



La participación de las mujeres migrantes en la
formación de las poblaciones mestizas de Chile:
Un estudio a partir del ADN mitocondrial

Tesis

Entregada A La
Universidad De Chile
En Cumplimiento Parcial De Los Requisitos
Para Optar Al Grado De

Magíster en Ciencias Biológicas

Facultad De Ciencias

Por

Margarita Belén Reyes Madrid

Marzo, 2023

Director de Tesis Dr. Mauricio Moraga

FACULTAD DE CIENCIAS

UNIVERSIDAD DE CHILE

INFORME DE APROBACIÓN

TESIS DE MAGÍSTER

Se informa a la Escuela de Postgrado de la Facultad de Ciencias que la Tesis de Magíster presentada por la candidata.

Margarita Belén Reyes Madrid

Ha sido aprobada por la comisión de Evaluación de la tesis como requisito para optar al grado de Magíster en Ciencias Biológicas, en el examen de Defensa Privada de Tesis rendido el día 30 de enero de 2023.

Director de Tesis:

Dr. Mauricio Moraga V.

Comisión de Evaluación de la Tesis

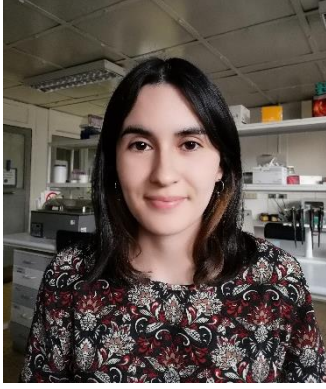
Dr. Elie Poulin C.

Dr. Luis Castañeda S.

DEDICATORIA

A mi hermana Isidora.

BIOGRAFÍA



Antropóloga física, titulada de la Universidad de Chile (2019). Durante mi pregrado, fue el área de genética aplicada al análisis en poblaciones humanas, la línea de investigación de mi interés y aquella que decidí seguir para la etapa final de mi carrera. Realicé mi memoria de pregrado en el Laboratorio de Genética de la Facultad de Ciencias Sociales, enfocada en caracterizar genéticamente dos poblaciones urbanas de la región de Antofagasta a partir de la región no recombinante del cromosoma Y. Esta experiencia me aproximó al estudio de la variabilidad genética humana analizando marcadores moleculares de herencia uniparental para evaluar procesos migratorios y dinámicas poblacionales. Con la intención de seguir esta línea de investigación, en marzo de 2020 ingresé al programa de Magíster en Ciencias Biológicas de la Facultad de Ciencias, donde cursé distintos ramos que profundizaron mis conocimientos en genética y genómica, ecología molecular y procesos filogenéticos. Realicé mi unidad de investigación en el Laboratorio de Bioinformática Integrativa de la Facultad de Ciencias Químicas y Farmacéuticas (U. de Chile), donde aprendí herramientas bioinformáticas para el análisis de variantes genómicas. La presente tesis realizada en el Laboratorio de Genética de Poblaciones y Evolución Humana (Facultad de Medicina, U. de Chile), bajo la tutela del Dr. Mauricio Moraga, surge a partir de un interés de ambos por dar a conocer el aporte de las mujeres migrantes a la diversidad mitocondrial de las poblaciones mestizas chilenas.

AGRADECIMIENTOS

A la Agencia Nacional de Investigación y Desarrollo (ANID) por financiar mi paso por el programa de Magíster en Ciencias Biológicas a través de la Beca de Magíster Nacional 2020.

Al Proyecto Fondecyt #1181889 por el financiamiento de los análisis que permitieron llevar a cabo la presente tesis.

Al profesor Mauricio Moraga por la paciencia, apoyo y comprensión durante los buenos y complejos momentos en el desarrollo de esta tesis.

A los profesores de la comisión revisora, Elie Poulin y Luis Castañeda, por su comprensión y apoyo cuando me reuní con ellos.

A mis compañeros del Laboratorio de Genética de Poblaciones y Evolución Humana por las buenas conversaciones, ánimo y contención en las distintas etapas de este proceso. Destaco a Carla Cerda por siempre brindarme su ayuda en el laboratorio, a Michael Orellana por responder a mis dudas constantes y a Daniela Castillo quien me ayudó en la crisis de los mitogenomas.

A mis amigas, especialmente María Paz por el apoyo brindado este año.

A Nicolás por siempre estar ahí y entregarme el apoyo y contención necesarios para seguir.

A mi familia, mis padres y hermana, por su apoyo y confianza tácitos, que me impulsaron a sacar esta tesis adelante.

ÍNDICE DE MATERIAS

DEDICATORIA	ii
BIOGRAFÍA	iii
AGRADECIMIENTOS	iv
ÍNDICE DE MATERIAS	v
LISTA DE TABLAS	ix
LISTA DE FIGURAS	x
RESUMEN	1
ABSTRACT	3
1. INTRODUCCIÓN	5
1.1. Antecedentes	6
1.1.1. <u>Antecedentes históricos</u>	6
1.1.1.1. Mujeres migrantes durante la conquista y colonia europea en América y Chile	6
1.1.1.2. Mujeres migrantes durante el tráfico de esclavos en América y Chile	9
1.1.1.3. Mujeres migrantes durante los siglos XIX y XX en Chile	14
1.1.2. <u>Antecedentes genéticos</u>	17
1.1.2.1. Composición genética de las poblaciones sudamericanas	17
1.1.2.2. ADN mitocondrial como marcador poblacional	19
1.2. Problematicación	24
1.3. Hipótesis	25
1.4. Objetivos	26

