i> (ENSG00000248079): DPH6 antisense RNA 1	PJL (14)

<i>PSG8</i> (ENSG00000124467): pregnancy specific beta-1-glycoprotein 8	PUR (4)
---	---------

A continuación, se muestran los genes con mayor número de SNP con desequilibrio mitonuclear y que a su vez están compartidos por más de dos poblaciones, de igual manera se indican con asteriscos aquellos pseudogenes dentro de los NUMT.

Tabla 17. Principales genes nucleares con el mayor número de SNP únicos con desequilibrio mitonuclear en más de dos poblaciones. Pseudogenes con desequilibrio mitonuclear marcados con doble asterisco.

Nombre del gen nuclear e ID de Ensembl	Población (número de SNP con DMT)
<i>AL669831.3</i> (ENSG00000230021): (Clone-based (Ensembl) gene)	JPT(36), CHB(15), CEU(7), FIN(1), MXL(34), PUR(12), YRI(130), GWD(83), BEB(7)
** <i>MTND2P28</i> (ENSG00000225630): Insertion from mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 2 (pseudogene 28)	JPT(18), CHB(9), CEU(6), MXL(6), YRI(23), GWD(1), BEB(5)
** <i>MTATP6P1</i> (ENSG00000248527): Insertion from mitochondrially encoded ATP synthase (6 pseudogene)	JPT(12), CHB(1), CEU(1), FIN(1), YRI(28), BEB(1)
** <i>MTCO1P12</i> (ENSG00000237973): Insertion from mitochondrially encoded cytochrome c oxidase I (pseudogene 12)	JPT(12), CHB(1), CEU(1), FIN(1), YRI(28), BEB(1)
** <i>MTCO3P12</i> (ENSG00000198744): Mitochondrially encoded cytochrome c oxidase III (pseudogene)	MXL(18), PUR(4)

La tabla siguiente muestra los genes mitocondriales con el mayor número de SNP con desequilibrio mitonuclear:

Tabla 18. Principales genes mitocondriales con el mayor número de SNP únicos con desequilibrio mitonuclear.

Nombre del gen mitocondrial e ID de Ensembl	Población (número de pares de SNP)
<i>MT-ATP6P1</i> (ENSG00000198899): Mitochondrially encoded ATP synthase membrane subunit 6	JPT(4), CHB(1), CEU(1), FIN(1), YRI(5), GWD(1), MXL(2), P JL(3), BEB(2)
<i>MT-ND1</i> (ENSG00000198888): Mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 1	JPT(2), CHB(1), CEU(1), YRI(6), GWD(3), MXL(1), PUR(4), P JL(2), BEB(2)
<i>MT-ND5</i> (ENSG00000198786): Mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 5	CHB(4), CEU(1), YRI(8), GWD(7), MXL(1), PUR(3), P JL(7)
<i>MT-RNR2</i> (ENSG00000210082): Mitochondrially encoded 16S RNA	JPT(2), CHB(2), YRI(4), GWD(7), MXL(1), PUR(2), P JL(2), BEB(2)
<i>MT-ND4</i> (ENSG00000198886): Mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 4	JPT(4), CHB(2), CEU(2), YRI(4), MXL(1), P JL(3), BEB(2)
<i>MT-ND2</i> (ENSG00000198763): Mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit	JPT(4), CHB(4), CEU(2), YRI(9), GWD(5), MXL(2), P JL(3), BEB(2)
<i>MT-CO1</i> (ENSG00000198804): Mitochondrially encoded cytochrome c oxidase I	JPT(3), YRI(8), GWD(7), MXL(1), PUR(5), P JL(3), BEB(1)
<i>MT-CYB</i> (ENSG00000198727): Mitochondrially encoded cytochrome b	JPT(12), CHB(1), YRI(9), GWD(2), PUR(2), P JL(5), BEB(2)
<i>MT-CO2</i> (ENSG00000198712): Mitochondrially encoded cytochrome c oxidase II	JPT(4), CHB(2), YRI(4), GWD(5), PUR(2), P JL(3), BEB(2)
	JPT(2), YRI(4), GWD(1), PUR(2),

<i>MT-ND6</i> (ENSG00000198695): Mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 6	MXL(1), P JL(1), BEB(1)
<i>MT-CO3</i> (ENSG00000198938):Mitochondrially encoded cytochrome c oxidase III	JPT(1), YRI(7), MXL(1), PUR(1), P JL(3), BEB(1)
<i>MT-ATP8</i> (ENSG00000228253):Mitochondrially Encoded ATP Synthase Membrane Subunit 8	JPT(1), YRI(3), GWD(2), P JL(2), BEB(1)
<i>MT-RNR1</i> (ENSG00000211459): Mitochondrially encoded 12S RNA	JPT(4), CEU(1), YRI(4), GWD(2), BEB(1)

5.4 Detectando desequilibrio mitonuclear en superpoblaciones

Habiendo detectado desequilibrio mitonuclear en poblaciones individuales, el siguiente paso fue realizar el análisis utilizando todos los individuos disponibles de la base de datos de 1000 genomas. Este análisis masivo se puede comparar con el realizado en poblaciones individuales para detectar desequilibrios mitonucleares atribuibles a adaptación local por un lado y de contexto global, por otro.

El agrupamiento de las poblaciones totales en cinco superpoblaciones: europea (EUR), asiática oriental (EAS), surasiática (SAS), africana (AFR) y americanos mestizos (AMR), dio como resultado un análisis de alta resolución con cinco veces más individuos que el utilizado en el análisis de poblaciones individuales, con una media de 477 individuos, brindando un panorama más general desde la perspectiva de superpoblación del desequilibrio mitonuclear detectado.

El contexto global de los resultados se comparó entre superpoblaciones para identificar las firmas de desequilibrio mitonucleares de cada superpoblación para identificar un fenómeno de coevolución basal en cada superpoblación.

Tabla 19. Número total de combinaciones analizadas por superpoblación

Superpoblación	Número total de parejas de SNP analizadas
Asiáticos del este (EAS)	4,897,365,095
Europeos (EUR)	3,425,885,382
Africana (AFR)	13,696,807,432
Americanos mestizos (AMR)	4,227,477,204
Surasiáticos (SAS)	6,185,400,879

El total de parejas de SNP nucleares y mitocondriales analizadas en el análisis de superpoblación tuvieron la misma magnitud (1×10^9) que el número total de parejas analizadas en las poblaciones individuales. En ambos análisis los parámetros de filtrado de los SNP fueron los mismos, por lo que el tamaño poblacional no parece haber afectado significativamente el número de parejas de SNP analizadas.

5.4.1 Resultados de análisis de desequilibrio mitonuclear en superpoblaciones

La siguiente tabla muestra el porcentaje total de valores de *tau* mayores a 0.1 en cada superpoblación:

Tabla 20. Total de valores de tau mayores a 0.1 por superpoblación.

Superpoblación	Número total de combinaciones	Comparaciones con <i>tau</i> > 0.1 (%)
Asia del este (EAS)	4,897,365,095	2,585,487 (0.05%)
Europeos (EUR)	3,425,885,382	1,528,609 (0.044%)
Africanos (AFR)	13,696,807,432	6,040,540 (0.044%)
Americanos mestizos (AMR)	4,227,477,204	8,714,276 (0.20%)
Asiáticos del sur (SAS)	6,185,400,879	4,009,788 (0.06%)

El porcentaje de valores totales con τ mayor a 0.1 obtenido en el análisis de superpoblaciones fue el mismo que el obtenido en el análisis de poblaciones individuales. Ambos valores se presentaron por debajo del 2% del total de combinaciones. En todas las superpoblaciones estos valores se encontraron por debajo del 1% del total.

El número de individuos no afectó el porcentaje de comparaciones con valores bajos de τ ($\tau < 0.1$). En ambos análisis, más del 90% de las comparaciones tenían un valor de τ despreciable.

Los siguientes gráficos de caja y bigotes muestran la distribución de los valores de τ por encima de 0.1 en cada superpoblación.

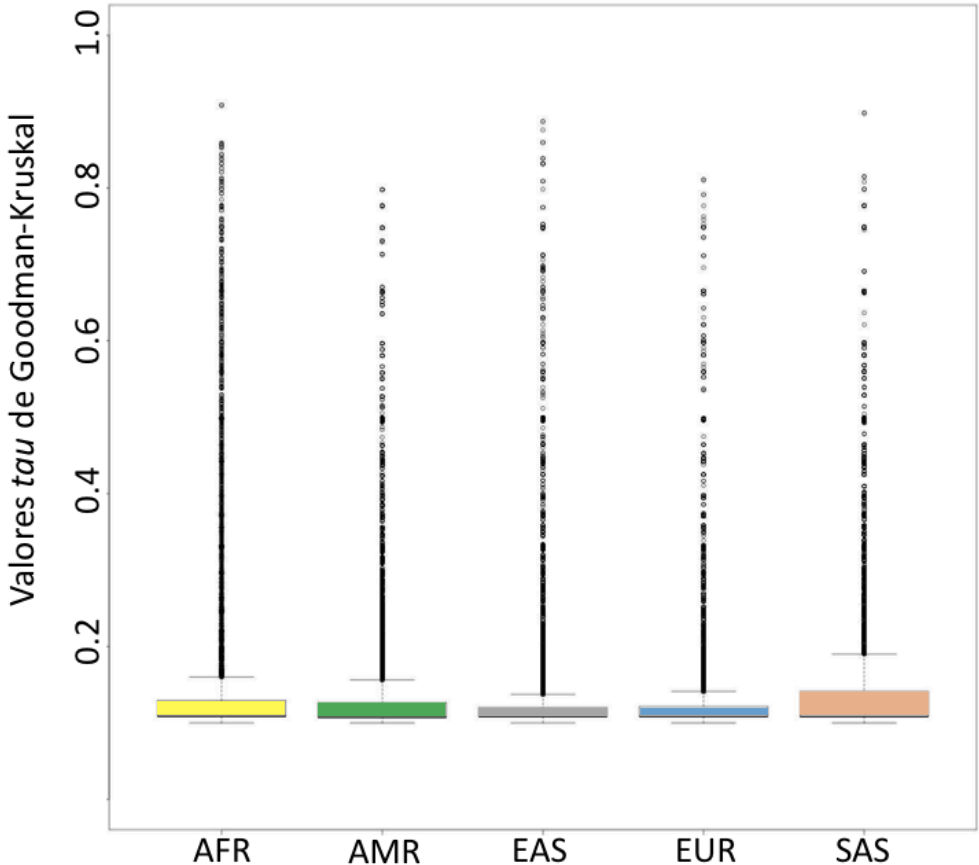


Figura 36. Gráfica de cajas y bigotes de distribución de valores tau mayores a 0.1 en superpoblaciones. AFR (africanos), AMR (americanos mestizos), EAS (asiáticos del este), EUR (europeos), SAS (asiáticos del sur).

La distribución de los valores de *tau* obtenidos en cada superpoblación mostró que más del 50% de los valores están por debajo de 0.2. Asimismo, no se alcanzaron los valores de *tau* iguales a 1. No obstante, se obtuvieron valores altos de asociación, superiores a 0.7 de *tau*.

La siguiente tabla muestra el número de valores para cada rango de *tau* en la distribución.

Tabla 21. Número de valores de tau por rango en cada superpoblación.

Superpoblación	Valores de <i>tau</i> de 0.1-0.5	Valores de <i>tau</i> de 0.5-0.7	Valores de <i>tau</i> de 0.7-0.9	Valores de <i>tau</i> =1
Asia del este (EAS)	2,585,011	319	157	0
Europeos (EUR)	1,528,316	199	94	0
Africanos (AFR)	6,038,018	1,681	841	0
Americanos mestizos (AMR)	8,711,677	1,807	792	0
Asiáticos del sur (SAS)	3,999,738	5,674	4,376	0

5.4.1.1 Resultados estadísticamente significativos

Como no se obtuvieron correlaciones perfectas ($tau = 1$), los resultados estadísticamente significativos se seleccionaron únicamente mediante el filtrado de los resultados crudos por un valor de corte, el cual se seleccionó en base al valor más alto obtenido por las simulaciones de variación aleatoria.

Tabla 22. Resultados de la prueba K-S contra la distribución simulada, p valor y máximo valor de tau obtenido en las simulaciones.

Superpoblación	Resultados prueba de K-S (D)	P valor	Valor máximo de tau
Asia del este (EAS)	0.5586	$2.2 \cdot 10^{-16}$	0.4425
Europeos (EUR)	0.68115	$2.2 \cdot 10^{-16}$	0.499
Africanos (AFR)	0.44269	$2.2 \cdot 10^{-16}$	0.442
Americanos mestizos (AMR)	0.44785	$2.2 \cdot 10^{-16}$	0.580
Asiáticos del sur (SAS)	0.56195	$2.2 \cdot 10^{-16}$	0.499

Los siguientes gráficos muestran la distribución acumulativa de los datos simulados contra los reales para las distintas superpoblaciones:

Europeos (EUR)

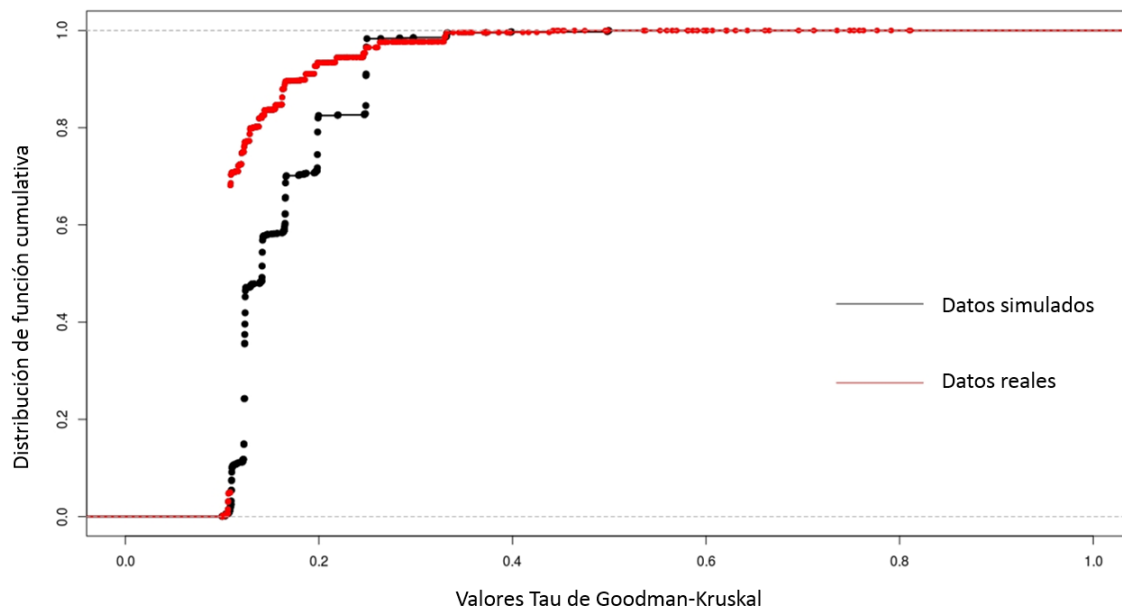


Figura 37. Distribuciones de función acumulativa de datos reales vs. simulados de tau en la superpoblación europea.

Asáticos del este (EAS)

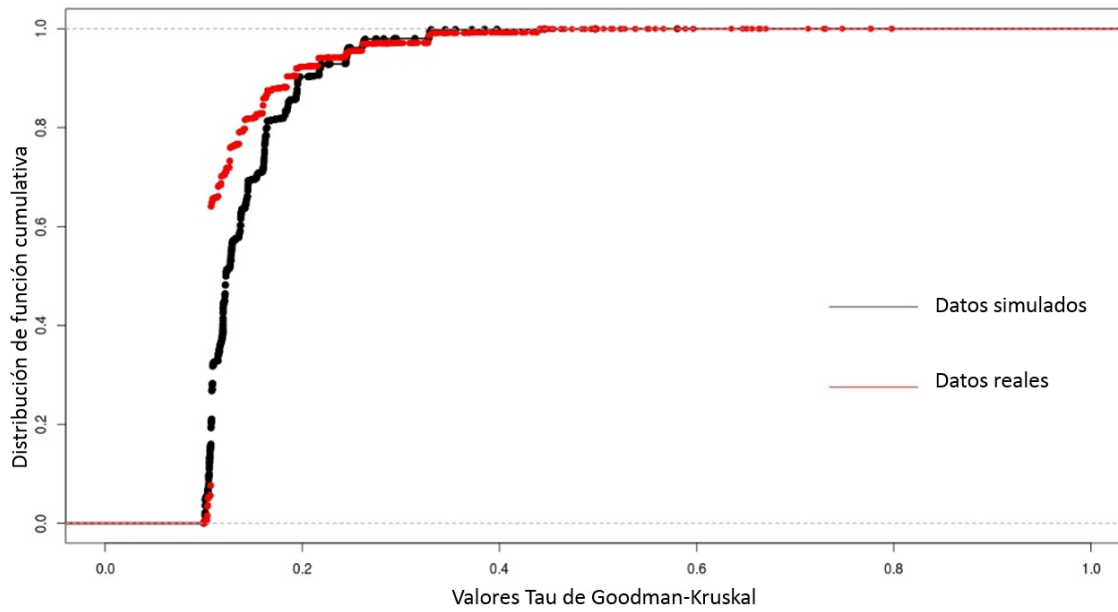


Figura 38. Distribuciones de función de distribución acumulativa de datos reales vs. simulados de tau en la superpoblación de Asia del este.

Africanos (AFR)

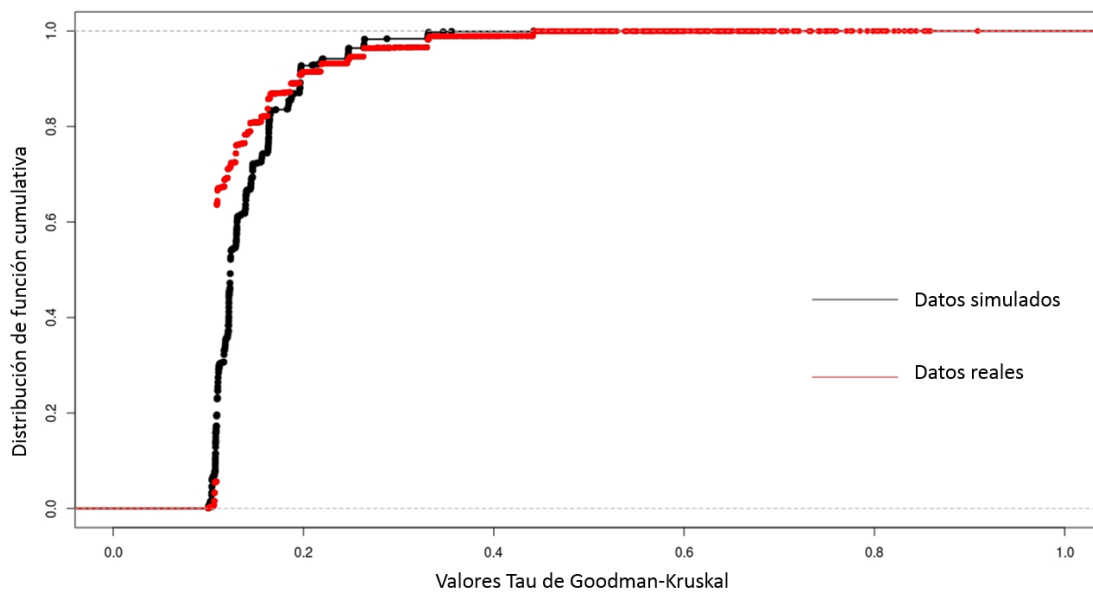


Figura 39. Distribuciones de función de distribución acumulativa de datos reales vs. simulados de tau en la superpoblación africana.

Americanos mestizos (AMR)

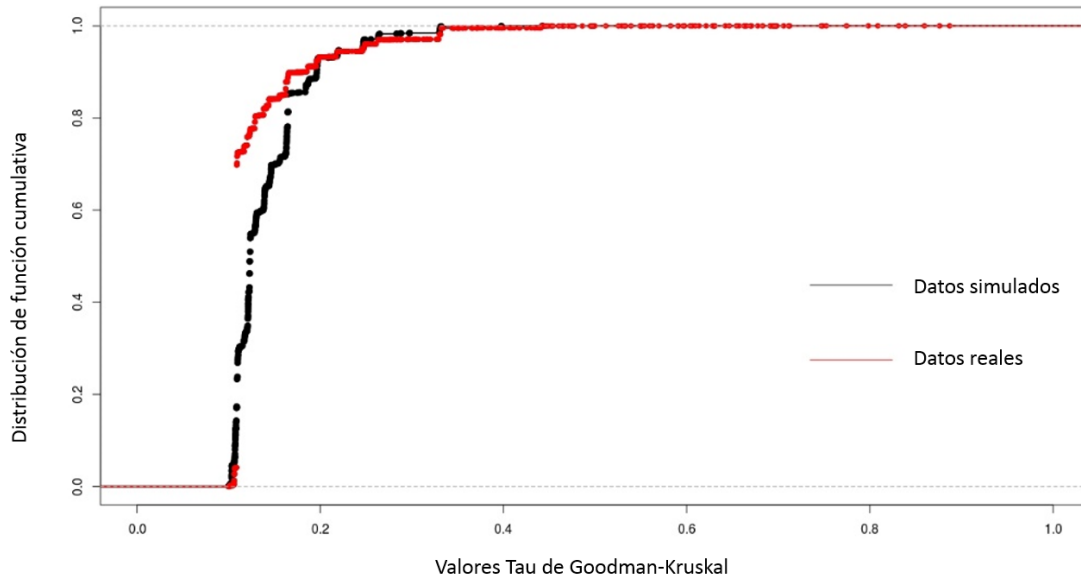


Figura 40. Distribuciones de función cumulativa de datos reales vs. simulados de tau en la superpoblación americana mestiza.

Surasiáticos (SAS)

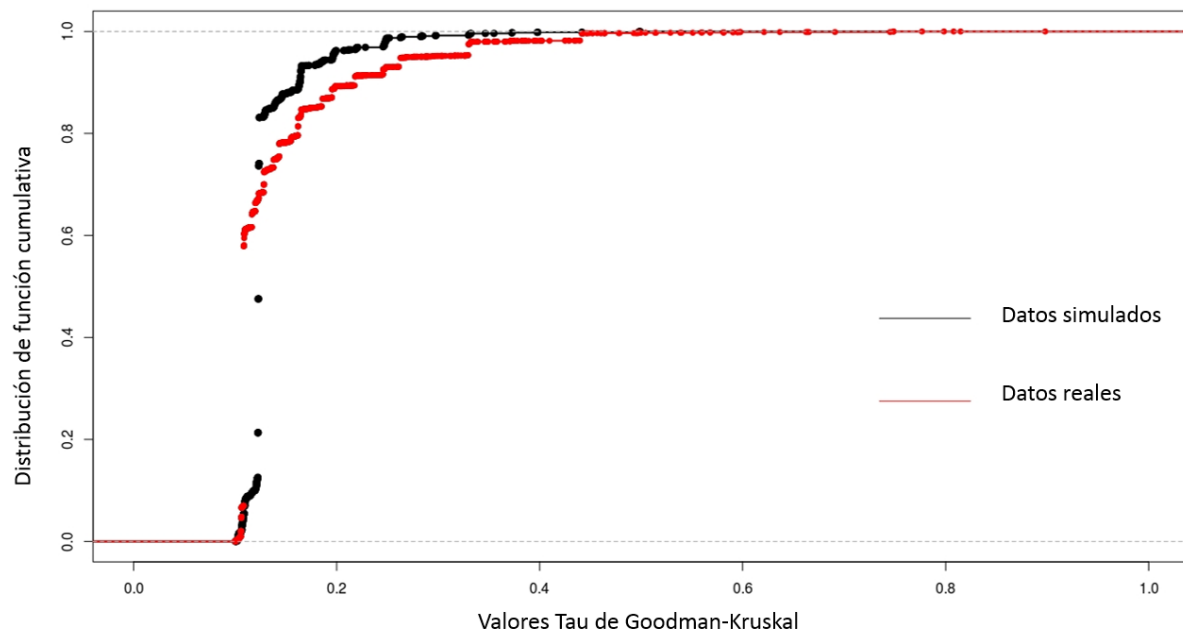


Figura 41. Distribuciones de función cumulativa de datos reales vs. simulados de tau en la superpoblación surasiática.

Los valores máximos de τ obtenidos en las simulaciones se encontraron en un rango de 0.44-0.58. En comparación, el efecto aleatorio en los valores de τ dadas las simulaciones fue más fuerte en las poblaciones individuales que en superpoblaciones, siendo que en la simulación de variación en poblaciones individuales los valores máximos fueron alrededor de 0.79.

La significancia estadística de los resultados del análisis de superpoblaciones puede establecerse en valores superiores a 0.58. Para el propósito del trabajo en la detección de desequilibrio mitonuclear se decidió seleccionar los SNP nucleares y mitocondriales con mayor asociación, los resultados del análisis por superpoblación se filtraron con un valor conservador de corte de 0.78, basado en los puntos de corte obtenidos en el análisis de poblaciones individuales.

Los valores p de la prueba de K-S fueron significativos en el análisis de superpoblaciones, lo que indica que la distribución de los valores reales no es resultado de un efecto aleatorio.

5.4.1.2 Resultados significativos y caracterización

Aplicando el parámetro de filtrado establecido en la sección anterior de 0.78, se obtuvieron los siguientes resultados significativos de τ :

Tabla 23. SNP con desequilibrio mitonuclear estadísticamente significativo en distintas superpoblaciones.

Superpoblación	Total de pares de SNP con desequilibrio mitonuclear estadísticamente significativos
Asia del este (EAS)	148
Europeos (EUR)	88
Africanos (AFR)	794
Americanos mestizos (AMR)	36
Asiáticos del sur (SAS)	220

Las superpoblaciones con mayor número de parejas de SNP con desequilibrio mitonuclear estadísticamente significativas fueron los africanos (AFR) y los asiáticos del sur (SAS), caso que de igual manera se observó en el análisis de poblaciones individuales.

Las superpoblaciones con menor desequilibrio fueron los americanos mestizos y europeos, lo cual también corresponde a lo obtenido en el análisis de poblaciones individuales.

5.4.2 Localización genómica de desequilibrio mitonuclear en superpoblaciones

En las siguientes Figuras podemos ver la densidad de los SNP con desequilibrio mitonuclear en el genoma nuclear (Figura 42) y el genoma mitocondrial (Figura 43).

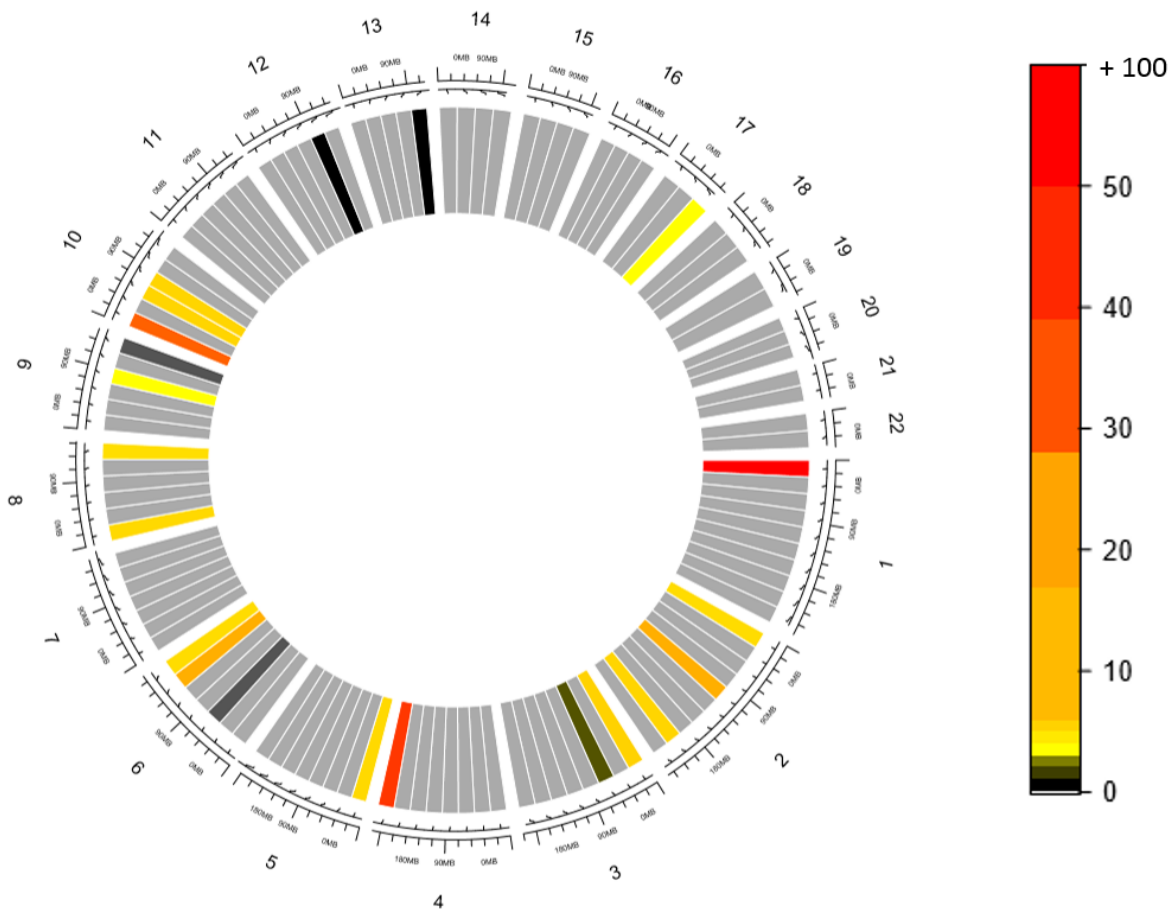


Figura 42. Densidad de los SNP nucleares con desequilibrio mitonuclear en regiones de 24 gigabases en el análisis con superpoblaciones. En este caso, los cromosomas con mayor densidad de SNP bajo desequilibrio son el 1, 4 y 10, seguidos del 2 y 6.

Todas las regiones de 24 gigabases poseen al menos un SNP bajo desequilibrio mitonuclear. En comparación al análisis de poblaciones individuales, el número de SNP con desequilibrio mitonuclear se encuentra menos dispersos en el análisis de

superpoblaciones, pero coincide en que la mayor densidad de estos SNP se encuentran en el primer bloque de 24 gigabases del cromosoma uno, con más de 100 SNP, lo cual como se verá en la caracterización de los loci, corresponde a regiones de inserciones mitocondriales. Los cromosomas 1, 2, 3, 4, 6 y 10, fueron los de mayor desequilibrio.

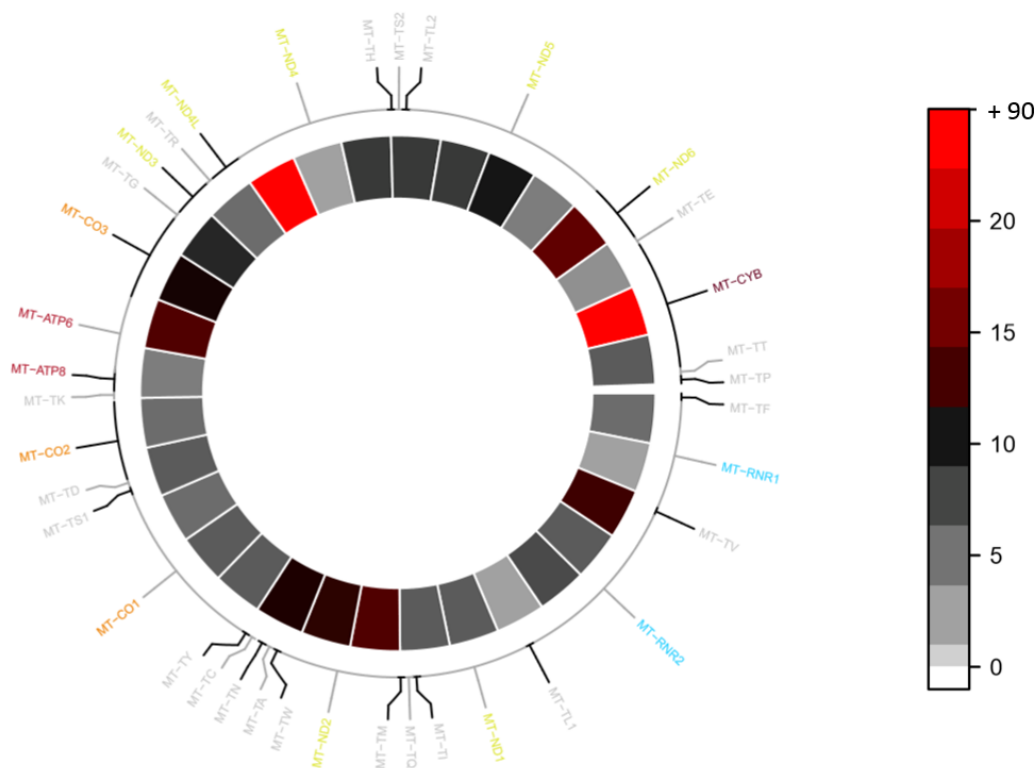


Figura 43. Densidad de los SNP mitocondriales con desequilibrio mitonuclear en ventanas de 500 bases en el análisis con superpoblaciones.

En la Figura 43 se aprecia que todas las regiones de 500 bases del genoma mitocondrial poseen al menos 1 SNP bajo desequilibrio mitonuclear. Al igual que en el caso de la densidad en el genoma nuclear, la densidad en el mtADN se observa menos dispersa que la observada en el análisis de poblaciones. En consecuencia, la concentración de desequilibrio mitonuclear fue en menos genes; en el caso de superpoblaciones los genes con mayor densidad fueron *citocromo b* (*MT-CYB*) y una subunidad de *NADH* ubiquinona oxidoreductasa (*MT-ND4*), que no fue de las destacadas en las poblaciones individuales.

5.4.3 Caracterización de regiones candidatas en el análisis de superpoblaciones

Más del 50% de los SNP en las poblaciones se lograron caracterizar, siendo que estos se ubicaron en regiones anotadas del genoma humano. La población surasiática tuvo el menor porcentaje de SNP caracterizados, en la población europea se logró caracterizar la totalidad de sus SNP. Lo cual correlaciona negativamente con el total de SNP obtenidos: las poblaciones con menor porcentaje de SNP caracterizados fueron aquellas con el mayor número de parejas de SNP (asiáticos del este, asiáticos del sur y africanos), y las de menor número de parejas de SNP, tuvieron el mayor porcentaje de SNP con alguna función descrita (americanos mestizos y europeos).

Tabla 24. Número y porcentajes de regiones caracterizadas totales en el análisis de superpoblaciones.

Superpoblación	Total de pares de SNP con desequilibrio mitonuclear	Total de pares de SNP caracterizados	Porcentaje de SNP caracterizados (%)
Asia del este (EAS)	128	78	60.93 %
Europeos (EUR)	8	8	100 %
Africanos (AFR)	794	480	60.45%
Americanos mestizos (AMR)	36	35	97.22%
Asiáticos del sur (SAS)	220	114	51.81%

Los porcentajes por categoría obtenidos en la caracterización de superpoblaciones (Figura 45) fueron muy similares a los obtenidos en el análisis de poblaciones individuales.

De igual manera, las inserciones mitocondriales en núcleo (Figura 45) conformaron más de $\frac{1}{4}$ de los datos, siendo 28% del total; los genes conformaron la mayoría con un 66%; sitios de unión a factores de transcripción un 7%; y regiones regulatorias 2%. Dentro del genoma mitocondrial, un 91% correspondió a genes y 9% a sitios de unión a factores de transcripción.

Tabla 25. Número de SNP nucleares únicos por categoría caracterizados en las superpoblaciones.

Categoría	Número de SNP nucleares únicos con desequilibrio caracterizados
Genes	454 (178 exones)
Sitios de unión a factores de transcripción	16
NUMT	172
Regiones regulatorias	45

Número de **SNP** nucleares únicos bajo desequilibrio caracterizados

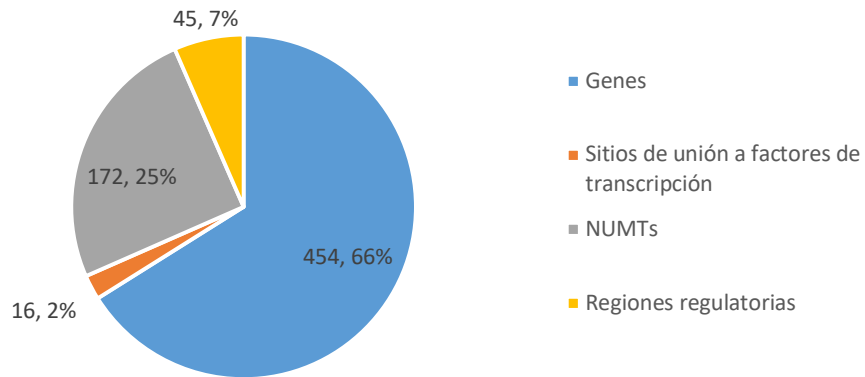


Figura 44. Gráfica de pastel representando porcentajes de los resultados de caracterización de SNP nucleares únicos bajo desequilibrio en las superpoblaciones analizadas.

En cuanto a los SNP localizados en el genoma mitocondrial, un 91% correspondieron a regiones exónicas, asimismo, detectamos que 9% del total de SNP se localizaron en regiones regulatorias, correspondientes a sitios de unión de factores de transcripción.

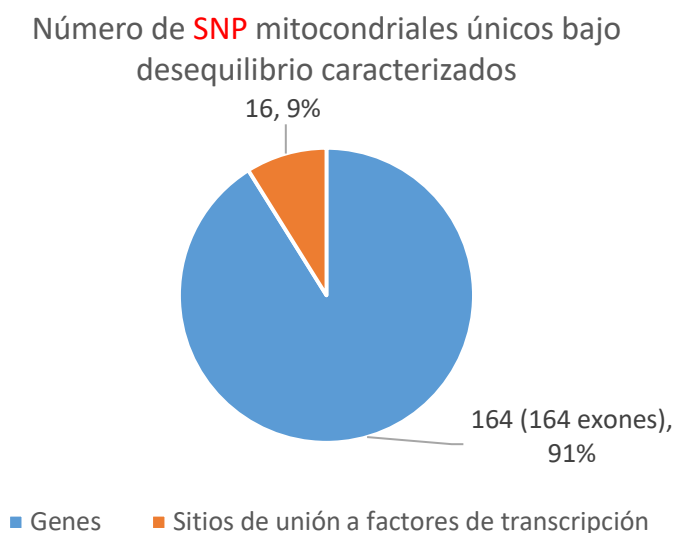


Figura 45. Gráfica de pastel representando porcentajes de los resultados de caracterización de SNP mitocondriales únicos bajo desequilibrio en superpoblaciones.

Tabla 26. Resultados de caracterización de regiones mitocondriales y nucleares en las superpoblaciones analizadas.

Superpoblación	Genes nucleares	Regiones regulatorias nucleares	Numt	Genes mitocondriales	Regiones regulatorias mitocondriales
Asia del este (EAS)	104	20	31	56	4
Europeos (EUR)	29	1	8	19	0
Africanos (AFR)	626	82	153	155	21
Americanos mestizos (AMR)	46	1	12	15	0
Asiáticos del sur (SAS)	123	17	10	21	3

La siguiente tabla muestra los genes nucleares con mayor número de SNP únicos con desequilibrio mitonuclear únicos en superpoblaciones, asimismo se indican aquellos pseudogenes que localizaron dentro de inserciones mitocondriales (NUMT) los cuales se han marcado con asteriscos

Tabla 27. Principales genes nucleares con mayor número de SNP únicos con desequilibrio mitonuclear en más de dos superpoblaciones. Pseudogenes correspondientes a NUMT marcados con doble asterisco.

Nombre del gen nuclear e ID de Ensembl	Población (número de pares de SNP)
AL669831.3 (ENSG00000230021): (Clone-based (Ensembl) gene)	EAS(31), CEU(8), AFR(135), AMR(12), SAS(10)
** <i>MTATP6P1</i> (ENSG00000248527): Insertion from mitochondrially encoded ATP synthase 6 pseudogene	EAS(11), CEU(2), AFR(42), SAS(8)
** <i>MTND2P28</i> (ENSG00000225630): Insertion from mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 2 (pseudogene 28)	EAS(9), CEU(5), AFR(46), SAS(2)
** <i>MTCO1P12</i> (ENSG00000237973): Insertion from mitochondrially encoded cytochrome c oxidase I (pseudogene 12)	EAS(2), AFR(26), AMR(4)
** <i>MTATP8P1</i> (ENSG00000240409): Insertion from mitochondrially encoded ATP synthase 8 (pseudogene 1)	EAS(2), AFR(7)
** <i>MTCO3P12</i> (ENSG00000198744): Insertion from mitochondrially encoded cytochrome c oxidase III (pseudogene)	AFR(7), AMR(3)
SGCZ (ENSG00000185053): Sarcoglycan zeta	AFR(11), SAS(2)

La siguiente tabla muestra los principales genes nucleares con mayor número de SNP únicos con desequilibrio, específicos de cada superpoblación. De igual manera, se indican con asteriscos aquellos pseudogenes dentro de los NUMT.

Tabla 28. Genes nucleares con mayor número de SNP únicos con desequilibrio mitonuclear en cada superpoblación. Pseudogenes con desequilibrio mitonuclear marcados con asterisco.

Nombre del gen nuclear e ID de Ensembl	Población (número de pares de SNP)
<i>AL031599.1</i> (ENSG00000224228): Clone-based (Ensembl) gene)	AFR(13)
** <i>MTCO1P40</i> (ENSG00000262902): Insertion from mitochondrially encoded cytochrome c oxidase I pseudogene 40	AFR(8)
** <i>MTND4P12</i> (ENSG00000247627): Insertion fomr mitochondrially Encoded NADH:Ubiquinone Oxidoreductase Core Subunit 4 Pseudogene 12	AFR(8)
<i>CSMD1</i> (ENSG00000183117): CUB and Sushi multiple domains 1	AFR(8)
<i>PCBD2</i> (ENSG00000132570): proterin-4 alpha-carbinolamine dehydratase 2	AFR(8)
<i>LINC01520</i> (ENSG00000230962): long intergenic non-protein coding RNA 1520	AMR(9)
<i>EGR2</i> (ENSG00000122877): early growth response 2	AMR(8)
<i>TMC1</i> (ENSG00000165091): transmembrane channel like 1	AMR(5)
<i>PCDH8P1</i> (ENSG00000225510): Protocadherin 8 Pseudogene	EUR(6)

SOX2-OT(ENSG00000242808): SOX2 overlapping transcript	EUR(3)
FAM184A(ENSG00000111879): family with sequence similarity 184 member A	EUR(2)
DPP10(ENSG00000175497): dipeptidyl peptidase like 10	SAS(16)
LINC024492(ENSG00000250590): Long Intergenic Non-Protein Coding RNA 2492	SAS(8)
CELF2(ENSG00000048740): CUGBP Elav-like family member 2	SAS(8)
PLCL2(ENSG00000154822): phospholipase C like 2	SAS(6)
PHACTR2(ENSG00000112419): phosphatase and actin regulator 2	SAS(6)
LRP2(ENSG00000081479): LDL receptor related protein 2	EAS(3)
LINC01481(ENSG00000257815)long intergenic non-protein coding RNA 1481	EAS(2)
CTNND2(ENSG00000169862): catenin delta 2	EAS(2)

La tabla siguiente muestra los genes mitocondriales con el mayor número de SNP con desequilibrio mitonuclear:

Tabla 29. Principales genes mitocondriales con el mayor número de SNP únicos con desequilibrio mitonuclear.