1.4.1. <u>Objetivo general</u>	26
1.4.2. <u>Objetivos específicos</u>	26
2. MATERIAL Y MÉTODOS	28
2.1. Muestra.....	28
2.2. Técnicas de recolección y procesamiento de datos	30
2.2.1. <u>Secuenciación Sanger de D-loop completo</u>	30
2.2.2. <u>Secuenciación de Nueva Generación (NGS) de mitogenomas</u>	32
2.2.2.1. Evaluación y diseño de protocolo de amplificación por PCR	34
2.2.2.2. Cuantificación y purificación de ADN	36
2.2.2.3. Construcción de librerías genómicas	37
2.2.2.4. Control de calidad y secuenciación de librerías genómicas.....	41
2.2.2.5. Procesamiento de lecturas obtenidas por secuenciación NGS.....	42
2.3. Métodos de análisis de datos	46
2.3.1. <u>Determinación de haplogrupos a partir de D-loop completo</u>	46
2.3.2. <u>Análisis de diversidad genética a partir de D-loop completo</u>	47
2.3.3. <u>Análisis de estructura poblacional a partir de D-loop completo</u>	48
2.3.4. <u>Análisis multivariados con información de ancestría genómica</u>	50
2.3.5. <u>Comparación con poblaciones europeas y africanas</u>	50
2.3.6. <u>Análisis filogenéticos a partir de genoma mitocondrial completo</u>	52
3. RESULTADOS	56
3.1. Resultados de trabajo de laboratorio y bioinformático	56
3.1.1. <u>Secuenciación de D-loop completo</u>	56
3.1.2. <u>Secuenciación NGS de mitogenomas</u>	57

3.2. Resultados de análisis de datos	66
<u>3.2.1. Caracterización de la diversidad de linajes mitocondriales no amerindios en Chile</u>	66
3.2.1.1. Frecuencias de haplogrupos mitocondriales no amerindios	66
3.2.1.2. Índices de diversidad molecular	73
3.2.1.3. Estructura poblacional: Análisis de FST y Dendrogramas	76
3.2.1.4. Estructura poblacional: Análisis de Componentes Principales	78
3.2.1.5. Redes de haplotipos	82
3.2.1.6. Comparación con marcadores informativos de ancestría (AIMs)	87
<u>3.2.2. Distribución geográfica y origen de linajes mitocondriales no amerindios</u>	90
3.2.2.1. Linajes euroasiáticos en Chile y Europa	90
3.2.2.2. Linajes africanos en Chile y África	95
<u>3.2.3. Filogenia de linajes mitocondriales no amerindios en Chile a partir de mitogenomas</u>	100
3.2.3.1. Caracterización de mitogenomas no amerindios encontrados en Chile	100
3.2.3.2. Reconstrucción filogenética de ADN mitocondrial	103
4. DISCUSIÓN	109
4.1. Diversidad mitocondrial no amerindia en Chile	109
4.2. Mujeres migrantes en el periodo de conquista y colonia española	111
4.3. Mujeres migrantes en el periodo de tráfico de esclavos	115
4.4. Mujeres migrantes en los siglos XIX y XX	119
4.5. Consideraciones finales y proyecciones	123

5. CONCLUSIONES	126
6. BIBLIOGRAFÍA	128
7. ANEXOS	142

LISTA DE TABLAS

Tabla 1: Poblaciones chilenas incluidas en este estudio.....	29
Tabla 2: Partidores utilizados para la amplificación del D-loop del ADN mitocondrial con sus respectivas temperaturas de alineamiento (T°a)	31
Tabla 3: Partidores utilizados para la secuenciación del D-loop de ADNmt	31
Tabla 4: Partidores utilizados en la amplificación de fragmentos para la secuenciación de mitogenomas con sus respectivas temperaturas de alineamiento (T° a)	35
Tabla 5: Frecuencias relativas de los macrohaplogrupos mitocondriales no amerindios por población chilena.....	70
Tabla 6: Índices de diversidad molecular y pruebas de neutralidad a nivel país y poblacional.....	75
Tabla 7: Proporciones de ancestría promedio para cada población chilena	88
Tabla 8: Muestras secuenciadas por NGS con su respectivo haplogrupo determinado a partir de D-loop y del genoma mitocondrial completo	101
Tabla 9: Modelos de sustitución para cada partición junto a sus parámetros obtenidos por IQ-Tree y que fueron ingresados en MrBayes	105
Tabla 10: Índices diagnósticos del análisis bayesiano, resumen de todos los parámetros ·	105
Tabla S1. Protocolos de amplificación de fragmentos largos para la secuenciación de mitogenomas.....	142
Tabla S2: Matriz de FST entre pares de poblaciones chilenas obtenida a partir de las frecuencias de haplotipos estimados por DNAsp	143
Tabla S3: Matriz de FST entre pares de poblaciones chilenas obtenido a partir de las frecuencias de 27 haplogrupos	144
Tabla S4: Poblaciones africanas de referencia utilizadas para los análisis comparativos	145
Tabla S5: Poblaciones europeas de referencia utilizadas en la red de haplotipos de HV	146
Tabla S6: Poblaciones africanas de referencia utilizadas en la red de haplotipos de L	146

LISTA DE FIGURAS

Figura 1: Regiones históricas africanas asociadas a la trata de esclavos con sus respectivos límites políticos actuales. (Fuente: Eltis y Richardson, 2010).....	11
Figura 2: Mapa de género y edad de esclavos traídos a América entre 1545-1864 (Fuente: Eltis y Richardson, 2010).....	13
Figura 3: Proporción de migrantes mujeres de distintas nacionalidades en Chile en el año 1907. Elaboración propia a partir del censo (INE, 1907)	16
Figura 4: Proporción de migrantes mujeres de distintas nacionalidades en Chile en el año 1930. Elaboración propia a partir del censo (INE, 1930)	17
Figura 5: Estructura de la molécula de ADN mitocondrial. Se indican los genes, los orígenes de replicación para la cadena pesada y liviana y hacia superior se destaca la región control o d-loop (Fuente: Jobling y col., 2013)	20
Figura 6: Flujo de trabajo de la secuenciación NGS de mitogenomas. A la izquierda se presentan los procedimientos de laboratorio para la preparación de las muestras, mientras que a la derecha se señalan los análisis bioinformáticos para el procesamiento de los datos. Además, al lado derecho de cada recuadro se indica el reactivo o herramienta utilizada en esa etapa y al lado izquierdo el formato del output generado	33
Figura 7: Ubicación de los siete fragmentos amplificados por PCR para la secuenciación del genoma mitocondrial completo. Se posicionan siguiendo el orden en que aparecen en la tabla XX y se indica el tamaño de cada uno	36
Figura 8: Gel de agarosa al 1,5% luego de electroforesis para 11 productos de PCR de D-loop. STD: estándar de peso molecular (en este caso de 1kb).....	56
Figura 9: Visor de programa Geneious con los cromatogramas de la muestra XL120 alineados a la secuencia de referencia. Aparecen destacados sitios donde se observa una doble señal de bases nitrogenadas	57
Figura 10: Gel de agarosa al 1,5% del mix de 7 fragmentos antes y después de la purificación de ADN. Las muestras purificadas se indican con una (P). STD: estándar de peso molecular (1kb)	58

Figura 11: Gel de agarosa al 1,5% posterior a la fragmentación de ADN por la acción de la dsDNA Fragmentase durante 5,5 minutos. STD: estándar de peso molecular (1kb en extremo izquierdo y 100pb en extremo derecho); pb: pares de bases	58
Figura 12: Gel de agarosa al 1,5% posterior a la purificación doble del ADN fragmentado. STD: estándar de peso molecular (1kb en extremo izquierdo y 100 pb en extremo derecho)	59
Figura 13: Electroferograma de TapeStation para tres librerías genómicas de ADN mitocondrial con la distribución de tamaños de fragmentos. Se destaca el peak de la distribución, mientras que los peaks en 25 pb y 1500 pb corresponden a estándares internos	60
Figura 14: Histograma de la calidad promedio de las secuencias.....	62
Figura 15: Nivel de duplicación de las secuencias	62
Figura 16: Visor del programa IGV con variantes en región del mitogenoma de la muestra SER117. Se incluye el track del alineamiento (BAM), la secuencia de referencia (NC_012920_1) y se indica la profundidad de cobertura para cada sitio nucleotídico ..	63
Figura 17: Número de lecturas mapeadas y no mapeadas para cada muestra	64
Figura 18: Gráfico de la cobertura en pb de las lecturas de secuenciación en genoma mitocondrial de referencia para dos muestras a) y b).	66
Figura 19: Gráfico de barras apiladas con la proporción de haplogrupos mitocondriales de origen euroasiático en las poblaciones chilenas	67
Figura 20: Gráfico de barras apiladas con la proporción de haplogrupos mitocondriales de origen africano en las poblaciones chilenas	71
Figura 21: Dendrogramas Neighbor-Joining construidos a partir de matrices de distancias de FST pareadas considerando las frecuencias de 211 haplotipos estimados a partir de las secuencias d-loop (a) y las frecuencias para 27 haplogrupos (b)	77
Figura 22: Análisis de Componentes Principales de las poblaciones chilenas (a). En (b) se representan las variables sobre los componentes. Este análisis se realizó en base a frecuencias de haplogrupos.....	79

Figura 23: Análisis de Componentes Principales de las poblaciones chilenas sin considerar el haplogrupo L (a). En (b) se representan las variables sobre los componentes. Este análisis se realizó en base a frecuencias de haplogrupos.....	81
Figura 24: Red de haplotipos de haplogrupos euroasiáticos en poblaciones chilenas ..	84
Figura 25: Red de haplotipos de macrohaplogrupo L en poblaciones chilenas.....	86
Figura 26: Red de haplotipos de haplogrupos asiáticos en poblaciones chilenas.....	87
Figura 27: Análisis factorial de datos mixtos para los grupos de linajes mitocondriales africanos, europeos y asiáticos (a). Se indican las variables cuantitativas sobre las dimensiones 1 y 2 (b).....	89
Figura 28: Análisis de Componentes Principales de poblaciones chilenas y europeas, incluidas dos poblaciones de Medio Oriente, a partir de frecuencias de haplogrupos mitocondriales euroasiáticos. En a) se grafican los componentes 1 y 2 y en b) los componentes 4 y 5	92
Figura 29: Red de haplotipos del macrohaplogrupo HV en poblaciones chilenas y europeas. Ver Anexo Tabla S5 para detalle de muestras europeas consideradas.....	94
Figura 30: Análisis de Componentes Principales de poblaciones chilenas y africanas, incluida la población europea de Andalucía, a partir de frecuencias de haplogrupos mitocondriales L-U6. Más detalle sobre las regiones a la que pertenecen los países en Figura 1 y Anexo Tabla S4. Se utilizó la función <i>Repel</i> para facilitar la lectura de las poblaciones en el plot	95
Figura 31: Red de haplotipos del macrohaplogrupo L en poblaciones chilenas y africanas. Ver Anexo Tabla S6 para detalle de las muestras africanas consideradas.....	98
Figura 32: Árbol filogenético de consenso de mayoría obtenido por medio de inferencia bayesiana para linajes mitocondriales no amerindios en Chile, a partir de mitogenomas. Los valores de los nodos representan en orden: la probabilidad posterior del análisis bayesiano y el soporte bootstrap de los análisis de máximo likelihood y parsimonia. El color de las ramas simboliza los haplogrupos asignados y a la derecha se indican los haplogrupos mayores.....	107
Figura 33: Mapas de isofrecuencia por método de <i>Kriging</i> para 3 linajes europeos H en base a la frecuencia de haplogrupos en poblaciones de la Península Ibérica y Chile ...	113