Nombre de gen mitocondrial e ID de Ensembl	Población (número de parejas de SNP)
<i>MT-ND5</i> (ENSG00000198786):Mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 5	EAS(5), EUR(2), AFR(20), AMR(2), SAS(2)
<i>MT-ND2</i> (ENSG00000198763):Mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit	EAS(7), EUR(6), AFR(13), AMR(2), SAS(2)
<i>MT-ATP6P1</i> (ENSG00000198899): Mitochondrially encoded ATP synthase membrane subunit 6	EAS(3), EUR(2), AFR(10), SAS(2)
<i>MT-CYB</i> (ENSG00000198727): Mitochondrially encoded cytochrome b	EAS(7), AFR(11), AMR(2), SAS(5)
<i>MT-ND4</i> (ENSG00000198886):Mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 4	EUR(2), AFR(12), SAS(2)
<i>MT-CO1</i> (ENSG00000198804):Mitochondrially encoded cytochrome c oxidase I	EUR(2), AFR(18), EAS(6)
<i>MT-ATP8</i> (ENSG00000228253): Mitochondrially encoded ATP synthase membrane subunit 8	EAS(3), AFR(4)

<i>MT-RNR1</i> (ENSG00000211459): Mitochondrially encoded 12S RNA	EAS(2), AFR(8)
<i>MT-TT</i> (ENSG00000210195): mitochondrially encoded tRNA threonine	EAS(3), AFR(6)
<i>MT-ND4</i> (ENSG00000198886):Mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 4	EUR(2), AFR(12), SAS(2)
<i>MT-CO1</i> (ENSG00000198804):Mitochondrially encoded cytochrome c oxidase I	EUR(2), AFR(18), EAS(6)

5.5 Resultados del análisis de enriquecimiento de términos GO en poblaciones y superpoblaciones

Utilizando el software de ShinyGO (Ge & Jung, 2018), no se encontraron términos GO enriquecidos significativos con un valor p de 0.05 en proceso biológico, componente celular y función molecular. No obstante, se presentan a continuación los grupos con mayor número de genes correspondientes a cada categoría de GO:

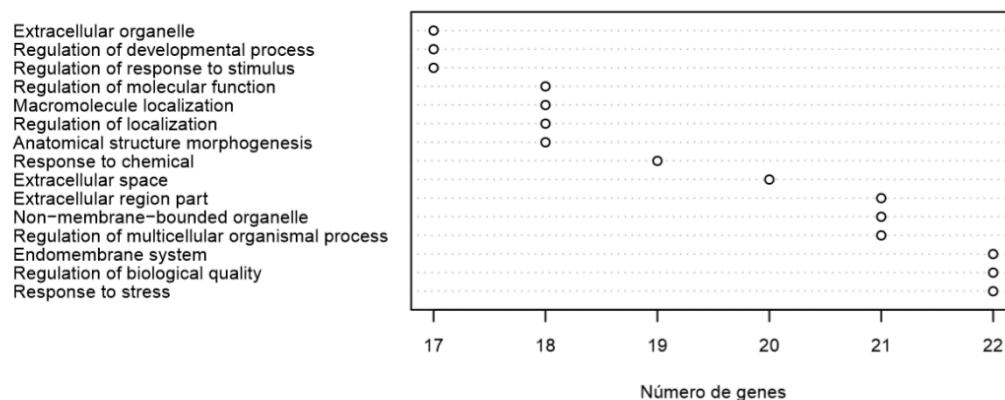


Figura 47. Términos GO con mayor número de genes en el análisis de poblaciones individuales.

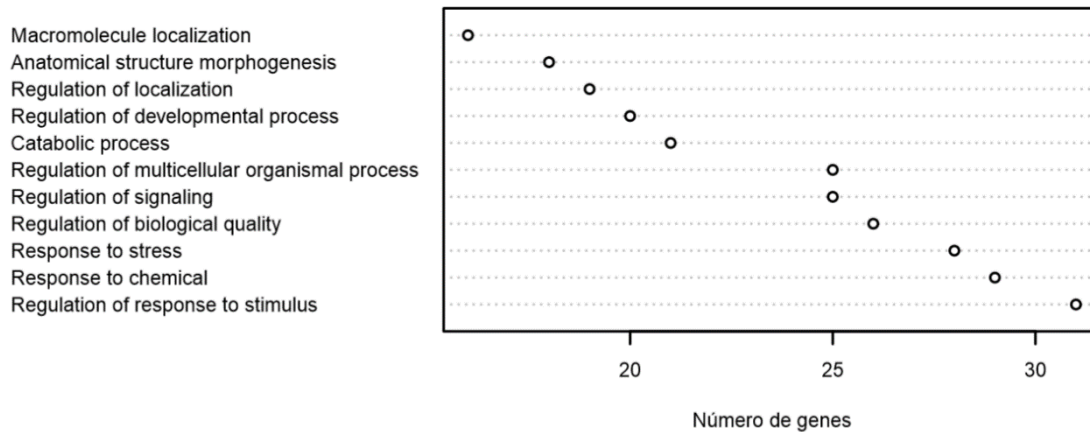


Figura 48. Términos GO con mayor número de genes en el análisis de superpoblaciones.

En ambos análisis, por poblaciones individuales y por superpoblaciones, no se obtuvieron términos GO enriquecidos. Esto puede deberse a que se cuenta con muy pocos genes, aproximadamente 145, en el análisis de poblaciones individuales y 166 en el de superpoblaciones. Asimismo, como se vio en las Tablas 26, 27, 38 y 39, los genes con mayor desequilibrio pertenecen a pseudogenes, por lo que dentro de los genes detectados existen algunos sin una función descrita.

Sin embargo, el software agrupó a los genes por categorías, coincidiendo entre ambos análisis en las categorías de “respuesta a químicos” y “respuesta a estrés” como las categorías con más genes.

6 Discusión

6.1 Poder de resolución de los análisis en búsqueda de desequilibrio mitonuclear

El número de parejas de SNP nucleares y mitocondriales analizadas tuvo una magnitud del orden de 1×10^9 ; dichas parejas se conformaron al utilizar toda la variación disponible dentro de la base de datos de 1000 genomas. Este volumen de comparaciones contrasta con los trabajos previos que intentan detectar firmas de desequilibrio mitonuclear (Rishishwar & Jordan, 2017; Sloan et al., 2015), en los que solo utilizan 350 SNP nucleares y haplotipos mitocondriales, despreciando así la variación total disponible. En contraste, el actual estudio posee una mayor resolución al analizar todas las combinaciones posibles de SNP mitonucleares en una población.

El total de valores con correlaciones altas fue muy bajo representando menos del 1% de las combinaciones totales, por lo que los resultados coinciden con el estudio de Sloan *et al.* (2015) en que la señal global de desequilibrio mitonuclear puede ser muy débil. Sin embargo, en nuestro estudio, dada la resolución, sí nos fue posible detectar asociaciones fuertes y significativas.

En comparación con el trabajo realizado por Rishishwar y col. (2017), en el cual la detección de asociación entre los genomas nucleares y mitocondriales se realizó mediante una correspondencia entre las distancias nucleares entre individuos y sus haplogrupos mitocondriales, el presente trabajo toma en cuenta toda la variación y detecta asociación directa entre alelos mitocondriales y nucleares, no limitándose a solo variantes pertenecientes a haplogrupos mitocondriales. Obteniendo un nivel de resolución a nivel SNP.

6.2 Efectos aleatorios en las asociaciones entre SNP nucleares y mitocondriales

En el análisis preliminar que se realizó con dos poblaciones se pudo observar que pueden producirse asociaciones perfectas por efectos meramente aleatorios.

Específicamente, en la Figura 22 se logra apreciar que el estadístico *tau* es muy sensible a variantes en muy baja frecuencia dentro de la población, dando como resultado que en aquellas parejas de SNP con tan solo un individuo variando, se esperasen más del 90% de valores perfectos de *tau* debidos al azar.

Dados los parámetros de filtrado establecidos en el análisis preliminar, se pudo observar que el utilizar un mínimo de tres individuos variando se logran filtrar las asociaciones perfectas no significativas, donde el número de asociaciones perfectas por azar aplicando los filtros en las poblaciones es menor a una pareja de SNP.

Dichas asociaciones perfectas no significativas representaron una gran parte del total de asociaciones perfectas detectadas, donde las poblaciones filtradas presentan una disminución en el número de correlaciones perfectas con una magnitud de 1×10^6 en comparación con las poblaciones china y japonesa (utilizadas en los análisis preliminares).

En el caso de las asociaciones altas, pero con $\tau < 1$, las simulaciones permitieron establecer un valor de corte de confianza y comparando entre los análisis de superpoblaciones y poblaciones, se observó que el utilizar un mayor número de individuos incrementa la confianza en valores más bajos, dado que los valores máximos obtenidos en las simulaciones fueron menores en las superpoblaciones.

6.3 Alto número de SNP con desequilibrio mitonuclear en poblaciones africanas y surasiáticas

En los análisis de búsqueda de desequilibrio mitonuclear, las poblaciones africanas y surasiáticas destacaron por ser las que presentaron un mayor número de SNP con desequilibrio mitonuclear, fenómeno que se observó en el análisis tanto con poblaciones individuales como con superpoblaciones.

Dentro del análisis de poblaciones se detectó un alto número de desequilibrio mitonuclear en poblaciones del sur de Asia (PJL y BEB), resultado que resalta ya que estas poblaciones poseen un número menor de parejas analizadas, en comparación con el resto de las poblaciones. La población Punjabi (PJL), tiene el segundo menor número de parejas analizadas (1,091,872,700 parejas) y son la segunda población con el mayor número de desequilibrio mitonuclear detectado (267 pares de SNP). Asimismo, la población bengalí (BEB) tiene el menor número de combinaciones (678,883,280 pares de SNP) y es la tercera población con el mayor número de SNP con desequilibrio mitonuclear detectado (138 pares de SNP). Los números de parejas de SNP con desequilibrio detectado en las poblaciones de Asia del sur, de 138 (BEB) y 267 (PJL), se asemejan a los números de parejas con desequilibrio detectadas en las poblaciones africanas, de 102 (GWD) y 341 (YRI).

En el análisis utilizando superpoblaciones se repitió el fenómeno, ya que las superpoblaciones africana y surasiática fueron las de mayor número de SNP con desequilibrio. Pero en este caso, ambas presentaban el mayor número de parejas analizadas: africanos con 13,696,807,432 parejas de SNP y surasiáticos con 6 185 400,879 parejas.

En las poblaciones africanas se podría esperar un número mayor de SNP con desequilibrio mitonuclear debido a su alta variación, en comparación con las poblaciones no africanas, influenciada por su historia demográfica (Campbell y

Tishkoff, 2008). En las poblaciones de Asia del sur, que no presentan tanta diversidad, una posible explicación es su historia demográfica y cultural, ya que se trata de una población con una estructura poblacional muy definida y en gran parte endogámica (Majumder, 2010), lo que puede repercutir en la detección de posibles firmas de interacción entre el genoma nuclear y mitocondrial detectadas. En una población muy endogámica, la variación observada en los SNP nucleares y mitocondriales puede presentar una covariación debido a que los individuos dentro del panel se encuentren muy relacionados parentalmente.

6.4 Regiones con firmas de desequilibrio mitonuclear

6.4.1 Regiones con firmas de desequilibrio mitonuclear en poblaciones individuales

Los SNP nucleares con firmas de desequilibrio compartidos entre poblaciones se encuentran en pseudogenes, que corresponden a una inserción mitocondrial en el núcleo (NUMT). Estos pseudogenes, localizados en el cromosoma 1, son copias de subunidades mitocondriales del citocromo c (*MTCO1P12* y *MTCO3P12*), del complejo NADH (*MTND2P28*) y de la ATP sintasa (*MTATP6*), y presentan una elevada densidad de SNP con desequilibrio mitonuclear (Figura 33). Asimismo, se observó que un 42% de los SNP únicos con desequilibrio mitonuclear corresponden a NUMT, siendo aproximadamente 274 de los 654 SNP identificados en total en las poblaciones. El NUMT en el cual se encuentra la mayor densidad de desequilibrio mitonuclear es el correspondiente al ID HSA_NumtS_001_b1, que se encuentra en la posición 564,464 a 570,304 del cromosoma 1, reportada por Lascaro *et al.*, (2008). También cabe mencionar que no fue la única inserción en la cual se detectó desequilibrio mitonuclear; el NUMT con ID HSA_NumtS_472_b1, ubicado en el cromosoma 13 en la posición 110 076 661 a 110 076 727, reportado también por Lascaro y col. (2008), presentó también diversos SNP con desequilibrio mitonuclear. Sin embargo, presentó un total de 22, lo que representa solo el 7% del total encontrado en los NUMT.

El porcentaje restante se ubicó en el NUMT HSA_NumtS_001_b1. No obstante, la alta concentración de desequilibrio mitonuclear en este NUMT podría deberse a un error de ensamblado, en el que múltiples secuencias mitocondriales hubiesen sido alineadas por error al genoma de referencia nuclear, en la base de datos de 1000 Genomas. La hipótesis alternativa sería la existencia de un fenómeno fuerte de coevolución en esta región, lo que implicaría que este NUMT fuese funcional. Cabe hacer notar que esta concentración de SNP no se observa en el resto de los NUMT reportados (Lascaro *et al.*, 2008). Esta cuestión será discutida más adelante (sección 6.7).

Todas las poblaciones presentaron más de un SNP con desequilibrio mitonuclear en el gen **AL669831.3**, el cual está clasificado como “*clone-based Ensembl gene*”, anotado de manera manual por parte del proyecto HAVANA (Hancock, 2004) como un pseudogen, sin función descrita pero con evidencia de que se transcribe a través de la observación de sus transcritos. Asimismo, se le clasificó como “*clone-based*” debido a que el transcrito no se asoció a ningún símbolo de HGNC (*Hugo Gene Nomenclature Committee*) (Yates *et al.*, 2017), por lo que se utilizó el nombre asignado por ENSEMBL.

En cuanto a las regiones con firmas de desequilibrio específicos de cada población, se obtuvieron de igual manera pseudogenes ubicados en los NUMT con IDs HSA_NumtS_001_b1 y HSA_NumtS_472_b1. Asimismo, se lograron identificar genes con firmas de desequilibrio específicos de cada población no pertenecientes a inserciones mitocondriales, pero con interacción mitocondrial descrita, lo que podría explicar su situación de desequilibrio mitonuclear:

TRPC6. Gen con desequilibrio mitonuclear en población japonesa, que codifica una proteína potencial receptora-transitoria de canal. Esta proteína de membrana permite el paso de iones de calcio y es activada por diacilglicerol (D’Esposito *et al.*, 1998). Se realizó un estudio enfocado en entender las funciones y propiedades de TRPC6 de este canal bajo los efectos de la hiperforina, antidepressivo cuyos efectos se centran en el transporte iónico, y se observó que colapsa el potencial de membrana mitocondrial, liberando iones de calcio y zinc de la mitocondria (Tu, Gibon, & Bouron, 2010), ligando el transporte mitocondrial de calcio con efectos neuronales.

FTO. Gen con desequilibrio mitonuclear en población yoruba. El gen recibe sus siglas por “*Fat mass-and obesity-associated*”, es decir es un gen asociado a obesidad e incremento de masa corporal, también conocido como dioxigenasa dependiente de alfa-cetoglutarato. Se hipotetiza que está involucrado en reparación de ADN/ARN (Tews *et al.*, 2013). Asimismo, se ha identificado que su regulación río abajo suprime la biogénesis mitocondrial y producción de energía (Wang *et al.*, 2017).

ABCB11. Gen con desequilibrio mitonuclear en población yoruba. Codifica para la proteína BSEP, por las siglas en inglés de “*Bile Salt Export Pump*”, que se encarga de transportar sales con la finalidad de mantener la homeostasis en el hígado, utilizando energía proporcionada por las moléculas de ATP. Se reporta que esta proteína se expresa exclusivamente en células de hígado (Strautnieks *et al.*, 1998), y su relación con la actividad metabólica mitocondrial, al observar que en ratones con el gen ABCB11 noqueado presentan una expresión reducida en genes involucrados en el metabolismo de ácido graso en la mitocondria en células hepáticas (Zhang *et al.*, 2012).

AVEN. Gen con desequilibrio mitonuclear en población mexicana. Codifica para la proteína AVEN, cuya función reportada es antiapoptótica, en donde interactúa de

manera sinérgica con la proteína Bcl-x(L), la cual es un regulador negativo de la apoptosis (Chau, Cheng, Kerr, & Hardwick, 2000; Figueroa, Chen, Oyler, Hardwick, & Betenbaugh, 2004), ruta en la cual está fuertemente involucrada la mitocondria, en cuya membrana se ubica la proteína Bcl-x(L) (Estaquier, Vallette, Vayssiere, & Mignotte, 2012).

Los genes mitocondriales con mayor número de SNP con desequilibrio mitonuclear identificados corresponden a subunidades de ATPasa (*MT-ATP61* y *MT-ATP8*), subunidades de NADH:ubiquinona oxidorreductasa (*MT-NAD1*, *MT-NAD4*, *MT-NAD5*, *MT-NAD6*), citocromo b oxidasa (*MT-CYB*), subunidades de citocromo c oxidasa (*MT-CO1*, *MT-CO2*, *MT-CO3*) y los rARN 12s y 16s (*MT-RNR1* y *MT-RNR2*). Todos los genes con desequilibrio mitonuclear detectados correspondieron a elementos de la cadena transportadora de electrones y a elementos que componen la maquinaria transcripcional mitocondrial.

6.4.2 Regiones con firmas de desequilibrio mitonuclear en superpoblaciones

En el caso de superpoblaciones, como fue discutido en el análisis de poblaciones, las principales regiones con desequilibrio mitonuclear compartidos entre superpoblaciones se localizaron en el NUMT HSA_NumtS_001_b1, correspondiente a pseudogenes de copias mitocondriales de subunidades del citocromo c (*MTCO1P12* y *MTCO3P12*), del complejo NADH mitocondrial (*MTND2P28*) y de la ATP sintasa (*MTATP6*). Los SNP correspondieron asimismo a otros NUMT reportados fuera del cromosoma 1, siendo los siguientes los que presentaron desequilibrio mitonuclear compartidos entre las superpoblaciones:

>**HSA_NumtS_001_b1**: Localizado en el cromosoma 1 de la posición 564 464 a la 570 304. Con un total de 407 SNP nucleares totales.

>**HSA_NumtS_472_b1**: Localizado en el cromosoma 13 de la posición 110 076 472 a la 110 076 727. Con un total de 4 SNP nucleares totales.

> **HSA_NumtS_513_b1**: Localizado en el cromosoma 17 de la posición 51 183 094 a la 51 183 746. Con un total de 12 SNP nucleares totales.

Al igual que en el análisis de poblaciones individuales, en las superpoblaciones, los NUMT representaron un porcentaje significativo de SNP únicos correspondientes a genes, el cual fue de 40%. Dado que este fenómeno se repitió en las mismas regiones en ambos análisis, se observó más a fondo el estado actual de estas regiones dentro del panel de 1000 genomas para tratar de diferenciar entre un posible error inherente a la base de datos y un fenómeno biológico real (ver sección 6.7).

El pseudogen **AL669831.3**, fue nuevamente el que más firmas de desequilibrio compartido entre superpoblaciones obtuvo. Asimismo, el gen **SGCZ** presentó firmas de desequilibrio mitonuclear entre las poblaciones africanas y surasiáticas. Dicho gen codifica para la proteína Sarcoglicano zeta, que forma parte del complejo de sarcoglicano, compuesto por proteínas transmembranales, y que a su vez es parte del complejo de la glucoproteína distrofina (DGC, por sus siglas en inglés), involucrada en mediar la estabilidad de la membrana muscular, y asociada con la distrofia muscular (Wheeler, Zarnegar, & McNally, 2002).

Los siguientes genes presentaron firmas de desequilibrio mitonuclear propios de cada población, y a su vez actividad reportada de interacción con mitocondria:

ERG2. Gen con desequilibrio mitonuclear identificado en la superpoblación americana “*admixed*”, el cual codifica para un factor de transcripción. Este gen se ha asociado a la enfermedad neurológica degenerativa de Charcot-Marie-Tooth (Pruitt *et al.*, 2014). Un estudio ha asociado la expresión de *EGR2* con la inducción de apoptosis; esto fue observado en varios cultivos celulares de cáncer, en donde la permeabilidad de la mitocondria se ve alterada liberando el citocromo c y activando las caspasas -3, -8 y -9. Asimismo se observó que *ERG2* induce la expresión de proteínas proapoptóticas (Unoki & Nakamura, 2003).

CTNND2. Gen con desequilibrio mitonuclear identificado en la superpoblación de Asia del este, que codifica para una proteína llamada delta-catenina, la cual tiene un papel de adhesión celular y está involucrada en el desarrollo de cáncer de ojos y cerebro (van Rootselaar *et al.*, 2017).