Figura 34: Mapas de isofrecuencia por método de <i>Kriging</i> para el linaje europeo HV0 en poblaciones de la Península Ibérica y de Chile, y para el haplogrupo V en la Península Ibérica	114
Figura 35: Mapas de isofrecuencia por método de <i>Kriging</i> para linaje africano L0a en base a su frecuencia en poblaciones de África y Chile	116
Figura 36: Mapas de isofrecuencia por método de <i>Kriging</i> para linajes africanos L1b y U6a en base a sus frecuencias en poblaciones de África y Chile	117
Figura 37: Mapas de isofrecuencia por método de <i>Kriging</i> para los linajes africanos L1c y L3e en base a sus frecuencias en poblaciones de África y Chile	119
Figura 38: Mapas de isofrecuencia por método de <i>Kriging</i> para linaje europeo K en base a su frecuencia en poblaciones de Europa y Chile	121
Figura 39: Mapas de isofrecuencia por método de <i>Kriging</i> para linaje europeo T2 en base a su frecuencia en poblaciones de Europa y Chile	122
Figura S1: Gráfico de barras con la contribución de cada variable al componente 1 (a) y 2 (b). PCA realizado a partir de las frecuencias de 27 haplogrupos	147
Figura S2: Gráfico de barras con la contribución de cada variable al componente 1 (a) y 2 (b). PCA realizado a partir de las frecuencias de 20 haplogrupos (sin haplogrupo L)	148
Figura S3: Análisis de Componentes Principales para poblaciones chilenas y españolas. Se utilizó la división regional de Barral-Arca y col. (2016)	149
Figura S4: Árbol filogenético realizado en Haplogrep a partir de mitogenomas	150
Figura S5: Árbol de consenso de mayoría obtenido por el análisis de Máxima Parsimonia para los linajes mitocondriales no amerindios en Chile, partir de mitogenomas. Se indican los valores de bootstrap	151
Figura S6: Árbol filogenético de Máximo Likelihood para los linajes mitocondriales no amerindios en Chile, a partir de mitogenomas. Se indican los valores de bootstrap	152
Figura S7: Mapas de isofrecuencias por método kriging para 5 linajes africanos. La simbología del nivel de frecuencia debe considerarse de manera separada	153
Figura S8: Mapas de isofrecuencias por método <i>kriging</i> para 3 linajes europeos. La simbología del nivel de frecuencia debe considerarse de manera separada	154

RESUMEN

La historia migratoria en el territorio que hoy conocemos como Chile ha estado marcada por diversos procesos socioculturales y demográficos. Aquí destacamos tres momentos principales que corresponden a la conquista europea, el tráfico trasatlántico de esclavos y las migraciones recientes de los siglos XIX y XX. En este contexto, la figura de la mujer como individuo y colectivo ha sido tradicionalmente invisibilizado. En este estudio se utilizó el ADN mitocondrial como marcador de herencia materna para explorar la diversidad mitocondrial no amerindia en distintas poblaciones chilenas con el objetivo de dar a conocer la participación de las mujeres migrantes en las dinámicas poblacionales de cada momento histórico. Se analizaron 520 muestras pertenecientes a 18 poblaciones de Chile y se aplicaron análisis de estructura poblacional a partir de frecuencias de haplogrupos y secuencias D-loop. Se realizaron comparaciones con poblaciones de Europa y África. Además, se secuenciaron y procesaron 45 mitogenomas obtenidos por NGS para análisis filogenéticos. La diversidad de linajes mitocondriales no amerindios, analizada a partir de frecuencias, F_{ST} , análisis de componentes principales y redes de haplotipos, evidenció una estructuración genética de las poblaciones chilenas, con diferencias latitudinales en la distribución de los linajes. Se distingue a las poblaciones del norte y centro norte de Chile (L(xL1b), HV0, U8b) de las demás poblaciones de centro y sur del país (H2a, K1a, T2b). En el caso de las poblaciones del centro y sur del país se observó una mayor frecuencia de linajes europeos que son frecuentes en la Península Ibérica. En el norte y centro de Chile se reportó una mayor diversidad y frecuencia de linajes africanos asociados al proceso de tráfico de esclavas bajo la gobernación del

Virreinato del Perú. Estos linajes africanos disminuyen hacia el sur, pero en Chillán y zonas aledañas se encontró un linaje africano particular que se asoció con un efecto fundador y que provino en el genoma de españolas durante el periodo de conquista y colonia. Por último, en el sur de Chile se observó una alta diversidad de linajes europeos y se relacionan con las migraciones de los siglos XIX y XX. Estos resultados fueron congruentes con los procesos descritos desde la historiografía. Sin embargo, al comparar con poblaciones europeas y africanas, la alta diversidad de dichas poblaciones limitó el estudio de los posibles patrones migratorios y origen geográfico de los linajes. A pesar de ello, se pudo concluir que el territorio chileno fue receptor de mujeres migrantes que se establecieron en el país, tuvieron descendencia y fueron parte de las dinámicas poblacionales que dieron origen a las poblaciones mestizas del país, reflejando una combinación particular de linajes mitocondriales no amerindios en las poblaciones. Finalmente, este estudio contribuyó en dar a conocer el aporte foráneo en la composición genética materna de las poblaciones en Chile.