Se identificaron dos ARN intergénicos largos no codificantes con desequilibrio mitonuclear: **LINC01520**, con firmas de desequilibrio mitonuclear en la superpoblación americana “*admixed*” y **LINC024492**, con firmas en la superpoblación surasiática.

Los principales genes mitocondriales con mayor número de SNP con desequilibrio mitonuclear en más de una superpoblación, corresponden a subunidades de NADH:ubiquinona oxidoreductasa (*MT-NAD1*, *MT-NAD4*, *MT-NAD5*, *MT-NAD2*), ATPasa (*MT-ATP61* y *MT-ATP8*), una subunidad de citocromo c oxidasa (*MT-CO1*), el sARN 12s (*MT-RNR1*), el tARN mitocondrial correspondiente a treonina (*MT-TT*) y citocromo b oxidasa (*MT-CYB*). Los cuales coinciden en su mayoría con los genes mitocondriales con mayor desequilibrio mitonuclear identificados en el análisis de poblaciones individuales. En donde los genes son los correspondientes a la ruta de fosforilación oxidativa y transporte de electrones.

Como se observa en la Figura 46, la densidad de SNP totales es mayor entre *MT-CYTB* y *MT-ND4*, lo que no coincide con los genes con más desequilibrio compartido entre superpoblaciones, siendo que *MT-CYYB* y *MT-ND4* poseen el mayor número de SNP nucleares asociados a un SNP mitocondrial.

6.5 Comparación de resultados de poblaciones y superpoblaciones

La notable diferencia entre ambos análisis es el número de SNP bajo desequilibrio mitonuclear detectado. En el análisis utilizando las poblaciones individuales se tuvo una media de 121 SNP con desequilibrio mitonuclear y en comparación, cuando se utilizaron las superpoblaciones, se obtuvo una media de 257 SNP con desequilibrio mitonuclear.

Las posiciones de SNP nucleares entre ambos análisis coinciden de manera única en SNP localizados en el NUMT HSA_NumtS_001_b1. En cuanto a los SNP mitocondriales un total de 120 SNP coinciden en ambos análisis, los cuales corresponden a los genes descritos en los resultados en las tablas 28 y 40.

Por otra parte, se observó que no existieron coincidencias en SNP nucleares entre poblaciones y superpoblaciones, salvo los SNP localizados en el NUMT HSA_NumtS_001_b1. Sin embargo, a pesar de no tratarse de la misma posición de SNP nucleares, sí existió coincidencia entre regiones, como se describió en las tablas 26 y 38. En cuanto a genes mitocondriales, sí existieron coincidencias en regiones génicas, que corresponden también a los genes descritos en las tablas 28 y 40.

Dada la concentración de desequilibrio y de los SNP compartidos localizados en el NUMT HSA_NumtS_001_b1, se optó por revisar a detalle esta región (ver sección 6.7) y poder descartar que este comportamiento obtenido no sea por algún artefacto.

El hecho de que no existan SNP compartidos entre poblaciones y superpoblaciones, incluyendo a las poblaciones individuales comparadas con la superpoblación en la que se agrupan, podría dar lugar a pensar que este fenómeno de desequilibrio mitonuclear es muy propio y característico de cada población a nivel de SNP, pero que sin embargo se llega a compartir entre poblaciones y superpoblaciones en los mismos genes, lo cual podría llegar significar que la coevolución entre genomas actúa sobre los genes, cuyas variantes compensatorias se distinguen entre poblaciones, y de existir una divergencia mayor entre estas poblaciones se esperaría ver una diferencia significativa en las frecuencias de dichas variantes entre las poblaciones. Por lo que el fenómeno puede presentar características únicas de desequilibrio en cada población, pero involucrase en las mismas interacciones genéticas entre núcleo y mitocondria.

Para observar con mayor resolución qué tan específico es este desequilibrio de cada población, a nivel de los SNP, haría falta realizar el análisis con cada una de las poblaciones restantes de la base de datos de 10000 genomas y observar de igual manera los SNP y genes compartidos con desequilibrio mitonuclear.

Considerando el hecho de que no se encontraron asociaciones perfectas en el análisis utilizando superpoblaciones, podría llegarse a formular que las asociaciones entre alelos nucleares y mitocondriales se detectan con mayor asociación en poblaciones con un menor número de individuos, y que por otra parte dado el incremento de número de individuos en superpoblaciones podría tener como efecto la improbabilidad de encontrar variantes con menos de tres o menos individuos variando, lo cual tiene consecuencia no detectar variantes perfectas, que en contraste al seleccionar una población específica de la superpoblación, al ser un número menor de individuos, las asociaciones perfectas tiene mayor probabilidad de suceder. Cabe mencionar que el fenómeno previamente descrito no afecta la significancia estadística, ya que se ha considerado el escenario de asociaciones perfectas de manera independiente a las asociaciones altas, por lo que la selección de SNP con desequilibrio se hace bajo parámetros basados en el número de individuos analizados y el tipo de asociación.

Por otra parte, esto podría significar que la manera en cómo se estableció significancia estadística en asociaciones perfectas y altas, no fue adecuada y los resultados por azar no fueron descartados de manera efectiva, por lo que para probar este punto se necesitaría simular con parámetros más representativos de la variación de cada población, diferentes a los propuestos. Por ejemplo, tomar en cuenta las probabilidades de ocurrencia de cada base en los SNP de las poblaciones. Asimismo, se podría replantear la fórmula [10] basado en un mayor número de escenarios, para obtener de manera más efectiva el número esperado de resultados con asociaciones perfectas por azar.

Cabe mencionar que en las Figuras 32, 33, 42 y 43 se observa que a pesar de que se obtuvieron un mayor número de SNP con desequilibrio mitonuclear en el análisis utilizando superpoblaciones, estos se encuentran menos dispersos en comparación con el análisis realizado con poblaciones individuales, dando una mayor justificación a la problemática planteada con anterioridad, sobre si el análisis utilizando poblaciones individuales provee una mayor resolución característica de cada población o si se trata de un fenómeno de falta de significancia estadística. Asimismo, se debe considerar que en el análisis de superpoblaciones no se utilizó el máximo simulado como punto de corte de *tau* a diferencia del análisis de poblaciones, dado que el punto de corte en el análisis de superpoblaciones fue menor al máximo simulado, este podría repercutir en la densidad de los SNP con desequilibrio mitonuclear

6.6 Contexto global poblacional de frecuencias alélicas en los SNP bajo desequilibrio mitonuclear

Logramos detectar más de doscientas parejas de SNP bajo desequilibrio mitonuclear en las superpoblaciones de 1000 genomas con más de 400 individuos en cada grupo. En su totalidad estos SNP resultaron específicos de cada superpoblación y bajo un modelo de coevolución, última responsable del desequilibrio mitonuclear, se esperaría que estos SNP incrementasen su frecuencia en esas poblaciones específicas para las cuales resultaron con desequilibrio. Para ver hasta qué grado se cumple esta predicción, procedimos a observar las frecuencias que presentan en todas las superpoblaciones los SNP con desequilibrio de determinada superpoblación, comparándola con sus respectivos valores de *tau*. Se esperaría que los SNP con valores altos de *tau* en una superpoblación, también presentaran frecuencias mayores en esa superpoblación, en comparación con las demás superpoblaciones.

Para facilitar la comparación, se prepararon *heatmaps* con las frecuencias alélicas por un lado y con los valores de *tau* por otro, para los SNP con desequilibrio de cada superpoblación, como se muestra en las Figuras siguientes.

Superpoblación americana mestiza:

En esta superpoblación se observa que, con excepción de la población africana, los SNP específicos de los americanos mestizos sí presentan una frecuencia mayor al resto, asimismo se observa que gran parte de los SNP de la superpoblación mestiza americana se encuentran en una frecuencia igual o mayor en la superpoblación africana. A pesar de que las frecuencias de dichos alelos no fueron específicas totalmente de mestizos americanos, las asociaciones de dichos SNP con variantes del genoma mitocondrial sí fueron específicas de mestizos americanos.

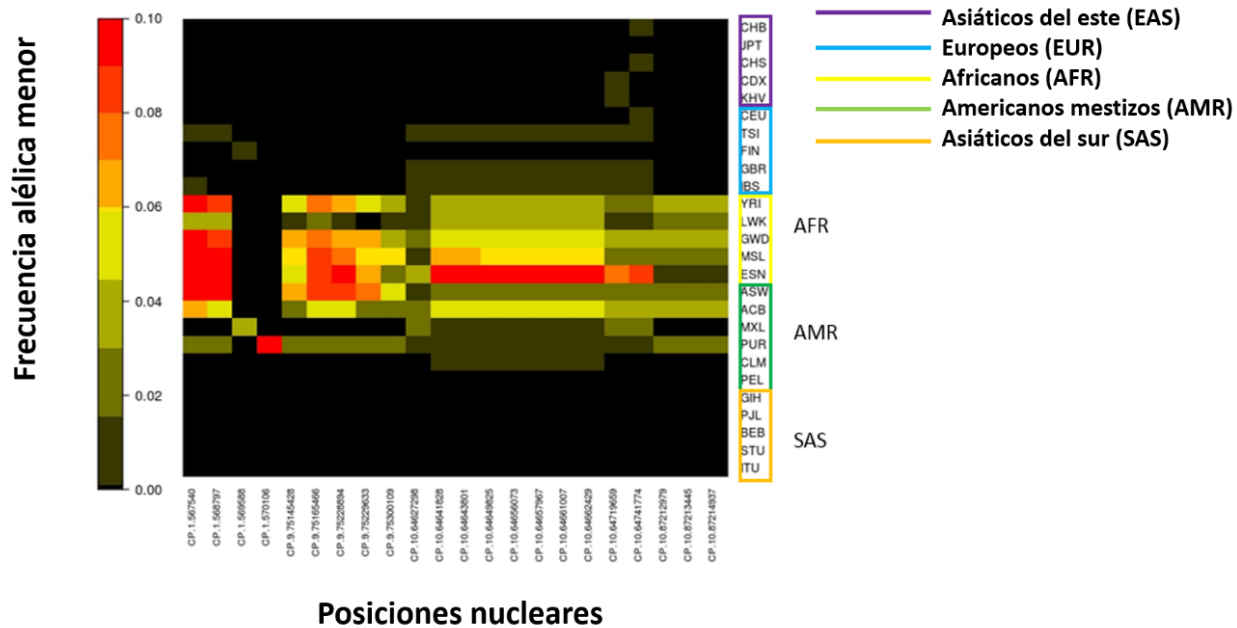
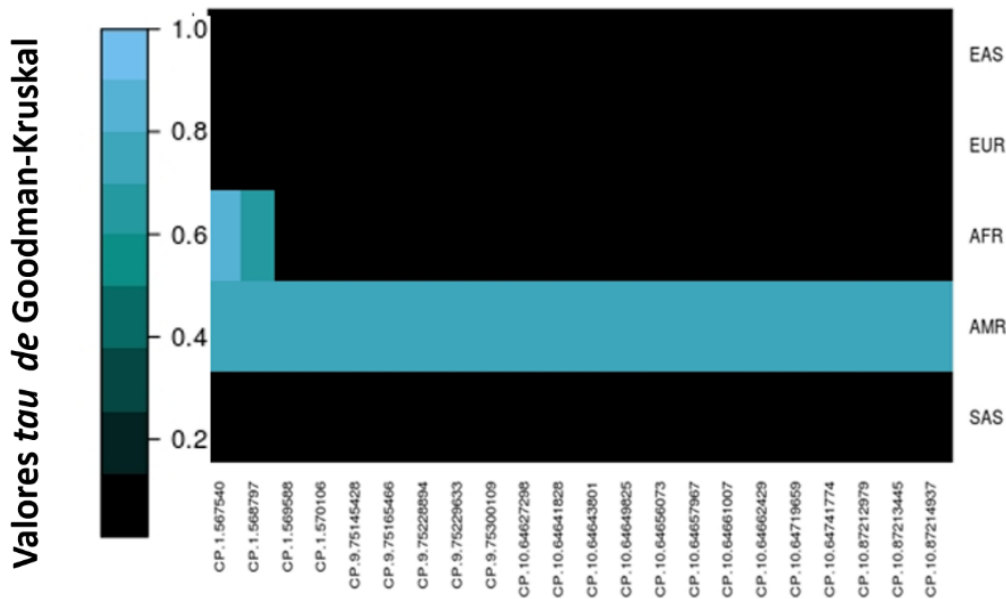


Figura 46. Heatmap de las frecuencias alélicas mínimas en distintas superpoblaciones para los SNP que presentaron desequilibrio mitonuclear en la superpoblación americana mestiza.



Posiciones nucleares

Figura 47. Heatmap de valores de tau de los SNP con desequilibrio mitonuclear en la superpoblación americana mestiza.

Superpoblación africana:

En esta superpoblación se observa que los SNP específicos se encuentran representados en las poblaciones africanas que conforman el bloque de superpoblación, asimismo se observan frecuencias menores de dichos SNP en todas las poblaciones de 1000 genomas. Las asociaciones altas de SNP nucleares y mitocondriales se ven representadas en la superpoblación africana.

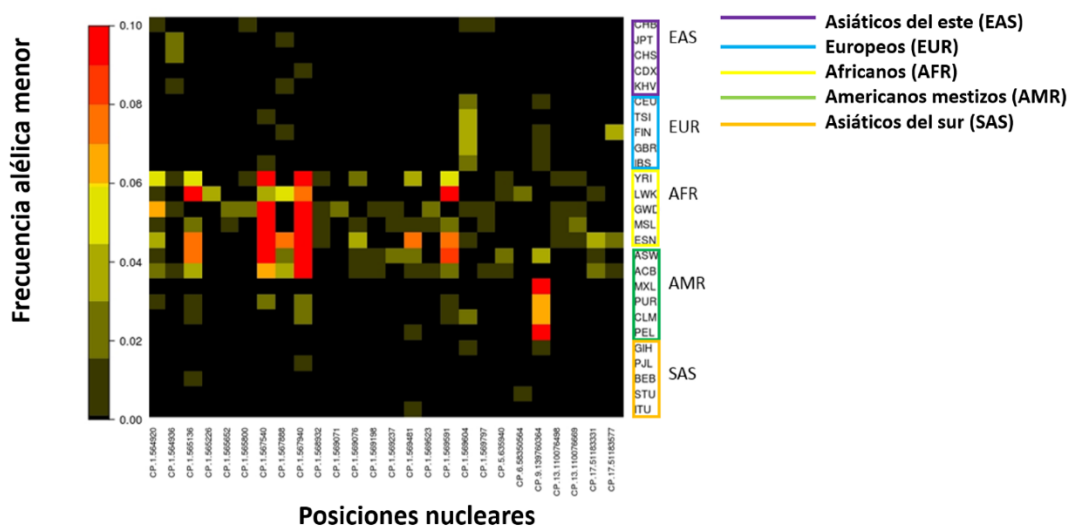


Figura 48. Heatmap de las frecuencias alélicas mínimas en distintas superpoblaciones para los SNP que presentaron desequilibrio mitonuclear en la superpoblación africana.

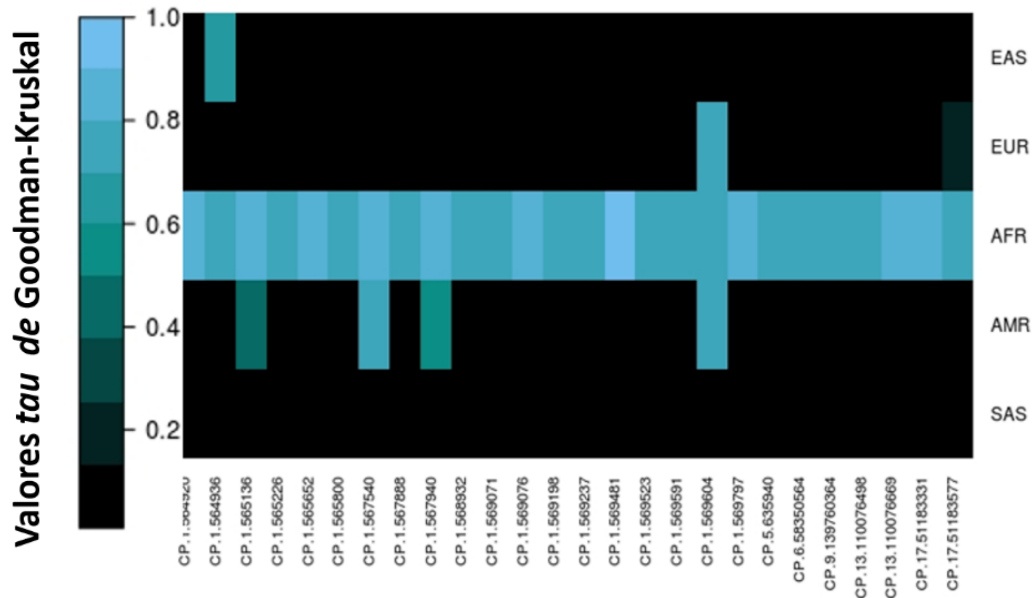


Figura 49. Heatmap de valores de tau de los SNP con desequilibrio mitonuclear en la superpoblación africana.

Superpoblación asiática del este:

En la superpoblación asiática del este se observa que las frecuencias de SNP con desequilibrio mitonuclear se encuentran en mayor frecuencia, siendo los asiáticos del este una de las poblaciones en la que se observa con más definición estas frecuencias propias de la superpoblación y su relación con el valor de alto de *tau* obtenido, seguida de la población de Asia de sur.

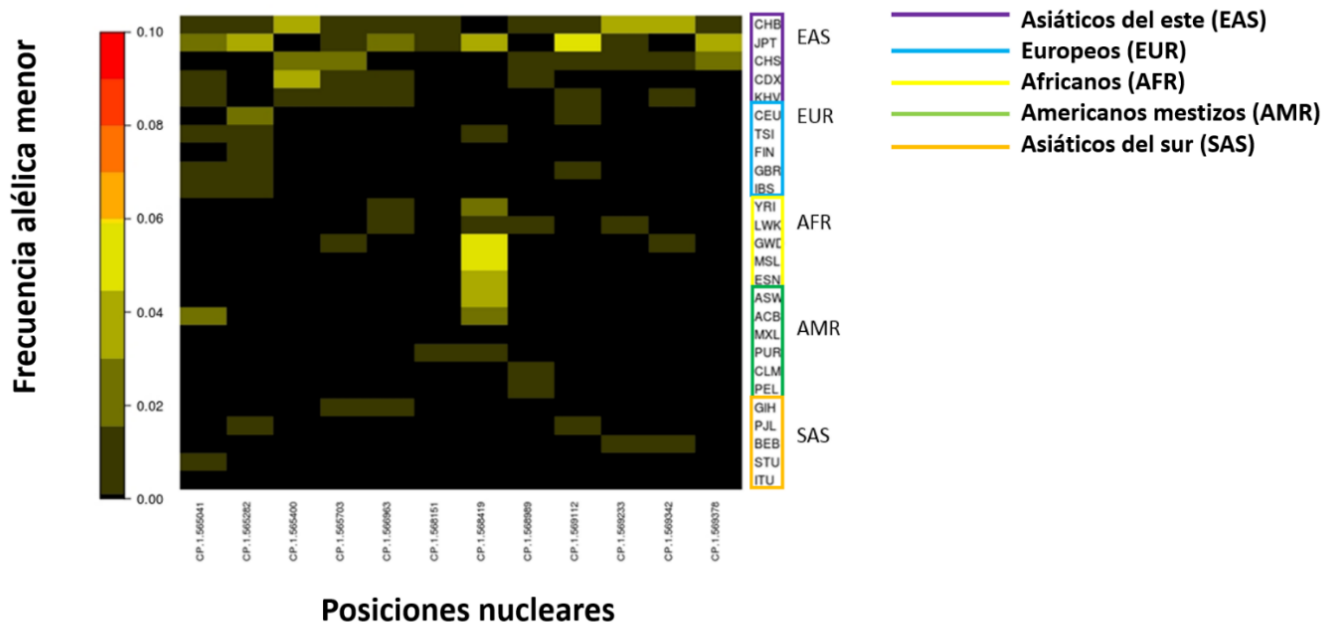


Figura 50. Heatmap de las frecuencias alélicas mínimas en distintas superpoblaciones para los SNP que presentaron desequilibrio mitonuclear en la superpoblación de Asia del este.

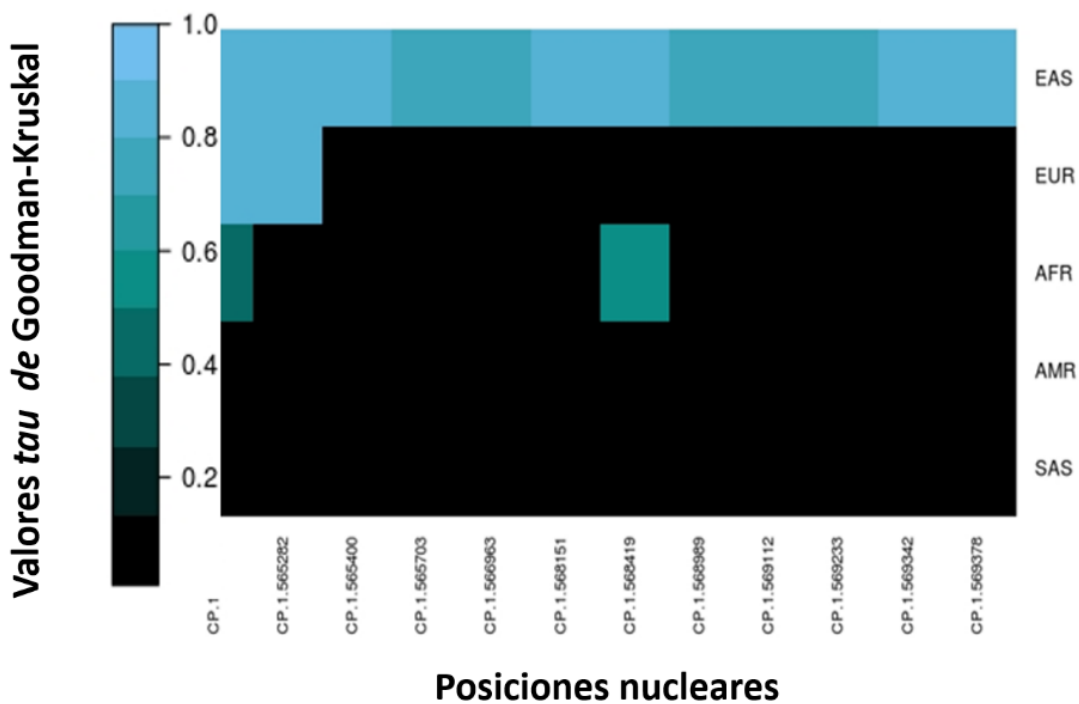


Figura 51. Heatmap de valores de tau de los SNP con desequilibrio mitonuclear en la superpoblación asiática del sur.