ABSTRACT

The migratory history in the territory we know today as Chile has been marked by various sociocultural and demographic processes. Here we highlight three main moments corresponding to the European conquest, the transatlantic slave trade, and the recent 19th and 20th centuries migrations. In this context, women's figure as an individual and collective has traditionally been made invisible. In this study, mitochondrial DNA was used as a marker of maternal inheritance to explore non-Amerindian mitochondrial diversity in different Chilean populations to reveal the participation of migrant women in the dynamic populations of each historical moment. 520 samples belonging to 18 Chilean populations were analyzed, and population structure analyzes were applied based on haplogroup frequencies and D-loop sequences. Comparisons with populations from Europe and Africa were made. In addition, 45 NGS-derived mitogenomes were sequenced and processed for phylogenetic analyses. The diversity of non-Amerindian mitochondrial lineages, analyzed from frequencies, FST, principal component analysis, and haplotype networks, showed a genetic structuring of the Chilean populations, with latitudinal differences in the distribution of the lineages. The populations of the north and central north of Chile (L(xL1b), HV0, U8b) are distinguished from the other populations of the center and south of the country (H2a, K1a, T2b). In the case of populations from the center and south of the country, a higher frequency of European lineages was observed, which are frequent in the Iberian Peninsula. In northern and central Chile, greater diversity and frequency of African lineages associated with the slave trade process was reported under the governorate of the Viceroyalty of Peru. These African lineages diminish towards the

south, but in Chillán and surrounding areas, a particular African lineage was found that was associated with a founder effect, and that arose in the genome of Spanish women during the period of conquest and colony. Finally, in southern Chile, a high diversity of European lineages was observed and is related to the migrations of the 19th and 20th centuries. These results were consistent with the processes described from historiography. However, compared with European and African populations, the high diversity of these populations limited the study of the possible migratory patterns and geographic origin of the lineages. Despite this, it was possible to conclude that the Chilean territory was the recipient of migrant women who settled in the country, had offspring, and were part of the dynamic populations that gave rise to the country's mestizo populations, reflecting a particular combination of mitochondrial lineages non-Amerindians in the towns. Finally, this study contributed to recognize the foreign contribution to the maternal genetic composition of the populations in Chile.

1. INTRODUCCIÓN

La migración es un fenómeno que ha caracterizado a la especie humana y es responsable de la formación de poblaciones mixtas con componentes genéticos diversos. A pesar de ello, actualmente suele poseer una connotación negativa y la historia tradicional se ha encargado de invisibilizar la participación de mujeres, personas indígenas y negras en la constitución de las poblaciones mestizas chilenas. En los últimos años, nuevos antecedentes han permitido reescribir la historia dando a conocer el rol de las mujeres inmigrantes en este proceso. Las evidencias historiográficas también muestran las principales oleadas migratorias que recibió el territorio chileno en el contacto inicial, así como aquellas más recientes. Por otra parte, a partir de herramientas moleculares como el uso de marcadores uniparentales, en especial el ADN mitocondrial, es posible analizar la composición genética de las poblaciones e inferir y caracterizar procesos migratorios. En Chile, los estudios con este marcador se han enfocado en el componente amerindio, mientras que el aporte foráneo en la composición genética materna de nuestro país no se ha explorado.

A continuación, se exponen en mayor detalle estos antecedentes, sección que se divide en antecedentes históricos y genéticos. En primer lugar, se profundiza en la participación de las mujeres migrantes durante tres momentos de la historia del país: conquista y colonia europea, tráfico de esclavos y oleadas migratorias en el siglo XIX y XX. Posteriormente se presentan los antecedentes genéticos que reflejan cómo estos procesos históricos han moldeado la composición genética de las poblaciones.

Al final de este capítulo de Introducción se plantea el problema de investigación

junto a la hipótesis y los objetivos que guiaron el desarrollo de esta tesis.

1.1. Antecedentes

1.1.1. Antecedentes históricos

1.1.1.1. Mujeres migrantes durante la conquista y colonia europea en América y Chile

La historia de las poblaciones nativa-americanas cambió radicalmente con la llegada, conquista y colonización española en el continente. La ambición por nuevos territorios y riquezas motivó la exploración del llamado “Nuevo Mundo” por parte de flotas ibéricas a partir del siglo XV. El contacto con estos grupos de inmigrantes generó una serie de consecuencias a nivel biológico, sociocultural y demográfico, siendo el mestizaje el proceso que daría origen a las poblaciones mestizas sudamericanas. El cruce se caracterizó por ser asimétrico, dado entre mujeres indígenas y hombres europeos, siendo excluido el hombre indígena de este proceso de mestizaje (Ruiz-Linares, 2015; Manríquez y col., 2016; Moraga, Pezo y de Saint Pierre, 2016). Este fenómeno se puede explicar por la llegada de un gran número de hombres europeos en comparación a las mujeres, y las relaciones de violencia y superioridad establecidas con los indígenas. Si bien el número de mujeres españolas en el proceso de conquista fue menor, la historia tradicional se ha encargado de invisibilizarlo. En las últimas décadas, nuevos estudios han repensado la participación de la mujer en la conquista y colonización de América (Sullivan-Beare, 1956; McEwan, 1991; Powers, 2005; Guardia, 2013; Gómez, 2014; Córdoba, 2018; Bisselle, 2019).

Durante las expediciones de Colón se constató que en el segundo viaje embarcó una mujer desde España (León, 2007) y así, desde los primeros viajes hacia América se registran a mujeres como parte de la tripulación, quienes acompañaban a sus cónyuges, familiares o viajaban por cuenta propia (Gómez, 2014; León, 2014). La Corona a principios del siglo XVI pretendió fomentar la migración femenina hacia las Américas, otorgando permisos, subsidios y órdenes de los gobernadores coloniales para traer mujeres con sus familias a los nuevos territorios (Gómez, 2014). Era necesario que los nuevos territorios del trono castellano tuviesen familias poblándolos para consolidar su conquista, siendo así que en 1530 Carlos I decretaría que los viajes hacia América hechos por españoles debían incluir a sus esposas e hijos en caso de tenerlos (Córdoba, 2018), arriesgándose a multas o la pérdida de posesiones en caso de no hacerlo (López, 2013).

La escasa información que menciona a las mujeres en esta época está contenida principalmente en cartas enviadas por los españoles a las mujeres de su familia, quienes eran alentadas a migrar hacia América (Gómez, 2014; López, 2013). El llamado real a hacer efectiva la colonización movilizó en sus inicios a españoles de distintas clases sociales y oficios como agricultores, mercaderes, soldados, entre otros, y financió en ocasiones el viaje de sus familias. En cuanto a cifras, se estima que la emigración femenina aumentó de un 6,3% a un 16,4% del total de migrantes españoles entre los años 1520-1539 y 1540-1559 (Bisselle, 2019). Hacia finales del siglo XVI se contabilizaban 45.327 viajeros registrados, donde 10.118 serían mujeres (Boyd-Bowman, 1976 en Córdoba, 2018). Para el periodo 1560-1579 se contabilizaron 5.013 mujeres que emigraron a las Américas, donde un 60% de estas (3.024) estaban en la categoría de solteras, en algunos casos con niños. Entre los años 1580 y 1600 se tienen registro de 2.472 mujeres migrantes

de España, provenientes en su mayoría de Andalucía (1.476 andaluzas) (Martínez, 1983).

En el caso del territorio chileno, este proceso se inició en el siglo XVI con la llegada de conquistadores como Diego de Almagro quien recorrió infructuosamente los territorios al sur de Copiapó siendo recibido con constantes ataques por los habitantes de la zona (Téllez, 1984). Con la expedición de Pedro de Valdivia, se definió a Santiago como el centro y capital del Reino de Chile, donde se concentró gran parte de la población migrante. En el norte de lo que hoy es Chile, la región que contemplaba al desierto de Atacama fue considerada como una zona aislada, pobre en recursos e incluso de peligro, por lo que no obtuvo el interés de los conquistadores españoles en los primeros siglos de la colonia (Hidalgo, 2012), mientras que las poblaciones de la costa fueron rápidamente reducidas a la lógica colonial (Bittman, 1977; Ballester, San Francisco y Gallardo, 2010). Por su parte, en la región al sur del río Bío-Bío, el pueblo mapuche mantuvo su resistencia sin dejar que los españoles conquistaran y consolidaran el orden colonial en esta zona (Bengoa, 2000). Durante los conflictos que se extendieron desde el siglo XVI al siglo XVIII se habría propiciado el secuestro y cautiverio de mujeres de los bandos beligerantes. Los primeros registros de mujeres españolas siendo tomadas como prisioneras de manera sistemática surgen en el 1568 (Cerdeña, 1988).

En cuanto a la procedencia de los primeros conquistadores, ésta fue variada, siendo *“oriundos de muchos sitios de la península ibérica: Andalucía, Extremadura, la región Vasca y las dos Castillas y León”* (Cussen, 2016, pp. 22). A medida que se expandía el imperio español, italianos, portugueses y filipinos también migraron hacia América como vasallos del rey, aunque en menor número (Cussen, 2016).

Las mujeres españolas en Chile se encuentran presentes desde los inicios de la época de la conquista, siendo Inés de Suarez una de las figuras más emblemáticas de este periodo. Sin embargo, al momento de hablar de las mujeres como colectivo existe una falta de registros que narren cuantas mujeres participaron de las expediciones iniciales de la conquista y las cifras estimadas por los historiadores resultan dispares. Parte de los registros sitúan la presencia de mujeres al sur de Santiago hacia 1554, al momento de hablar del grupo conquistador que se asienta en la recién fundada Concepción, contabilizando entre 20 y 30 mujeres entre los habitantes de la ciudad (Cerde, 1988). De acuerdo con Encina (1940) un total de 50 mujeres vivían en el Reino de Chile hacia el año 1583 en comparación a 1.100 hombres estimados. Vergara (1981), por su parte, reconoce un total de 165 mujeres y más de 550 hombres españoles hasta el año 1570. A partir de 1570 se observa una disminución de la migración de españolas y europeas, aumentando el número de descendientes nacidas en América y Chile (Maldonado, 2019).