Superpoblación europea:

En la población europea, al igual que en las superpoblaciones descritas hasta el momento, se logra apreciar un patrón de frecuencias de SNP un poco más resaltado en su bloque de poblaciones correspondiente, sin embargo no es tan notorio, y se observa que comparte frecuencias con otras superpoblaciones del panel, al igual que los valores altos de τ , en el que existen SNP con asociaciones altas en más superpoblaciones.

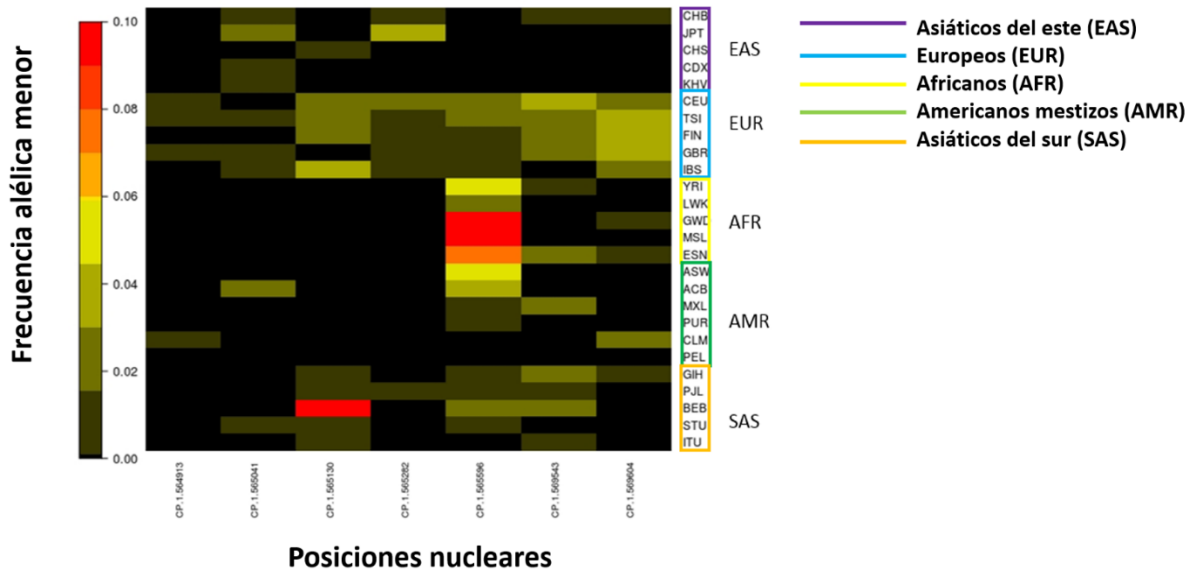


Figura 52. Heatmap de las frecuencias alélicas mínimas en distintas superpoblaciones para los SNP que presentaron desequilibrio mitonuclear en la superpoblación europea.

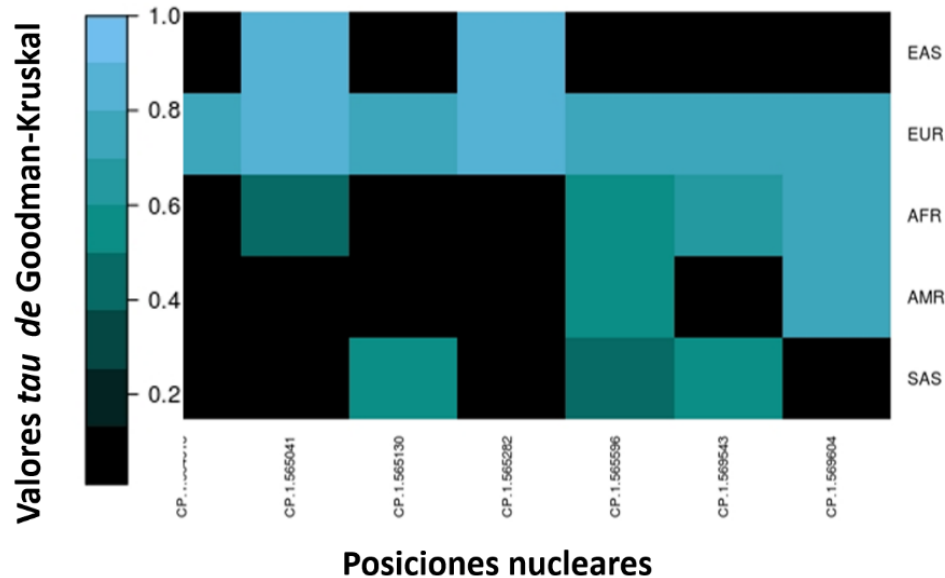


Figura 53. Heatmap de valores de tau de los SNP con desequilibrio mitonuclear en la superpoblación europea.

Superpoblación surasiática:

La población surasiática presentó el patrón más notorio de frecuencias altas de SNP con desequilibrio mitonuclear en su bloque poblacional, donde se observa que comparte pocas frecuencias de SNP con el resto de la población, de igual manera sucede con los valores de τ , donde solo se comparte un SNP con desequilibrio con un τ mayor a 0.5

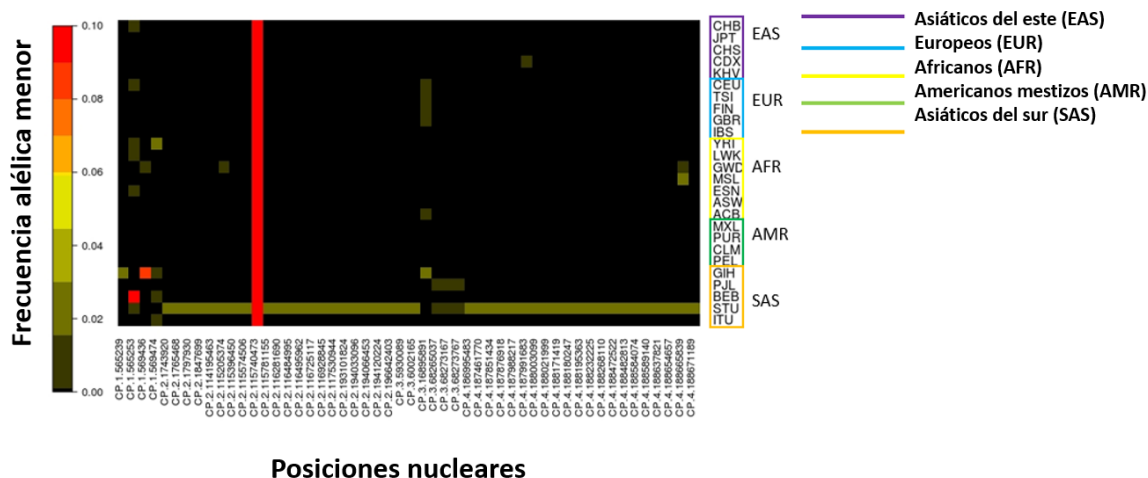


Figura 54. Heatmap de las frecuencias alélicas mínimas en distintas superpoblaciones para los SNP que presentaron desequilibrio mitonuclear en la superpoblación surasiática.

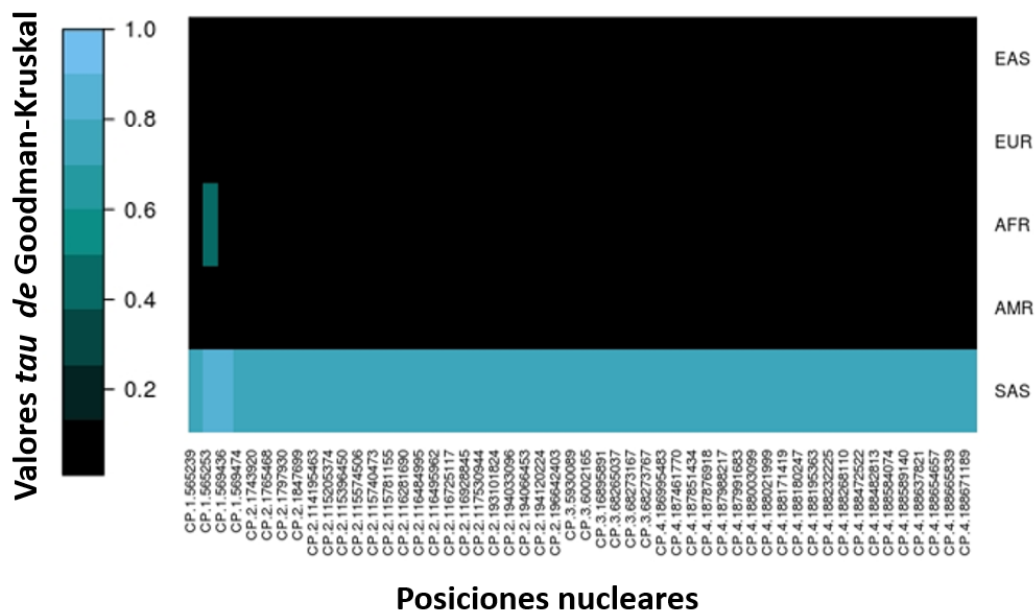


Figura 55. Heatmap de valores de tau de los SNP con desequilibrio mitonuclear en la superpoblación surasiática.

En los *heatmaps* de las Figuras 46 a 55 podemos ver una sobrerrepresentación de la frecuencia alélica menor, en los SNP nucleares con desequilibrio mitonuclear en cada superpoblación. Si bien estos SNP no son exclusivos de cada superpoblación, sí se encuentran en mayor frecuencia en las superpoblaciones en donde se logra detectar el desequilibrio mitonuclear. Esto podría ser un indicio de que en estas superpoblaciones estos SNP están incrementando su frecuencia por coadaptación con variantes mitocondriales específicas, manteniendo un ajuste sutil de la función mitocondrial. El hecho de que no estén en mayor frecuencia o fijados podría explicarse por el poco tiempo de divergencia de las poblaciones humanas. Por otra parte, el hecho de que se encuentren en bajas frecuencias a nivel mundial, indica que el sustrato de la variación con potencial coadaptativo lo constituye la variación existente y que en pocos casos, si alguno, han surgido *de novo* (por mutación reciente). Aunque queda abierta la posibilidad de que esos casos no sea posible detectarlos con estos números de muestra por estar en muy baja frecuencia.

Asimismo, con la comparación entre los gráficos de frecuencia alélica menor y de *tau*, podemos concluir que todas las poblaciones tienen sus SNP nucleares con desequilibrio mitonuclear único. Sin embargo, los genotipos en esos SNP están presentes en todas las poblaciones en una frecuencia menor, lo cual apoyaría a lo antes mencionado de una posible coadaptación entre ambos genomas, dado el ingreso de una variante en el genoma nuclear existente en otra población.

La población con un patrón menos claro de valores de *tau* y frecuencias alélicas menores fueron los europeos. Esta población fue la que presentó el menor número de SNP bajo desequilibrio. Por el contrario, las poblaciones africanas y surasiáticas fueron las que presentaron un mayor número de SNP en desequilibrio mitonuclear y una correlación característica entre la frecuencia de alelos mínimos y el desequilibrio mitonuclear detectado.

Los patrones de frecuencia de alelos menores encontrados en la población africana y su relación con todas las poblaciones podrían explicarse por la historia de migración de las poblaciones humanas como ya fue anteriormente discutido en la sección 6.3, ya que podemos observar que todos los alelos con desequilibrio mitonuclear en cada población están presentes en la población africana, sin embargo no bajo desequilibrio. Lo que podría llevar al planteamiento de una coevolución entre el genoma nuclear y mitocondrial independiente entre poblaciones dado un mismo genotipo nuclear, y que el sustrato de variación para la coadaptación es la previamente existente, sugiriendo que dichas variaciones surgieron en la población africana y posteriormente se dispersaron

De igual manera, pueden existir regiones con alta asociación entre alelos nucleares y mitocondriales, tan fuerte, que cualquier variación podría ser letal o perjudicial y que por efectos de selección no han ingresado variantes, en este caso, el presente análisis sería incapaz de detectar dichas regiones, ya que es necesaria la variación en dichas regiones para poder aplicar detectar asociaciones utilizando *tau* por lo que

se plantea utilizar información genética de individuos con alguna enfermedad asociada directamente a regiones de interacción núcleo-mitocondria y considerando que aquellos individuos poseen variación en esa regiones, sería posible detectarlas con el análisis descrito,

6.7 Inserciones mitocondriales en núcleo (NUMT)

En el genoma nuclear, las firmas de desequilibrio mitonuclear se concentraron en inserciones de fragmentos mitocondriales en núcleo (NUMT), específicamente en el NUMT HSA_NumtS_001_b1, que se encuentra en la posición 564 464 a 570 304 del cromosoma 1, y en donde se encuentran diversos pseudogenes, copias de genes mitocondriales. Lo cual nos lleva a cuestionar la relevancia biológica de esto, o si se trata de algún error o artefacto.

Al realizar un alineamiento del pseudogen MTATP61 contra su versión mitocondrial *MT-ATP6*, se observa que en ambos existen SNP con desequilibrio mitonuclear, los cuales están localizados en la misma posición y corresponden al mismo aminoácido (Figura 59).

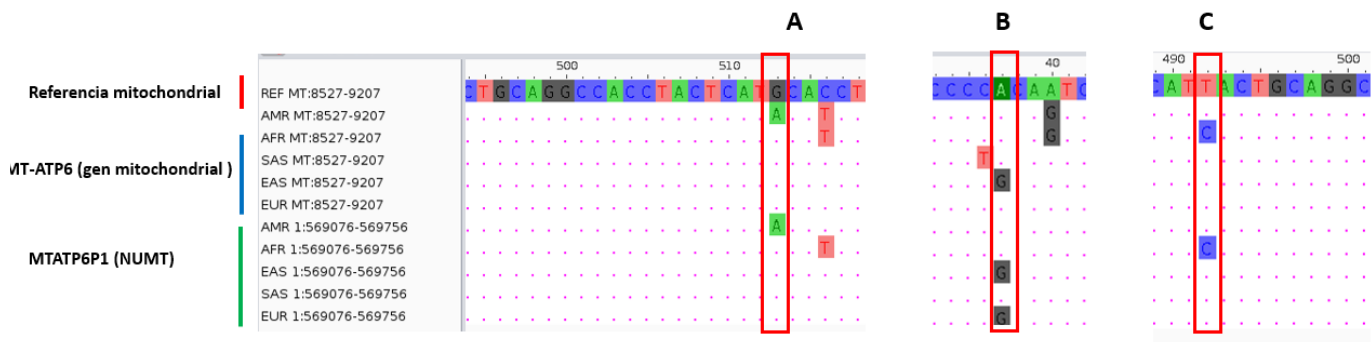


Figura 56. Alineamiento del pseudogen MTATP6P1 y el gen MT-ATP6. Las letras A, B y C indican los SNP con desequilibrio mitonuclear encontrados.

En la Figura 56 podemos también apreciar que los SNP nucleares y mitocondriales identificados con desequilibrio mitonuclear en el gen y pseudogen son específicos de alguna superpoblación. Por ejemplo, el SNP con desequilibrio mitonuclear marcado con la letra A solo se encuentra en la superpoblación americana (AMR) y de igual manera sucede con el SNP marcado con la letra C, el cual es único de la superpoblación de África (AFR). Sin embargo, existen SNP, como el marcado con la letra B, en donde la variante no es única de una superpoblación, como en este ejemplo, en el que el desequilibrio se identificó en la superpoblación de Asia del este, lo cual podría significar que esa variable está presente en otras poblaciones debido a efectos migratorios.

El observar que las variantes son exactamente las mismas en gen y pseudogen, dio luz a considerar que este resultado se debiese a algún tipo de error, ya sea por secuenciación de fragmentos mitocondriales de las muestras que posteriormente fuesen alienados al genoma de referencia generando así un falso SNP, o por un posible error en el ensamblado original del genoma de referencia.

Se generó una gráfica representando con colores rojo (base A) y negro (base B) las regiones identificadas bajo desequilibrio en población africana en el pseudogen *MTAP6P1* y su alineamiento contra su contraparte mitocondrial *MT-ATP6*, con la finalidad de observar si en todas las regiones se trata de la misma variante en ambos genomas.

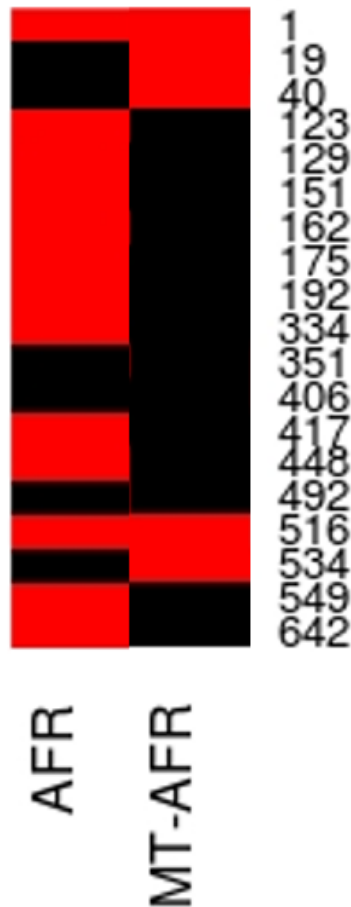


Figura 57. Esquema representando la presencia de dos bases en las posiciones del pseudogen *MTATP6P1* y el gen *MT-ATP6* con desequilibrio mitonuclear identificado en la superpoblación africana.

En la Figura 57 observamos que no en todas las parejas de SNP bajo desequilibrio se trata de la misma base en ambos genomas, ya que en las columnas de MT-AFR y AFR, indicando respectivamente gen y pseudogen, no se observa el mismo color ya sea negro y rojo como lo son por ejemplo las posiciones 19 y 40 y muchas otras.

No obstante, sí se observan casos en los que se trata de la misma base como es el caso de las posiciones 1 y 516. Llegando a concluir que puede no tratarse de exactamente el mismo fragmento el cual habría sido mal alineado en la muestra.

Por otra parte, se observó en el llamado de variantes, cuántas lecturas alineaban en la región del NUMT HSA_NumtS_001_b1., con la finalidad de identificar alguna peculiaridad.

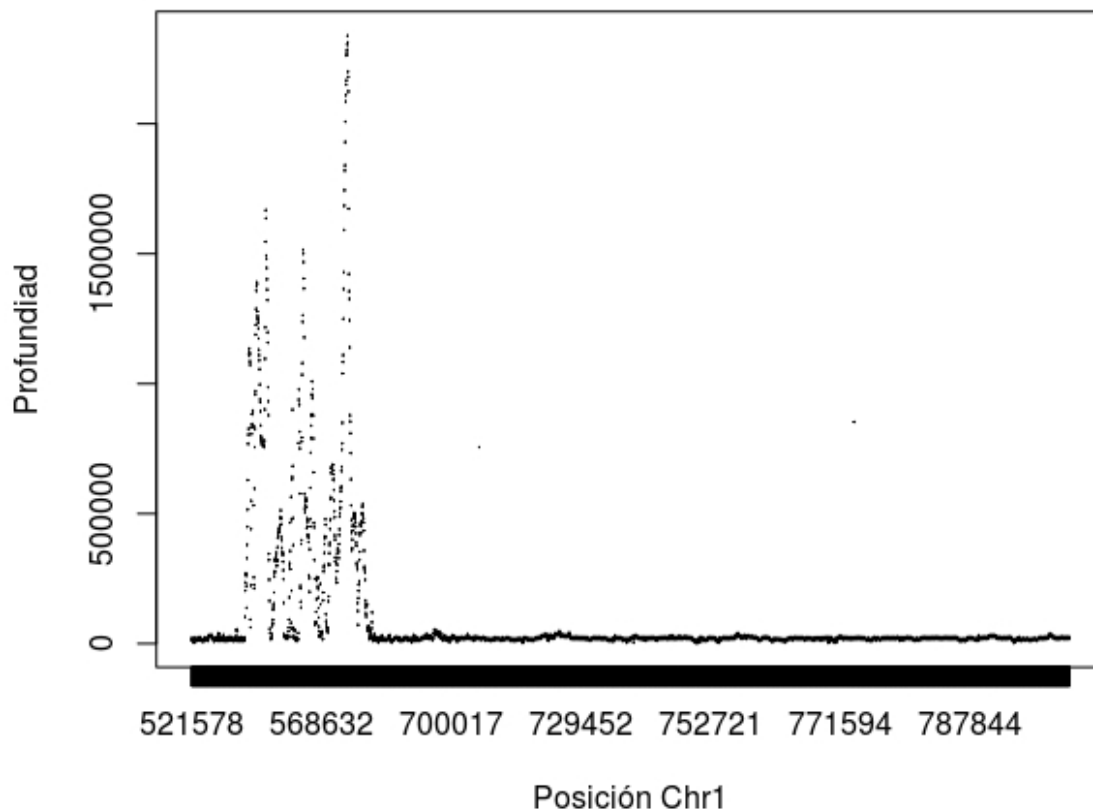


Figura 58. Profundidad de alineamiento en la región de 266,266 pares de bases del cromosoma 1 correspondiente a la base de datos de 1000 genomas

Se graficó la profundidad de alineamiento en una región de 266 266 pares de bases del cromosoma 1, la cual comprendía la región del NUMT y fue posible observar una clara elevación de la profundidad en una magnitud de más de 1 000 000× con respecto al resto de la región (Figura 61), indicando la posibilidad de que hayan sido alienadas secuencias del ADN mitocondrial de manera errónea en esta región, dando como resultado errores en el llamado de los SNP nucleares. Esto puede deberse a que el mtADN se encuentra en mayor número de copias en la célula en

relación al ADN nuclear y es difícil eliminarlo, representando un problema bien conocido en genómica (Thalman, Hebler, Poinar, Pääbo, & Vigilant, 2004).

Por otro lado, un error en el ensamblado del genoma de referencia surgiría si cada uno de estos NUMT existiera como elemento repetitivo en tándem, que al ser idénticos no pudiesen ser identificados como distintos, colapsándolos en el ensamblado en una única copia. Esto produciría una sobrerrepresentación de fragmentos de secuenciación en el alineamiento en estas zonas, es decir, resultaría en un aumento local de la profundidad de secuenciación.

Para observar si este fenómeno era generalizado, graficamos las profundidades de todos los NUMT reportados hasta el momento (Figura 59) en el genoma de humano.

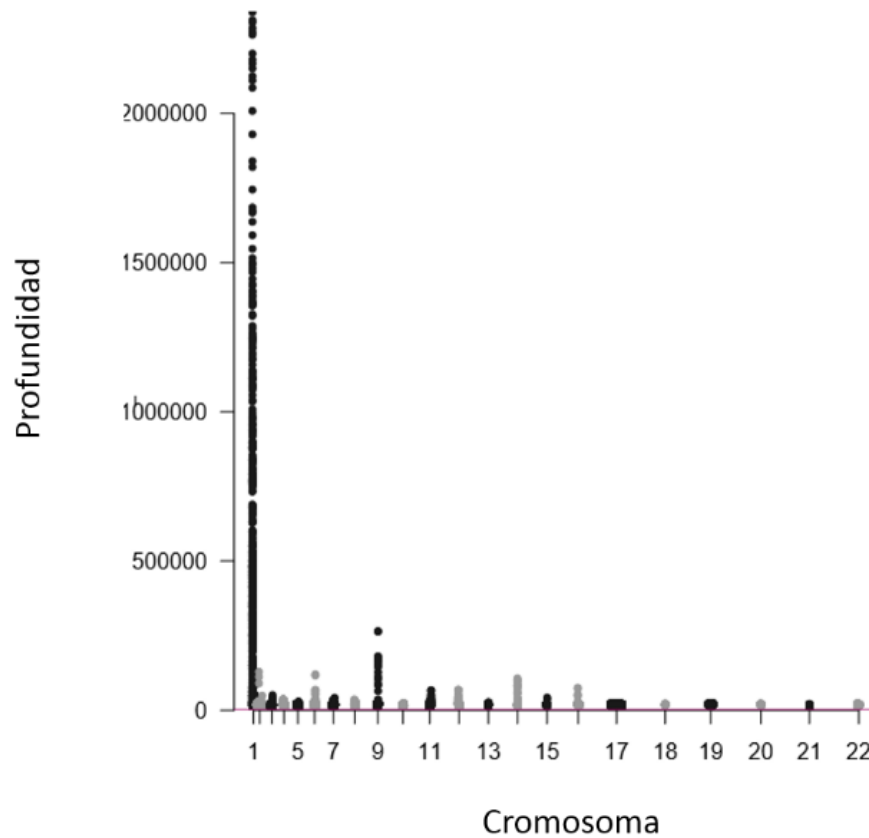


Figura 59. Profundidades de los NUMT por cromosoma en el genoma humano.

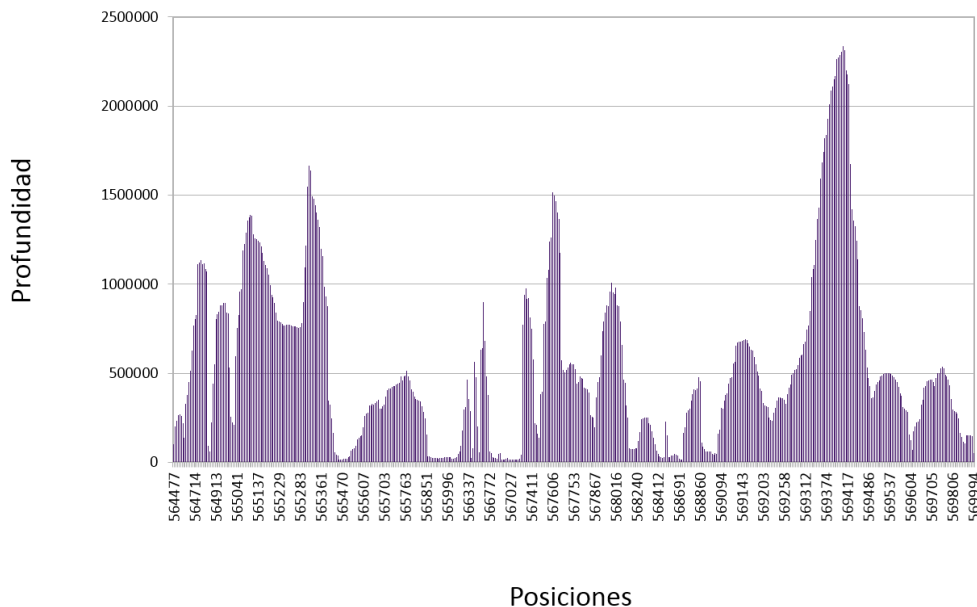


Figura 60. Profundidades locales a lo largo del NUMT HSA_NumtS_001_b1.

Como puede observarse, el NUMT ubicado en el cromosoma 1 presenta un nivel de profundidad único, pues la profundidad del resto de los NUMT difiere en una magnitud de $1\ 000\ 000\times$. Esto excluye un error general en el ensamblado de los NUMT en el genoma de referencia, y quizá también la posibilidad de que se esté mapeando ADN mitocondrial en lugar de nuclear de manera general, ya que si ese fuera el caso, todos los NUMT deberían presentar una profundidad alta y similar, lo cual no sucede.

Viendo a detalle la región correspondiente al NUMT HSA_NumtS_001_b1, en la Figura 63, observamos que la profundidad tampoco es homogénea.

Por lo que se hipotetiza que dichas inserciones mitocondriales pueden encontrarse de manera repetida en el genoma nuclear, siendo de un tamaño menor al NUMT HSA_NumtS_001_b1. Al ser estas regiones repetidas de tamaño menor a 5 kb pudieron haberse perdido durante la realización del ensamblado del genoma de referencia, error que ya ha sido reportado en estas regiones repetidas (Salzberg & Yorke, 2005), y al no ser identificadas en el ensamblado, las secuencias de ADN correspondiente a las regiones repetidas se alinearon en una única región.

Para poder generar un planteamiento de una posible relevancia funcional en la interacción núcleo-mitocondria en esta inserción, es primordial descartar posibles errores en el ensamblado. Para averiguar si dichas regiones repetidas se

encuentran en *tándem*, se propone realizar una secuenciación dirigida a la región NUMT y buscar de manera bioinformática dichos repetidos y visualizarlos utilizando el algoritmo propuesto por Buchner (2003), el cual se basa en la transformación de periodicidad, en donde se detectan caracteres recurrentes dado un intervalo (Buchner & Janjarasjitt, 2003).

Por otra parte, en la publicación de Ricchetti y col. (2004) se identificaron los NUMT específicos de humanos, pero no se menciona esta inserción en el cromosoma uno. Lo cual genera discrepancia, ya que algunos SNP presentan variación idéntica entre la inserción y el gen mitocondrial, lo cual se esperaría de un evento de inserción muy reciente. No obstante, utilizando la herramienta *phyloP* (Pollard, Hubisz, Rosenbloom, & Siepel, 2010), la cual nos permite ver la conservación entre 100 vertebrados (Figura 64), y que forma parte de la paquetería de herramientas disponibles en el navegador de genomas UCSC (Kent *et al.*, 2002), se encontró que esta región es altamente conservada y ancestral a humanos, por lo cual sería necesaria una búsqueda de desequilibrio mitonuclear en esta región en otros organismos, para establecer si la posible relevancia funcional en la interacción núcleo-mitocondria que posee esta región se adquirió de manera reciente en humanos.

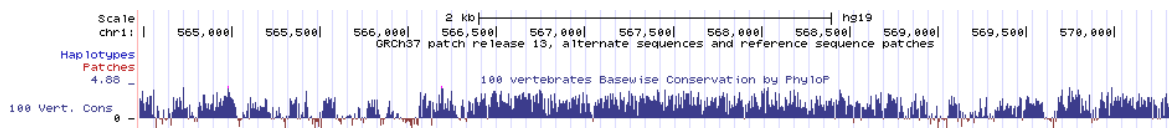


Figura 61. Visualización de la región NUMT HSA_NumtS_001_b1 y los resultados de PhyloP en donde las barras azules representan una alta conservación.

El parecido entre las secuencias de la inserción y del gen original, dan como resultado que se detecte una alta covariación y por tanto un desequilibrio mitonuclear. Dado el fenómeno expuesto en esta sección es de suma importancia considerar esta región en el cromosoma uno como una posible problemática para la búsqueda y el estudio de desequilibrio mitonuclear en el genoma humano en un futuro.

7 Conclusiones

La comparación por parejas de SNP entre el genoma nuclear y mitocondrial nos permitió realizar un análisis de alta resolución con una magnitud de 1×10^9 de asociaciones calculadas, siendo un análisis de mayor resolución en poblaciones humanas en comparación a los previamente realizados.

El estadístico *tau* resultó ser bastante sensible a asociaciones azarosas, más notorio con los números de muestra más bajos utilizados. Sin embargo, se lograron establecer los parámetros necesarios para detectar desequilibrio mitonuclear estadísticamente significativo, así como cuantificar y filtrar el efecto del azar en el análisis.

Fueron detectadas parejas de SNP con desequilibrio mitonuclear con significancia estadística, en donde las poblaciones y superpoblaciones de Asia del sur y africanas presentaron el mayor número de SNP con desequilibrio, hipotetizando un posible efecto coevolutivo dado su historia poblacional en el caso de la población africana y en el caso de los surasiáticos un posible efecto de asociaciones entre SNP nucleares y mitocondriales dado su estructura poblacional y altamente endogámica.

El análisis utilizando superpoblaciones obtuvo un mayor número de SNP con desequilibrio en comparación con el análisis realizado con poblaciones individuales, asimismo, la densidad de estos SNP en su ubicación genómica fue mayor en superpoblaciones.

Al comparar los resultados de poblaciones individuales y superpoblaciones, se hipotetiza que el análisis realizado con poblaciones individuales nos permite detectar desequilibrio más característico de cada población. Este efecto no parece ser debido a procesos demográficos, considerando que las frecuencias alélicas de los SNP con desequilibrio mitonuclear son mayores en las poblaciones donde se detectó el desequilibrio mitonuclear, lo que apunta más a un efecto coevolutivo.

Se identificaron SNP nucleares con desequilibrio mitonuclear en genes con evidencia de interacción indirecta entre núcleo y mitocondria, involucrados en biogénesis mitocondrial y metabolismo; asimismo, se detectó desequilibrio mitonuclear en genes mitocondriales pertenecientes a los complejos de la cadena transportadora de electrones, ARN de transferencia y ribosomal. Recientemente se han caracterizado funciones de mantenimiento y metabolismo del ADN mitocondrial realizadas por genes con funciones canónicas distintas (ver sección 1.1.6), abriendo así la posibilidad de que los genes detectados aquí, puedan tener también una doble función, que los involucre de manera más directa con el mtADN. Esta hipótesis queda abierta para futuros ensayos experimentales.

El mayor desequilibrio mitonuclear fue detectado en regiones de inserciones mitocondriales (NUMTS), específicamente en una inserción localizada en el

cromosoma uno. Con base en nuestros análisis, se concluye que podría tratarse de un posible error de ensamblado y cuya región debe tomarse muy en cuenta para el estudio de asociaciones entre alelos nucleares y mitocondriales.

8 Perspectivas

Se enlistan a continuación una serie de análisis y vías de investigación que ayudarían a esclarecer en mayor profundidad el fenómeno de desequilibrio mitonuclear y sus bases evolutivas y funcionales:

- Realizar el análisis con todas las poblaciones de la base de datos de 1000 genomas de manera individual.
- Simular el fenómeno de desequilibrio mitonuclear dada variación por azar en el genoma nuclear y mitocondrial en una población con parámetros más estrictos para descartar posibles errores de significancia estadística.
- Mediante secuenciación dirigida descartar errores de ensamblado por repetidos en la región NUMT que presentó alto desequilibrio mitonuclear.
- Analizar datos genómicos de individuos con enfermedades relacionadas a mitocondria y comparar humanos con especies cercanas.
- Realizar una validación de los resultados *in silico*. Se podrían realizar experimentos mediante el uso de híbridos en células humanas en donde se observen los posibles efectos de una incompatibilidad mitonuclear, al introducir combinaciones de alelos “no asociados” dado los SNP candidatos con desequilibrio mitonuclear.

Tablas suplementarias

Tabla suplementaria 1 Número total de parejas de SNP nucleares y mitocondriales por población

Población	Número total de parejas de SNP nucleares y mitocondriales
Residentes de Utah con ascendencia del Norte y Oeste de Europa (CEU)	4,691,650,656
Finlandeses en Finlandia (FIN)	1,332,291,638
Yorubas en Ibadan, Nigeria (YRI)	9,732,971,952
Gambianos en divisiones del oeste de Gambia (GWD)	3,592,186,062
Residentes de los Ángeles, con ascendencia mexicana (MXL)	3,674,471,920
Puertorriqueños de Puerto Rico (PUR)	1,700,921,208
Punjabi de Lahore, Pakistan (PJL)	1,091,872,700
Hablantes de Bengali de Bangladesh (BEB)	678,883,280

Tabla suplementaria 2 Porcentaje de parejas de SNP nucleares y mitocondriales con tau mayor a 0.1 en el análisis de poblaciones individuales

Población	Número total de parejas de SNP nucleares y mitocondriales	Combinaciones con <i>tau</i> > 0.1 (%)
Residentes de Utah con ascendencia del Norte y Oeste de Europa (CEU)	4,691,650,656	63,450,470 (1.35%)
Finlandeses en Finlandia (FIN)	1,332,291,638	2,959,385 (0.22%)
Yorubas en Ibadan, Nigeria (YRI)	6,525,423,168	153,183,610 (2.34%)
Gambianos en divisiones del oeste de Gambia (GWD)	3,592,235,062	28,194,546 (0.78%)
Residentes de los Ángeles, con ascendencia mexicana (MXL)	3,674,471,920	57,642,279 (1.56%)
Puertorriqueños de Puerto Rico (PUR)	1,700,921,208	8,098,095 (0.47%)
Punjabi de Lahore, Pakistan (PJI)	1,091,872,700	4,767,362 (0.43%)
Hablantes de Bengali de Bangladesh (BEB)	678,883,280	3,021,599 (0.44%)

Tabla suplementaria 3 Número de resultados de tau por rango en el análisis de poblaciones individuales.

Población	Valores de tau de 0.1-0.5	Valores de tau de 0.5-0.7	Valores de tau de 0.7-0.9	Valores de tau =1
Chinos Han en Beijing, China (CHB)	70,801,299	31,711	666	13,049,156
Japoneses en Tokyo, Japón (JPT)	56,735,868	33,867	612	6,425,479
Residentes de Utah con ancestría del Norte y Oeste de Europa (CEU)	63,408,082	41,593	723	72
Finlandeses en Finlandia (FIN)	2,955,667	3,010	708	1
Yorubas en Ibadan, Nigeria (YRI)	153,060,250	119,975	3,157	228
Gambianos en divisiones del oeste de Gambia (GWD)	28,174,836	18,917	773	20
Residentes de los Ángeles, con ancestría mexicana (MXL)	57,593,620	47,988	570	101
Puertorriqueños de Puerto Rico (PUR)	8,084,986	9,993	3,116	1
Punjabi de Lahore, Pakistan (PJL)	4,760,503	3,743	3,116	263
Hablantes de Bengalí de Bangladehs (BEB)	3,017,845	2,676	956	122

Tabla suplementaria 4 Número esperado de asociaciones perfectas debido a efectos aleatorios en distintas poblaciones

Población	Número total de parejas de SNP con tres individuos variando	Número de individuos	Número esperado de parejas de SNP nucleares y mitocondriales con correlación perfecta debida al azar con ≥ 3 individuos variando
Chinos Han en Beijing, China (CHB)	952,727	99	0.04
Japoneses en Tokyo, Japón (JPT)	928,267	99	0.04
Residentes de Utah con ascendencia del Norte y Oeste de Europa (CEU)	1,832,572	99	0.07
Finlandeses en Finlandia (FIN)	1,843,942	113	0.05
Yorubas en Ibadan, Nigeria (YRI)	1,025,603	64	0.14
Gambianos en divisiones del oeste de Gambia (GWD)	1,578,515	104	0.05
Residentes de los Ángeles, con ascendencia mexicana (MXL)	1,069,774	96	0.04
Puertorriqueños de Puerto Rico (PUR)	1,044,061	86	0.06
Punjabi from Lahore, Pakistan (PJL)	1,069,774	96	0.04
Bengali from Bangladesh (BEB)	1,044,061	86	0.06

Tabla suplementaria 5 Número total de asociaciones perfectas detectadas por población con más de tres individuos variando

Población	Número de SNP bajo desequilibrio con valor de $\tau = 1$ no atribuibles al azar
Residentes de Utah con ascendencia del Norte y Oeste de Europa (CEU)	72
Finlandeses en Finlandia (FIN)	1
Yorubas en Ibadan, Nigeria (YRI)	228
Gambianos en divisiones del oeste de Gambia (GWD)	20
Residentes de los Ángeles, con ascendencia mexicana (MXL)	101
Puertorriqueños de Puerto Rico (PUR)	0
Punjabi de Lahore, Pakistán (PJL)	263
Hablantes de Bengali de Bangladesh (BEB)	122

Tabla suplementaria 6 Resultados de simulaciones en poblaciones representativas

Población	Resultados prueba de K-S (D)	P-valor	Valor máximo de τ
Residentes de Utah con ascendencia del Norte y Oeste de Europa (CEU)	0.1296	$2.2 \cdot 10^{-16}$	0.791
Yorubas en Ibadan, Nigeria (YRI)	0.23	$2.2 \cdot 10^{-16}$	0.772
Residentes de los Ángeles, con ascendencia mexicana (MXL)	0.23	$2.2 \cdot 10^{-16}$	0.771
Punjabi de Lahore, Pakistán (PJL)	0.22	$2.2 \cdot 10^{-16}$	0.777

Tabla suplementaria 7 Número total de resultados significativos de valores de tau en cada población.

Población	<i>tau</i> = 1 (en los SNP con ≥ 3 individuos variando)	Valores de <i>tau</i> mayores al máximo valor simulado	Total
Chinos Han en Beijing, China (CHB)	37	14	51
Japoneses en Tokyo, Japón (JPT)	90	21	111
Residentes de Utah con ascendencia del Norte y Oeste de Europa (CEU)	72	1	73
Finlandeses en Finlandia (FIN)	0	1	1
Yorubas en Ibadan, Nigeria (YRI)	228	113	341
Gambianos en divisiones del oeste de Gambia (GWD)	20	82	102
Residentes de los Ángeles, con ascendencia mexicana (MXL)	101	12	113
Puertorriqueños de Puerto Rico (PUR)	0	22	22
Punjabi de Lahore, Pakistan (PJL)	263	4	267
Hablantes de Bengali de Bangladesh (BEB)	122	16	138

Literatura citada

- Aho, K., Alfons, A., An-deregg, N., Aragon, T., Arppe, A., Baddeley, A., ... Zeileis Maintainer Andri Signorell, A. (2018). DescTools. Retrieved from <https://cran.r-project.org/web/packages/DescTools/DescTools.pdf>
- Amunts, A., Brown, A., Toots, J., Scheres, S. H. W., & Ramakrishnan, V. (2015). Ribosome. The structure of the human mitochondrial ribosome. *Science (New York, N.Y.)*, *348*(6230), 95–98. <https://doi.org/10.1126/science.aaa1193>
- Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., ... Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. <https://doi.org/10.1038/nature15393>
- Brown, T. A. (2002). The Human Genome. Retrieved from <https://www.ncbi.nlm.nih.gov/books/NBK21134/>
- Buchner, M., & Janjarasjitt, S. (2003). Detection and visualization of tandem repeats in dna sequences. *IEEE Transactions on Signal Processing*, *51*(9), 2280–2287. <https://doi.org/10.1109/TSP.2003.815396>
- Burton, R. S., Pereira, R. J., & Barreto, F. S. (2013). Cytonuclear Genomic Interactions and Hybrid Breakdown. *Annual Review of Ecology, Evolution, and Systematics*, *44*(1), 281–302. <https://doi.org/10.1146/annurev-ecolsys-110512-135758>
- Case, A. L., & Willis, J. H. (2008). Hybrid male sterility in mimulus (*phrymaceae*) is associated with a geographically restricted mitochondrial rearrangement. *Evolution*, *62*(5), 1026–1039. <https://doi.org/10.1111/j.1558-5646.2008.00360.x>
- Chau, B. N., Cheng, E. H., Kerr, D. A., & Hardwick, J. M. (2000). Aven, a novel inhibitor of caspase activation, binds Bcl-xL and Apaf-1. *Molecular Cell*, *6*(1), 31–40. Retrieved from [http://doi.org/10.1016/S1097-2765\(05\)00021-3](http://doi.org/10.1016/S1097-2765(05)00021-3)
- Chen, J.-Q., Cammarata, P. R., Baines, C. P., & Yager, J. D. (2009). Regulation of mitochondrial respiratory chain biogenesis by estrogens/estrogen receptors and physiological, pathological and pharmacological implications. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1793*(10), 1540–1570. <https://doi.org/10.1016/J.BBAMCR.2009.06.001>
- Chen, X. J., Wang, X., Kaufman, B. A., & Butow, R. A. (2005). Aconitase couples metabolic regulation to mitochondrial DNA maintenance. *Science (New York, N.Y.)*, *307*(5710), 714–717. <https://doi.org/10.1126/science.1106391>
- Chou, J.-Y., & Leu, J.-Y. (2010). Speciation through cytonuclear incompatibility: Insights from yeast and implications for higher eukaryotes. *BioEssays*, *32*(5), 401–411. <https://doi.org/10.1002/bies.200900162>
- Chou, J.-Y., & Leu, J.-Y. (2015). The Red Queen in mitochondria: cyto-nuclear co-evolution, hybrid breakdown and human disease. *Frontiers in Genetics*, *6*, 187. <https://doi.org/10.3389/fgene.2015.00187>
- Clark, K. A., Howe, D. K., Gafner, K., Kusuma, D., Ping, S., Estes, S., & Denver, D. R. (2012). Selfish Little Circles: Transmission Bias and Evolution of Large Deletion-Bearing Mitochondrial DNA in *Caenorhabditis briggsae* Nematodes. *PLoS ONE*, *7*(7), e41433.