1.1.1.2. Mujeres migrantes durante el tráfico de esclavos en América y Chile

Durante el periodo colonial también atravesaron el Atlántico esclavos africanos, quienes fueron forzados a migrar desde los primeros momentos de la colonia como parte de los bienes personales de los conquistadores españoles (Cussen, 2016). Entre los siglos XVI y XIX, se estima que cerca de 10 millones de hombres, mujeres y niños desembarcaron en los puertos de América central, el Caribe y Sudamérica (Eltis y Richardson, 2010). Si bien en los primeros años de la trata de esclavos, el número y procedencia de quienes eran traídos a América estaba limitado a aquellos que hubiesen nacido de padres cristianos en la península (por el temor a la difusión de herejías entre los

indígenas del continente), a principios del siglo XVI la Corona española finalmente cedió a la migración masiva de esclavos negros ante la disminución de población indígena y la falta de mano de obra en las plantaciones y centros mineros (Mellafe, 1984). En cuanto a los lugares de procedencia de esclavos, la principal región de embarque fue la costa de África occidental, comprendiendo la zona entre el río Senegal y Kunene (Mellafe, 1984; Vial, 1957).

La composición étnica y origen geográfico de los esclavos que llegaron a América es consistentemente limitada, producto de las circunstancias en las que se llevó a cabo esta práctica (Fortes-Lima, 2015). Las bases de datos sobre el origen de los esclavos se centraban en los puertos de donde salían las embarcaciones de la costa africana, sin considerar el origen de las personas que transportaban. La reconstrucción histórica sobre el lugar de embarque de los esclavos que llegaron a América (Figura 1) ha permitido establecer que desde 1501 a 1866, alrededor de un 46.3% de los esclavos provenían de la Costa Centro Occidental de África, principalmente desde lo que hoy es Angola, junto a la zona del Congo y Gabón. Un 16.12% de los esclavos provenían de la Ensenada de Benín (o también llamada Costa de Esclavos), que actualmente comprende las costas de Togo, Benín y Nigeria. Un 12.31% de la Ensenada de Biafra, abarcando lo que hoy son las costas de Nigeria, Camerún y Gabón. Un 9,6% de la Costa de Oro que hoy corresponde a Costa de Marfil y Ghana. Para la región de Senegambia (Senegal y Gambia) un 5,7% de los esclavos embarcaron de esa zona, un 3,2% de Sierra Leona y un 2,7% de la Costa de Barvolento (también conocida como la Costa del Grano) que hoy comprende Costa de Marfil y Liberia. Por último, un 4,1% embarcó del Sudeste de África (Eltis y Richardson, 2010, 2013). Durante las primeras décadas de 1800 frente a la alta demanda de esclavos

africanos en Sudamérica se expandió la búsqueda de personas hacia Mozambique y el sudeste africano, convirtiéndose en la segunda fuente de esclavos africanos para el 1830, después de la Costa Occidental (Eltis y Richardson, 2013).



Figura 1: Regiones históricas africanas asociadas a la trata de esclavos con sus respectivos límites políticos actuales. (Fuente: Eltis y Richardson, 2010).

Para el caso de Chile, los primeros registros se remontan a Hernando de Magallanes, quien fue el primer europeo en pisar territorio chileno durante 1520 sin intento de colonización (Martinić, 1977, 2002). En su viaje se documenta la presencia de esclavos negros, africanos libres y moriscos (descendientes de musulmanes españoles) que habrían visto el país por primera vez, tripulantes que se vuelven usuales en las

siguientes expediciones hacia Chile (Vial, 1957). Tanto en la hueste de Diego de Almagro como de Pedro de Valdivia se registra la presencia de esclavos, pero no es posible estimar un número, aunque se cree sea menor a 10 en el último caso (Vial, 1957). Sin embargo, se pudo identificar la unión de conquistadores con esclavas negras, ya que se sabe que 97 fundadores dejaron una descendencia tanto legítima como no reconocida de 392 hijos, de los cuales 7 corresponden a mulatos y negros para los cuales se conoce que dos nacieron de una madre negra (Mellafe, 1984).

En general, los registros de la época consignan poco detalle sobre los africanos esclavizados, ya que por lo general figuran como “criados”, lo que impide diferenciarlos de los indígenas y estimar una cifra real de su representatividad en el país. Tampoco existió un interés en registrar los lugares de origen de éstos (Cussen, 2016; Vial, 1957). De todas formas, se estima una cifra de cerca de 6.000 hombres y mujeres de África subsahariana que fueron forzados a migrar a Chile entre los siglos XVI y XIX (de Ramón, 2000). La diferencia en la proporción por sexo no fue tan pronunciada como en el caso de las migrantes españolas y se estima que en general la presencia africana femenina fue de un tercio (Mellafe, 1984). En la figura 2 se muestra la proporción de esclavos africanos por sexo y edad que fueron llevados a América desde las distintas regiones de África y coincide con lo descrito para Chile con un porcentaje de mujeres cercano al 30%. La composición de género de los esclavos durante este tiempo se ha caracterizado tradicionalmente en ser principalmente masculina, lo que se ha explicado por las demandas de los dueños de plantaciones a lo largo de las Américas de trabajadores fuertes y resistentes. Sin embargo, la literatura moderna ha mostrado que hasta el 1800 existía una alta migración forzada de mujeres y niños a lo largo de las regiones africanas (Fortes-

Lima, 2015).

Con respecto a las rutas utilizadas para traer esclavos hacia Chile se distinguen 3, la primera corresponde al viaje directo desde España acompañando a nobles, clérigos y burócratas (Vial, 1957). La segunda sería la ruta del Pacífico que conecta Cartagena de Indias, Portobelo, Guayaquil, Lima y Valparaíso (Mellafe, 1984), donde por mar o tierra se movilizaban hasta Lima donde podían ser comprados por personas del Reino de Chile (Cussen, 2016). La tercera y más utilizada es aquella conocida como continental que atraviesa Buenos Aires, Córdoba, Mendoza y Santiago, donde debían cruzar la cordillera de los Andes (Cussen, 2016).