<https://doi.org/10.1371/journal.pone.0041433>

- Clay, K., & Kover, P. X. (1996). The red queen hypothesis and plant/pathogen interactions. *Annu. Rev. Phytopathol*, *34*, 29–50. Retrieved from <https://www.annualreviews.org/doi/pdf/10.1146/annurev.phyto.34.1.29>
- D’Esposito, M., Strazzullo, M., Cuccurese, M., Spalluto, C., Rocchi, M., D’Urso, M., & Ciccodicola, A. (1998). Identification and assignment of the human transient receptor potential channel 6 gene TRPC6 to chromosome 11q21–>q22. *Cytogenetics and Cell Genetics*, *83*(1–2), 46–47. <https://doi.org/10.1159/000015165>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... 1000 Genomes Project Analysis Group, 1000 Genomes Project Analysis. (2011). The variant call format and VCFtools. *Bioinformatics (Oxford, England)*, *27*(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Dhir, A., Dhir, S., Borowski, L. S., Jimenez, L., Teitell, M., Rötig, A., ... Proudfoot, N. J. (2018). Mitochondrial double-stranded RNA triggers antiviral signalling in humans. *Nature*, *560*(7717), 238–242. <https://doi.org/10.1038/s41586-018-0363-0>
- Du Mee, D. J. M., Ivanov, M., Parker, J. P., Buratowski, S., & Marquardt, S. (2018). Efficient termination of nuclear lncRNA transcription promotes mitochondrial genome maintenance. *ELife*, *7*. <https://doi.org/10.7554/eLife.31989>
- Duarte, F. V., Palmeira, C. M., & Rolo, A. P. (2014). The Role of microRNAs in Mitochondria: Small Players Acting Wide. *Genes*, *5*(4), 865–886. <https://doi.org/10.3390/genes5040865>
- Ellison, C. K., & Burton, R. S. (2008a). Genotype-dependent variation of mitochondrial transcriptional profiles in interpopulation hybrids. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(41), 15831–15836. <https://doi.org/10.1073/pnas.0804253105>
- Ellison, C. K., & Burton, R. S. (2008b). Interpopulation hybrid breakdown maps to the mitochondrial genome. *Evolution*, *62*(3), 631–638. <https://doi.org/10.1111/j.1558-5646.2007.00305.x>
- Estaquier, J., Vallette, F., Vayssiere, J.-L., & Mignotte, B. (2012). The Mitochondrial Pathways of Apoptosis (pp. 157–183). Springer, Dordrecht. https://doi.org/10.1007/978-94-007-2869-1_7
- Fields, P. D., McCauley, D. E., McAssey, E. V, & Taylor, D. R. (2014). Patterns of cyto-nuclear linkage disequilibrium in *Silene latifolia*: genomic heterogeneity and temporal stability. *Heredity*, *112*(2), 99–104. <https://doi.org/10.1038/hdy.2013.79>
- Figueroa, B., Chen, S., Oyler, G. A., Hardwick, J. M., & Betenbaugh, M. J. (2004). Aven and Bcl-xL enhance protection against apoptosis for mammalian cells exposed to various culture conditions. *Biotechnology and Bioengineering*, *85*(6), 589–600. <https://doi.org/10.1002/bit.10913>
- Friedman, J. R., & Nunnari, J. (2014). Mitochondrial form and function. *Nature*, *505*(7483), 335–343. <https://doi.org/10.1038/nature12985>
- Gao, Y., Katyal, S., Lee, Y., Zhao, J., Rehg, J. E., Russell, H. R., & McKinnon, P. J. (2011). DNA ligase III is critical for mtDNA integrity but not Xrcc1-mediated nuclear DNA repair. *Nature*, *471*(7337),

240–244. <https://doi.org/10.1038/nature09773>

- Ge, S., & Jung, D. (2018). ShinyGO: a graphical enrichment tool for animals and plants. *BioRxiv*, 315150. <https://doi.org/10.1101/315150>
- Genetical Society (Great Britain), D. P., & Scarpulla, R. C. (1987). *Genes & development*. Cold Spring Harbor Laboratory in association with the Genetical Society of Great Britain. Retrieved from <http://doi.org/10.1101/gad.1177604>
- Goodman, L. A., & Kruskal, W. H. (1972). Measures of Association for Cross Classifications, IV: Simplification of Asymptotic Variances. *Journal of the American Statistical Association*, 67(338), 415. <https://doi.org/10.2307/2284396>
- Greber, B. J., Boehringer, D., Leitner, A., Bieri, P., Voigts-Hoffmann, F., Erzberger, J. P., ... Ban, N. (2014). Architecture of the large subunit of the mammalian mitochondrial ribosome. *Nature*, 505(7484), 515–519. <https://doi.org/10.1038/nature12890>
- Hancock, J. M. (2004). HAVANA (Human and Vertebrate Analysis and Annotation). In *Dictionary of Bioinformatics and Computational Biology*. Chichester, UK: John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780471650126.dob0940>
- Hanson, M. R., & Bentolila, S. (2004). Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *The Plant Cell*, 16 Suppl(Suppl), S154-69. <https://doi.org/10.1105/tpc.015966>
- Hazkani-Covo, E., Zeller, R. M., & Martin, W. (2010). Molecular Poltergeists: Mitochondrial DNA Copies (numts) in Sequenced Nuclear Genomes. *PLoS Genetics*, 6(2), e1000834. <https://doi.org/10.1371/journal.pgen.1000834>
- He, J., Cooper, H. M., Reyes, A., Di Re, M., Sembongi, H., Litwin, T. R., ... Holt, I. J. (2012). Mitochondrial nucleoid interacting proteins support mitochondrial protein synthesis. *Nucleic Acids Research*, 40(13), 6109–6121. <https://doi.org/10.1093/nar/gks266>
- Karaman, M. W. (2003). Genomes, 2nd edition. *Journal of Heredity*, 94(5), 432-a-433. <https://doi.org/10.1093/jhered/esg082>
- Kasashima, K., Sumitani, M., & Endo, H. (2012). Maintenance of mitochondrial genome distribution by mitochondrial AAA+ protein ClpX. *Experimental Cell Research*, 318(18), 2335–2343. <https://doi.org/10.1016/j.yexcr.2012.07.012>
- Kenney, M. C., Chwa, M., Atilano, S. R., Falatoonzadeh, P., Ramirez, C., Malik, D., ... Udar, N. (2014). Inherited mitochondrial DNA variants can affect complement, inflammation and apoptosis pathways: insights into mitochondrial-nuclear interactions. *Human Molecular Genetics*, 23(13), 3537–3551. <https://doi.org/10.1093/hmg/ddu065>
- Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, D. (2002). The human genome browser at UCSC. *Genome Research*, 12(6), 996–1006. <https://doi.org/10.1101/gr.229102>
- Konovalova, S., & Tynismaa, H. (2013). Mitochondrial aminoacyl-tRNA synthetases in human disease. *Molecular Genetics and Metabolism*, 108(4), 206–211. <https://doi.org/10.1016/j.ymgme.2013.01.010>
- Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., ... Weirauch, M. T. (2018).

- The Human Transcription Factors. *Cell*, 172(4), 650–665.
<https://doi.org/10.1016/j.cell.2018.01.029>
- Lang, M., Sazzini, M., Calabrese, F. M., Simone, D., Boattini, A., Romeo, G., ... Gasparre, G. (2012). Polymorphic NumtS trace human population relationships. *Human Genetics*, 131(5), 757–771. <https://doi.org/10.1007/s00439-011-1125-3>
- Lascaro, D., Castellana, S., Gasparre, G., Romeo, G., Saccone, C., & Attimonelli, M. (2008). The RHNumtS compilation: Features and bioinformatics approaches to locate and quantify Human NumtS. *BMC Genomics*, 9(1), 267. <https://doi.org/10.1186/1471-2164-9-267>
- Latta, R. G., Linhart, Y. B., & Mitton, J. B. (2001). Cytonuclear Disequilibrium and Genetic Drift in a Natural Population of Ponderosa Pine. *Genetics*, 158(2).
- Leigh-Brown, S., Enriquez, J. A., & Odom, D. T. (2010). Nuclear transcription factors in mammalian mitochondria. *Genome Biology*, 11(7), 215. <https://doi.org/10.1186/gb-2010-11-7-215>
- Lewontin, R. C. (1964). The interaction of selection and linkage. I. general considerations; heterotic models. *Genetics*, 49(1).
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Majumder, P. P. (2010). The Human Genetic History of South Asia. *Current Biology*, 20(4), R184–R187. <https://doi.org/10.1016/J.CUB.2009.11.053>
- Malik, D., Hsu, T., Falatoonzadeh, P., Cáceres-del-Carpio, J., Tarek, M., Chwa, M., ... Kenney, M. C. (2014). Human retinal transmitochondrial cybrids with J or H mtDNA haplogroups respond differently to ultraviolet radiation: implications for retinal diseases. *PloS One*, 9(2), e99003. <https://doi.org/10.1371/journal.pone.0099003>
- Matsushima, Y., Goto, Y., & Kaguni, L. S. (2010). Mitochondrial Lon protease regulates mitochondrial DNA copy number and transcription by selective degradation of mitochondrial transcription factor A (TFAM). *Proceedings of the National Academy of Sciences of the United States of America*, 107(43), 18410–18415. <https://doi.org/10.1073/pnas.1008924107>
- McKenzie, M., Chiotis, M., Pinkert, C. A., & Trounce, I. A. (2003). Functional Respiratory Chain Analyses in Murid Xenomitochondrial Cybrids Expose Coevolutionary Constraints of Cytochrome b and Nuclear Subunits of Complex III. *Molecular Biology and Evolution*, 20(7), 1117–1124. <https://doi.org/10.1093/molbev/msg132>
- Meisinger, C., Sickmann, A., & Pfanner, N. (2008). The Mitochondrial Proteome: From Inventory to Function. *Cell*, 134(1), 22–24. <https://doi.org/10.1016/j.cell.2008.06.043>
- Nicholls, D. G. (2002). Mitochondrial function and dysfunction in the cell: its relevance to aging and aging-related disease. *The International Journal of Biochemistry & Cell Biology*, 34(11), 1372–1381. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12200032>
- Ohtsuji, M., Katsuoka, F., Kobayashi, A., Aburatani, H., Hayes, J. D., & Yamamoto, M. (2008). Nrf1 and Nrf2 play distinct roles in activation of antioxidant response element-dependent genes. *The Journal of Biological Chemistry*, 283(48), 33554–33562. <https://doi.org/10.1074/jbc.M804597200>

- Orr, H. A. (1996). Dobzhansky, Bateson, and the genetics of speciation. *Genetics*, 144(4), 1331–1335. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8978022>
- Pakendorf, B., & Stoneking, M. (2005). Mitochondrial DNA and human evolution. *Annual Review of Genomics and Human Genetics*, 6(1), 165–183. <https://doi.org/10.1146/annurev.genom.6.080604.162249>
- Parr, R. L., Maki, J., Reguly, B., Dakubo, G. D., Aguirre, A., Wittcock, R., ... Thayer, R. E. (2006). The pseudo-mitochondrial genome influences mistakes in heteroplasmy interpretation. *BMC Genomics*, 7, 185. <https://doi.org/10.1186/1471-2164-7-185>
- Petros, J. A., Baumann, A. K., Ruiz-Pesini, E., Amin, M. B., Sun, C. Q., Hall, J., ... Wallace, D. C. (2005). mtDNA mutations increase tumorigenicity in prostate cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 102(3), 719–724. <https://doi.org/10.1073/pnas.0408894102>
- Phillips, W. S., Coleman-Hulbert, A. L., Weiss, E. S., Howe, D. K., Ping, S., Wernick, R. I., ... Denver, D. R. (2015). Selfish Mitochondrial DNA Proliferates and Diversifies in Small, but not Large, Experimental Populations of *Caenorhabditis briggsae*. *Genome Biology and Evolution*, 7(7), 2023–2037. <https://doi.org/10.1093/gbe/evv116>
- Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., & Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Research*, 20(1), 110–121. <https://doi.org/10.1101/gr.097857.109>
- Pruitt, K. D., Brown, G. R., Hiatt, S. M., Thibaud-Nissen, F., Astashyn, A., Ermolaeva, O., ... Ostell, J. M. (2014). RefSeq: an update on mammalian reference sequences. *Nucleic Acids Research*, 42(Database issue), D756–63. <https://doi.org/10.1093/nar/gkt1114>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., ... Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81(3), 559–575. <https://doi.org/10.1086/519795>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- R: A Language and Environment for Statistical Computing. (2008). Retrieved from <http://www.gnu.org/copyleft/gpl.html>.
- Rand, D. M., Haney, R. A., & Fry, A. J. (2004). Cytonuclear coevolution: the genomics of cooperation. *Trends in Ecology & Evolution*, 19(12), 645–653. <https://doi.org/10.1016/j.tree.2004.10.003>
- Reyes, A., He, J., Mao, C. C., Bailey, L. J., Di Re, M., Sembongi, H., ... Holt, I. J. (2011). Actin and myosin contribute to mammalian mitochondrial DNA maintenance. *Nucleic Acids Research*, 39(12), 5098–5108. <https://doi.org/10.1093/nar/gkr052>
- Ricchetti, M., Tekaia, F., & Dujon, B. (2004). Continued Colonization of the Human Genome by Mitochondrial DNA. *PLoS Biology*, 2(9), e273. <https://doi.org/10.1371/journal.pbio.0020273>
- Rishishwar, L., & Jordan, I. K. (2017). Implications of human evolution and admixture for mitochondrial replacement therapy. *BMC Genomics*, 18(1), 140.

<https://doi.org/10.1186/s12864-017-3539-3>

- Roach, J. C., Glusman, G., Smit, A. F. A., Huff, C. D., Hubley, R., Shannon, P. T., ... Galas, D. J. (2010). Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science (New York, N.Y.)*, *328*(5978), 636–639. <https://doi.org/10.1126/science.1186802>
- Salas, A., Yao, Y.-G., Macaulay, V., Vega, A., Carracedo, Á., & Bandelt, H.-J. (2005). A Critical Reassessment of the Role of Mitochondria in Tumorigenesis. *PLoS Medicine*, *2*(11), e296. <https://doi.org/10.1371/journal.pmed.0020296>
- Salzberg, S. L., & Yorke, J. A. (2005). Beware of mis-assembled genomes. *Bioinformatics*, *21*(24), 4320–4321. <https://doi.org/10.1093/bioinformatics/bti769>
- Santos, C., Martínez, M., Lima, M., Hao, Y.-J., Simões, N., & Montiel, R. (2008). Mitochondrial DNA mutations in cancer: a review. *Current Topics in Medicinal Chemistry*, *8*(15), 1351–1366. Retrieved from <https://doi.org/10.2174/156802608786141151>
- Schnabel, A., & Asmussen, M. A. (1989). Definition and properties of disequilibria within nuclear-mitochondrial-chloroplast and other nuclear-dicytoplasmic systems. *Genetics*, *123*(1), 199–215. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2806883>
- Sharbrough, J., Havird, J. C., Noe, G. R., Warren, J. M., & Sloan, D. B. (2017). The Mitonuclear Dimension of Neanderthal and Denisovan Ancestry in Modern Human Genomes. *Genome Biology and Evolution*, *9*(6), 1567–1581. <https://doi.org/10.1093/gbe/evx114>
- Singh, B., Li, X., Owens, K. M., Vanniarajan, A., Liang, P., & Singh, K. K. (2015). Human REV3 DNA Polymerase Zeta Localizes to Mitochondria and Protects the Mitochondrial Genome. *PLOS ONE*, *10*(10), e0140409. <https://doi.org/10.1371/journal.pone.0140409>
- Slatkin, M. (2008). Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nature Reviews. Genetics*, *9*(6), 477–485. <https://doi.org/10.1038/nrg2361>
- Sloan, D. B., Fields, P. D., & Havird, J. C. (2015). Mitonuclear linkage disequilibrium in human populations. *Proceedings. Biological Sciences*, *282*(1815). <https://doi.org/10.1098/rspb.2015.1704>
- Smedley, D., Haider, S., Durinck, S., Pandini, L., Provero, P., Allen, J., ... Kasprzyk, A. (2015). The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Research*, *43*(W1), W589–W598. <https://doi.org/10.1093/nar/gkv350>
- Smirnov, A., Entelis, N., Martin, R. P., & Tarassov, I. (2011). Biological significance of 5S rRNA import into human mitochondria: role of ribosomal protein MRP-L18. *Genes & Development*, *25*(12), 1289–1305. <https://doi.org/10.1101/gad.624711>
- St. John, J. C. (2013). Mitochondrial DNA: Fate of the Paternal Mitochondrial Genome. In *eLS*. Chichester, UK: John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470015902.a0006165.pub2>
- Strautnieks, S. S., Bull, L. N., Knisely, A. S., Kocoshis, S. A., Dahl, N., Arnell, H., ... Thompson, R. J. (1998). A gene encoding a liver-specific ABC transporter is mutated in progressive familial intrahepatic cholestasis. *Nature Genetics*, *20*(3), 233–238. <https://doi.org/10.1038/3034>
- Taanman, J.-W. (1999). The mitochondrial genome: structure, transcription, translation and replication. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1410*(2), 103–123.

[https://doi.org/10.1016/S0005-2728\(98\)00161-3](https://doi.org/10.1016/S0005-2728(98)00161-3)

- Tews, D., Fischer-Posovszky, P., Fromme, T., Klingenspor, M., Fischer, J., Rütter, U., ... Wabitsch, M. (2013). FTO Deficiency Induces UCP-1 Expression and Mitochondrial Uncoupling in Adipocytes. *Endocrinology*, *154*(9), 3141–3151. <https://doi.org/10.1210/en.2012-1873>
- Thalmann, O., Hebler, J., Poinar, H. N., Pääbo, S., & Vigilant, L. (2004). Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes. *Molecular Ecology*, *13*(2), 321–335. Retrieved from <https://doi.org/10.1046/j.1365-294X.2003.02070.x>
- Tu, P., Gibon, J., & Bouron, A. (2010). The TRPC6 channel activator hyperforin induces the release of zinc and calcium from mitochondria. *Journal of Neurochemistry*, *112*(1), 204–213. <https://doi.org/10.1111/j.1471-4159.2009.06446.x>
- Unoki, M., & Nakamura, Y. (2003). EGR2 induces apoptosis in various cancer cell lines by direct transactivation of BNIP3L and BAK. *Oncogene*, *22*, 2172–2185. <https://doi.org/10.1038/sj.onc.1206222>
- van Rootselaar, A.-F., Groffen, A. J., de Vries, B., Callenbach, P. M. C., Santen, G. W. E., Koelewijn, S., ... van den Maagdenberg, A. M. J. M. (2017). δ -Catenin (*CTNND2*) missense mutation in familial cortical myoclonic tremor and epilepsy. *Neurology*, *89*(23), 2341–2350. <https://doi.org/10.1212/WNL.0000000000004709>
- Wang, X., Huang, N., Yang, M., Wei, D., Tai, H., Han, X., ... Xiao, H. (2017). FTO is required for myogenesis by positively regulating mTOR-PGC-1 α pathway-mediated mitochondria biogenesis. *Cell Death & Disease*, *8*(3), e2702. <https://doi.org/10.1038/cddis.2017.122>
- Warnes, G., Gorjanc, G., Leisch, F., & Man, M. (2012). Genetics: Population genetics. R package version 1.3. 6. Retrieved from <https://scholar.google.com/scholar?cluster=6973167904226562869&hl=en&oi=scholar>
- Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., ... Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Research*, *44*(W1), W58–W63. <https://doi.org/10.1093/nar/gkw233>
- Wheeler, M. T., Zarnegar, S., & McNally, E. M. (2002). Zeta-sarcoglycan, a novel component of the sarcoglycan complex, is reduced in muscular dystrophy. *Human Molecular Genetics*, *11*(18), 2147–2154. Retrieved from <https://doi.org/10.1093/hmg/11.18.2147>
- Yamamoto, Y. (2001). D-Loop. In *Encyclopedia of Genetics* (pp. 539–540). Elsevier. <https://doi.org/10.1006/rwgn.2001.0346>
- Yates, B., Braschi, B., Gray, K. A., Seal, R. L., Tweedie, S., & Bruford, E. A. (2017). Genenames.org: the HGNC and VGNC resources in 2017. *Nucleic Acids Research*, *45*(D1), D619–D625. <https://doi.org/10.1093/nar/gkw1033>
- Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J., ... Flicek, P. (2018). Ensembl 2018. *Nucleic Acids Research*, *46*(D1), D754–D761. <https://doi.org/10.1093/nar/gkx1098>
- Zhang, Y., Li, F., Patterson, A. D., Wang, Y., Krausz, K. W., Neale, G., ... Schuetz, J. D. (2012). Abcb11

Deficiency Induces Cholestasis Coupled to Impaired-Fatty Acid Oxidation in Mice * □ S.
<https://doi.org/10.1074/jbc.M111.329318>

ZHAO, H., NETTLETON, D., & DEKKERS, J. C. M. (2007). Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between single nucleotide polymorphisms. *Genetical Research*, 89(01), 1.
<https://doi.org/10.1017/S0016672307008634>

Anexos

Anexo 1: Todos los genes reportados con SNP bajo desequilibrio mitonuclear

Poblaciones:

Genes	Descripción
NCOA1	Nuclear receptor coactivator 1;NCOA1;ortholog
SFI1	Protein SFI1 homolog;SFI1;ortholog
PSG8	Pregnancy-specific beta-1-glycoprotein 8;PSG8;ortholog
RNF157	RING finger protein 157;RNF157;ortholog
STPG1	O(6)-methylguanine-induced apoptosis 2;STPG1;ortholog
CSRP2	Cysteine and glycine-rich protein 2;CSRP2;ortholog
TRPC6	Short transient receptor potential channel 6;TRPC6;ortholog
RBM18	Probable RNA-binding protein 18;RBM18;ortholog
AL139099	DNA polymerase epsilon subunit 2;POLE2;ortholog
TRIM2	Tripartite motif-containing protein 2;TRIM2;ortholog
CPM	Carboxypeptidase M;CPM;ortholog
COL9A1	Collagen alpha-1(IX) chain;COL9A1;ortholog
RORB	Nuclear receptor ROR-beta;RORB;ortholog
RGS6	Regulator of G-protein signaling 6;RGS6;ortholog
PRR14L	Protein PRR14L;PRR14L;ortholog
NHSL1	NHS-like protein 1;NHSL1;ortholog
STX11	Syntaxin-11;STX11;ortholog
HAO1	Hydroxyacid oxidase 1;HAO1;ortholog
GALNTL6	Polypeptide N-acetylgalactosaminyltransferase-like 6;GALNTL6;ortholog
SSBP3	Single-stranded DNA-binding protein 3;SSBP3;ortholog
AC022893,TERF1	Telomeric repeat-binding factor 1;TERF1;ortholog
NCBP3	Nuclear cap-binding protein subunit 3;NCBP3;ortholog
SPATA16	Spermatogenesis-associated protein 16;SPATA16;ortholog

PDE1C	Calcium/calmodulin-dependent 3',5'-cyclic nucleotide phosphodiesterase 1C;PDE1C;ortholog
AC022893	Somatomedin-B and thrombospondin type-1 domain-containing protein;SBSPON;ortholog
TENM3	Teneurin-3;TENM3;ortholog
MGAT2	Alpha-1,6-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase;MGAT2;ortholog
SOX5	Transcription factor SOX-5;SOX5;ortholog
DPP10	Inactive dipeptidyl peptidase 10;DPP10;ortholog
RBM20	RNA-binding protein 20;RBM20;ortholog
SCGB1A1	Uteroglobin;SCGB1A1;ortholog
RAB36	Ras-related protein Rab-36;RAB36;ortholog
RALGPS2	Ras-specific guanine nucleotide-releasing factor RalGPS2;RALGPS2;ortholog
VASH2	Vasohibin-2;VASH2;ortholog
CACNA1D	Voltage-dependent L-type calcium channel subunit alpha-1D;CACNA1D;ortholog
AC048338	Leucine-rich repeat-containing protein 43;LRRC43;ortholog
DGKB	Diacylglycerol kinase beta;DGKB;ortholog
OCA2	P protein;OCA2;ortholog
CLYBL	Citrate lyase subunit beta-like protein, mitochondrial;CLYBL;ortholog
HERC2	E3 ubiquitin-protein ligase HERC2;HERC2;ortholog
CD36	Platelet glycoprotein 4;CD36;ortholog
AC048338	Uncharacterized protein;unassigned;ortholog
PDGFC	Platelet-derived growth factor C;PDGFC;ortholog
AC090518	Zinc finger protein 280D;ZNF280D;ortholog
GNA14	Guanine nucleotide-binding protein subunit alpha-14;GNA14;ortholog
CDH12	Cadherin-12;CDH12;ortholog
HMGB1	High mobility group protein B1;HMGB1;ortholog
AL139099	Protein kintoun;DNAAF2;ortholog
LRRC31	Leucine-rich repeat-containing protein 31;LRRC31;ortholog
CNTN1	Contactin-1;CNTN1;ortholog
LUC7L3	Luc7-like protein 3;LUC7L3;ortholog
ADGRL2	Adhesion G protein-coupled receptor L2;ADGRL2;ortholog
C3orf67	Uncharacterized protein C3orf67;C3orf67;ortholog

NPAS2	Neuronal PAS domain-containing protein 2;NPAS2;ortholog
SPC25	Kinetochore protein Spc25;SPC25;ortholog
TRIM41	E3 ubiquitin-protein ligase TRIM41;TRIM41;ortholog
ABCB11	Bile salt export pump;ABCB11;ortholog
CD226	CD226 antigen;CD226;ortholog
WDR7	WD repeat-containing protein 7;WDR7;ortholog
SLC22A2	Solute carrier family 22 member 2;SLC22A2;ortholog
PPARG	Peroxisome proliferator-activated receptor gamma;PPARG;ortholog
DYSF	Dysferlin;DYSF;ortholog
CDK8	Cyclin-dependent kinase 8;CDK8;ortholog
BCAS3	Breast carcinoma-amplified sequence 3;BCAS3;ortholog
CHMP3	Charged multivesicular body protein 3;CHMP3;ortholog
ZNF346	Zinc finger protein 346;ZNF346;ortholog
AL139099	40S ribosomal protein S29;RPS29;ortholog
PISD	Phosphatidylserine decarboxylase proenzyme, mitochondrial;PISD;ortholog
CCNYL2	Cyclin-Y-like protein 2;CCNYL2;ortholog
CARMIL1	F-actin-uncapping protein LRRC16A;CARMIL1;ortholog
COBL	Protein cordon-bleu;COBL;ortholog
DLEU1	Leukemia-associated protein 1;DLEU1;ortholog
JAKMIP2	Janus kinase and microtubule-interacting protein 2;JAKMIP2;ortholog
ZNF287	Zinc finger protein 287;ZNF287;ortholog
DPH6	Diphthine--ammonia ligase;DPH6;ortholog
TRPM3	Transient receptor potential cation channel subfamily M member 3;TRPM3;ortholog
MAN1A1	Mannosyl-oligosaccharide 1,2-alpha-mannosidase IA;MAN1A1;ortholog
EFCAB6	EF-hand calcium-binding domain-containing protein 6;EFCAB6;ortholog
CAST	DNA-directed RNA polymerase I subunit RPA34;CD3EAP;ortholog
FTO	Alpha-ketoglutarate-dependent dioxygenase FTO;FTO;ortholog
CAST	Calpastatin;CAST;ortholog
WDFY3	WD repeat and FYVE domain-containing protein 3;WDFY3;ortholog
AVEN	Cell death regulator Aven;AVEN;ortholog
RORA	Nuclear receptor ROR-alpha;RORA;ortholog

PDE4D	cAMP-specific 3',5'-cyclic phosphodiesterase 4D;PDE4D;ortholog
DRG1	Protein NDRG1;NDRG1;ortholog
AKR1C2	Aldo-keto reductase family 1 member C2;AKR1C2;ortholog
CCDC15	Coiled-coil domain-containing protein 15;CCDC15;ortholog
FAH	Fumarylacetoacetase;FAH;ortholog
ERBB4	Receptor tyrosine-protein kinase erbB-4;ERBB4;ortholog
DRG1	Developmentally-regulated GTP-binding protein 1;DRG1;ortholog
SEPT9	Septin-9;SEPT9;ortholog
SGO2	Shugoshin 2;SGO2;ortholog
AC012413,PXDNL	Peroxidasin-like protein;PXDNL;ortholog
MGAT2	2-acylglycerol O-acyltransferase 2;MOGAT2;ortholog
CLIC5	Chloride intracellular channel protein 5;CLIC5;ortholog
SPC25	Signal peptidase complex subunit 2;SPCS2;ortholog
AC012158	Uncharacterized protein C12orf56;C12orf56;ortholog
FREM2	FRAS1-related extracellular matrix protein 2;FREM2;ortholog
TMCO4	Transmembrane and coiled-coil domain-containing protein 4;TMCO4;ortholog
LINC00299	Putative uncharacterized protein encoded by LINC00299;LINC00299;ortholog
CDH23	Cadherin-23;CDH23;ortholog

Superpoblaciones:

Genes	Descripción
UBE2L6	Ubiquitin/ISG15-conjugating enzyme E2 L6;UBE2L6;ortholog
AP001258	Glycine-N-acyltransferase-like 1 pseudogene 3;GLYATL1P3;ortholog
AC012574	Homeobox protein Nkx-2.6;NKX2-6;ortholog
SP1	DAN domain family member 5;DAND5;ortholog
LINC00301	Putative uncharacterized protein encoded by LINC00301;LINC00301;ortholog
FGF5	Fibroblast growth factor 5;FGF5;ortholog

CRADD	Death domain-containing protein CRADD;CRADD;ortholog
STX18	Syntaxin-18;STX18;ortholog
TMEM200A	Transmembrane protein 200A;TMEM200A;ortholog
DYDC2	DPY30 domain-containing protein 2;DYDC2;ortholog
VPS13C	Vacuolar protein sorting-associated protein 13C;VPS13C;ortholog
FSTL4	Follistatin-related protein 4;FSTL4;ortholog
CTDSP2	Carboxy-terminal domain RNA polymerase II polypeptide A small phosphatase 2;CTDSP2;ortholog
AC105001,SOX7	Transcription factor SOX-7;SOX7;ortholog
AF228730	Putative protein FAM90A5P;FAM90A5P;ortholog
IFNGR1	Interferon gamma receptor 1;IFNGR1;ortholog
AC073611	Separin;ESPL1;ortholog
UNC5D	Netrin receptor UNC5D;UNC5D;ortholog
TOX4	TOX high mobility group box family member 4;TOX4;ortholog
GLYATL1	Glycine N-acyltransferase-like protein 1;GLYATL1;ortholog
CCSER2	Serine-rich coiled-coil domain-containing protein 2;CCSER2;ortholog
AL353807	Endogenous retrovirus group K member 7 Pol protein;ERVK-7;ortholog
MED23	Mediator of RNA polymerase II transcription subunit 23;MED23;ortholog
HDAC4	Histone deacetylase 4;HDAC4;ortholog
MCRS1	Microspherule protein 1;MCRS1;ortholog
ABCC8	ATP-binding cassette sub-family C member 8;ABCC8;ortholog
PLXNC1	Plexin-C1;PLXNC1;ortholog
AC090193	F-box only protein 32;FBXO32;ortholog
VSIG10	V-set and immunoglobulin domain-containing protein 10;VSIG10;ortholog
FAM213A	Redox-regulatory protein FAM213A;FAM213A;ortholog
KRT79	Keratin, type II cytoskeletal 79;KRT79;ortholog
AC073611	Transcription factor Sp7;SP7;ortholog
MOXD1	DBH-like monooxygenase protein 1;MOXD1;ortholog
SOHLH2	Spermatogenesis- and oogenesis-specific basic helix-loop-helix-containing protein 2;SOHLH2;ortholog
TENM4	Teneurin-4;TENM4;ortholog

NHSL1	NHS-like protein 1;NHSL1;ortholog
HSD17B4	Peroxisomal multifunctional enzyme type 2;HSD17B4;ortholog
SP1,AC073611	Transcription factor Sp1;SP1;ortholog
CTNND1	Catenin delta-1;CTNND1;ortholog
AP000777	Coiled-coil domain-containing protein 86;CCDC86;ortholog
EHBP1L1	EH domain-binding protein 1-like protein 1;EHBP1L1;ortholog
RASSF4	Ras and Rab interactor 2;RIN2;ortholog
DCN	Decorin;DCN;ortholog
MPZL1	Myelin protein zero-like protein 1;MPZL1;ortholog
REPS1	RalBP1-associated Eps domain-containing protein 1;REPS1;ortholog
OSBPL6	Oxysterol-binding protein-related protein 6;OSBPL6;ortholog
AC105389	LIM and calponin homology domains-containing protein 1;LIMCH1;ortholog
SDK1	Protein kinase C delta type;PRKCD;ortholog
SLC24A3	Sodium/potassium/calcium exchanger 3;SLC24A3;ortholog
MTUS2	Microtubule-associated tumor suppressor candidate 2;MTUS2;ortholog
TPGS2	Tubulin polyglutamylase complex subunit 2;TPGS2;ortholog
AC069437	Ankyrin repeat and sterile alpha motif domain-containing protein 1B;ANKS1B;ortholog
APBA2	Amyloid beta A4 precursor protein-binding family A member 2;APBA2;ortholog
MYB	Transcriptional activator Myb;MYB;ortholog
SERPING1	Plasma protease C1 inhibitor;SERPING1;ortholog
SLCO1B1	Solute carrier organic anion transporter family member 1B1;SLCO1B1;ortholog
CPA2	Carboxypeptidase A2;CPA2;ortholog
LAMA1	Laminin subunit alpha-1;LAMA1;ortholog
EME2	Probable crossover junction endonuclease EME2;EME2;ortholog
AC089983	Host cell factor 2;HCFC2;ortholog
KSR2	Kinase suppressor of Ras 2;KSR2;ortholog
AC089987	Olfactory receptor 8S1;OR8S1;ortholog
OR4A8	Olfactory receptor 4A8;OR4A8;ortholog
AC105001	Uncharacterized protein;unassigned;ortholog
KRT6C	Keratin, type II cytoskeletal 6C;KRT6C;ortholog
TRPS1	Zinc finger transcription factor Trps1;TRPS1;ortholog

APLNR	Apelin receptor;APLNR;ortholog
AC073611	Aladin;AAAS;ortholog
GDAP1L1	Ganglioside-induced differentiation-associated protein 1-like 1;GDAP1L1;ortholog
AVIL	Advillin;AVIL;ortholog
AC073611	Prefoldin subunit 5;PFDN5;ortholog
CHST11	Carbohydrate sulfotransferase 11;CHST11;ortholog
ARG1	Arginase-1;ARG1;ortholog
ERICH1	Glutamate-rich protein 1;ERICH1;ortholog
POLE3	DNA polymerase epsilon subunit 3;POLE3;ortholog
CLN8	Protein CLN8;CLN8;ortholog
IGF1	Insulin-like growth factor I;IGF1;ortholog
AP000777	Prostaglandin D2 receptor 2;PTGDR2;ortholog
STAB2	Stabilin-2;STAB2;ortholog
AC008575	Uncharacterized protein;unassigned;ortholog
AC089983,TXNRD1	Thioredoxin reductase 1, cytoplasmic;TXNRD1;ortholog
GEMIN5	Gem-associated protein 5;GEMIN5;ortholog
CCDC171	Coiled-coil domain-containing protein 171;CCDC171;ortholog
P2RX3	P2X purinoceptor 3;P2RX3;ortholog
PACRG	Parkin coregulated gene protein;PACRG;ortholog
CRY1	Cryptochrome-1;CRY1;ortholog
LIPM	Lipase member M;LIPM;ortholog
CCDC28A	Coiled-coil domain-containing protein 28A;CCDC28A;ortholog
CCDC169	Coiled-coil domain-containing protein 169;CCDC169;ortholog
RASSF4	Ras association domain-containing protein 4;RASSF4;ortholog
PTPRK	Receptor-type tyrosine-protein phosphatase kappa;PTPRK;ortholog
SLC4A10	Sodium-driven chloride bicarbonate exchanger;SLC4A10;ortholog
GOLT1B	Vesicle transport protein GOT1B;GOLT1B;ortholog
TIGAR	Fructose-2,6-bisphosphatase TIGAR;TIGAR;ortholog
ERN2	Serine/threonine-protein kinase/endoribonuclease IRE2;ERN2;ortholog
TAAR1	Trace amine-associated receptor 1;TAAR1;ortholog
ATAD2	ATPase family AAA domain-containing protein 2;ATAD2;ortholog

IL20RA	Interleukin-20 receptor subunit alpha;IL20RA;ortholog
ZFP91	E3 ubiquitin-protein ligase ZFP91;ZFP91;ortholog
MSTO1,AL353807	Protein misato homolog 1;MSTO1;ortholog
PRPSAP2	Phosphoribosyl pyrophosphate synthase-associated protein 2;PRPSAP2;ortholog
AF228730	Putative protein FAM90A15P;FAM90A15P;ortholog
CEP72	Centrosomal protein of 72 kDa;CEP72;ortholog
EPHB1	Ephrin type-B receptor 1;EPHB1;ortholog
GRID1	Glutamate receptor ionotropic, delta-1;GRID1;ortholog
PRCP	Lysosomal Pro-X carboxypeptidase;PRCP;ortholog
AF228730	Ubiquitin carboxyl-terminal hydrolase 17-like protein 1;USP17L1;ortholog
SPSB3	SPRY domain-containing SOCS box protein 3;SPSB3;ortholog
AC008575	Signal recognition particle 19 kDa protein;SRP19;ortholog
ZFP91-CNTF	HCG2042749, isoform CRA_d;ZFP91-CNTF;ortholog
SLC43A1	Large neutral amino acids transporter small subunit 3;SLC43A1;ortholog
FAM114A2	Protein FAM114A2;FAM114A2;ortholog
AC008575	Uncharacterized protein (Fragment);unassigned;ortholog
ALDH1L2	Mitochondrial 10-formyltetrahydrofolate dehydrogenase;ALDH1L2;ortholog
AC121338	Vitamin D3 receptor;VDR;ortholog
LETMD1	LETM1 domain-containing protein 1;LETMD1;ortholog
LIPK	Lipase member K;LIPK;ortholog
SLC30A10	Zinc transporter 10;SLC30A10;ortholog
KCNIP4	Kv channel-interacting protein 4;KCNIP4;ortholog
EYA4	Eyes absent homolog 4;EYA4;ortholog
ASB3	Ankyrin repeat and SOCS box protein 3;ASB3;ortholog
ACBD3	Golgi resident protein GCP60;ACBD3;ortholog
CELF1	CUGBP Elav-like family member 1;CELF1;ortholog
OR5M10	Olfactory receptor 5M10;OR5M10;ortholog
AL353807	Histone-lysine N-methyltransferase ASH1L;ASH1L;ortholog
LPXN	Leupaxin;LPXN;ortholog

ECHDC1	Ethylmalonyl-CoA decarboxylase;ECHDC1;ortholog
AC008575	Receptor expression-enhancing protein 5;REEP5;ortholog
ATP5G2	ATP synthase F(0) complex subunit C2, mitochondrial;ATP5G2;ortholog
EML5	Echinoderm microtubule-associated protein-like 5;EML5;ortholog
KCNU1	Potassium channel subfamily U member 1;KCNU1;ortholog
PDGFRL	Platelet-derived growth factor receptor-like protein;PDGFRL;ortholog
SDK1	Protein sidekick-1;SDK1;ortholog
MSR1	Macrophage scavenger receptor types I and II;MSR1;ortholog
AF228730	Putative protein FAM90A13P;FAM90A13P;ortholog
AL356124	Laminin subunit alpha-2;LAMA2;ortholog
MRGPRG-AS1	Putative uncharacterized protein MRGPRG-AS1;MRGPRG-AS1;ortholog
AF228730	Putative protein FAM90A20P;FAM90A20P;ortholog
C12orf74	Uncharacterized protein C12orf74;C12orf74;ortholog
AC105001	Uncharacterized protein C8orf74;C8orf74;ortholog
OR9Q1	Olfactory receptor 9Q1;OR9Q1;ortholog
BTBD11	Ankyrin repeat and BTB/POZ domain-containing protein BTBD11;BTBD11;ortholog
HS3ST2	Heparan sulfate glucosamine 3-O-sulfotransferase 2;HS3ST2;ortholog
AC073611,C12orf10	UPF0160 protein MYG1, mitochondrial;C12orf10;ortholog
AC105001	Retinitis pigmentosa 1-like 1 protein;RP1L1;ortholog
MB21D2	Protein MB21D2;MB21D2;ortholog
OR8K5	Olfactory receptor 8K5;OR8K5;ortholog
HBS1L	HBS1-like protein;HBS1L;ortholog
ZDHHC2	Palmitoyltransferase ZDHHC2;ZDHHC2;ortholog
QDPR	Dihydropteridine reductase;QDPR;ortholog