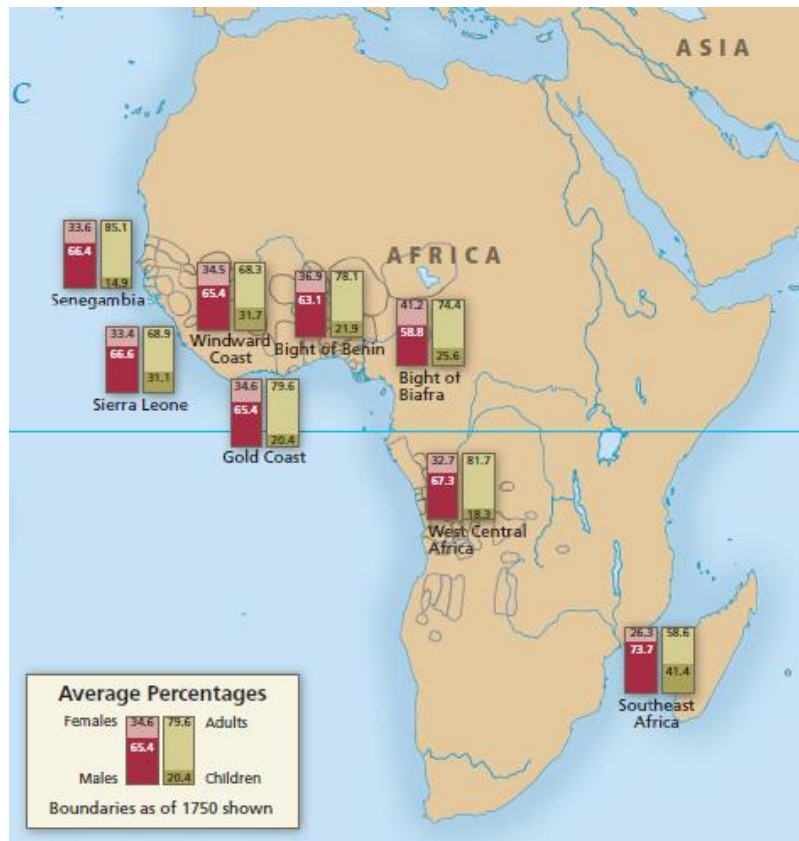


Figura 2: Mapa de género y edad de esclavos traídos a América entre 1545-1864 (Fuente: Eltis y Richardson, 2010).

Entre 1580 y 1640 se produjo el mayor auge de migración de esclavos hacia Chile cuando se necesitaba mano de obra para el trabajo rural, quienes en su mayoría provenían del Congo, Angola y Guinea y eran movilizados a través de los Andes (Mellafe, 1984). Distintas empresas marítimas se disputaban el tráfico de esclavos, siendo Portugal quien tenía el monopolio hasta que luego de la Guerra con España (1640-1668) cesara su tráfico con varias colonias de América. Durante el siglo XVIII, las potencias francesa (1702-1713) e inglesa (1713-1738) obtuvieron el monopolio del negocio esclavista en América en sus respectivos años en los que se les concedieron los “asientos” (Mellafe, 1984). De esta forma, Chile quedó entre Lima y Buenos Aires, transformándose en ruta de paso, ya que debido a que el Virreinato del Perú mandó a llamar a esclavos para trabajar en las minas y plantaciones, gran parte de los esclavos desde Buenos Aires y Chile tuvo que migrar hacia la zona norte del país que en esa época formaba parte del virreinato del Perú (Mellafe, 1984, Vial, 1957). Finalmente, en 1823 se abolió la esclavitud negra en Chile por la moción de José Miguel Infante (Feliú, 1973).

1.1.1.3. Mujeres migrantes durante los siglos XIX y XX en Chile

Posterior a los procesos por la independencia y de la anexión, por parte del Estado Chileno, de los territorios que hoy corresponden a las regiones de Arica y Parinacota, Tarapacá y Antofagasta a consecuencia de la Guerra del Pacífico (1879-1883), se vivió un periodo de importantes flujos migratorios transfronterizos y extra-continetales. Estos fenómenos migratorios también ocurrieron en Sudamérica, donde políticas gubernamentales fomentaron el arribo de europeos, con la idea de que contribuirían al progreso del país en las décadas finales del siglo XIX (Cussen, 2016). Además, la inestabilidad económica y

política en sus lugares de origen más las repercusiones de la Primera Guerra Mundial impulsaron en gran medida las distintas oleadas migratorias (González, Lufin y Galeno, 2016). Fue en el año 1907 cuando la tasa de inmigración en Chile alcanzó su cifra más alta correspondiendo a un 4,2% y se registraban 134.524 extranjeros.

A partir de datos censales de los años 1907 y 1930 se elaboraron dos gráficos para ilustrar la proporción de inmigrantes mujeres en las distintas provincias de Chile (Figuras 3 y 4). Considerando solo países fuera de América, se observa que a lo largo del país son las españolas quien poseen una mayor presencia, junto a mujeres de Francia, Gran Bretaña, Inglaterra y Alemania. En cuanto a las diferencias por género, estas oleadas migratorias se caracterizaron por poseer un patrón masculinizado, pero si se revisan las cifras censales (INE, 1908, 1931) se observa que las mujeres representan aproximadamente la mitad de la cantidad de hombres migrantes.

De acuerdo con la bibliografía, el norte de Chile en el contexto de la minería del salitre y el cobre recibió a migrantes europeos como ingleses, franceses, italianos, españoles, croatas, y del continente asiático a chinos y árabes (González, 2011; González y col., 2016). En ciudades de esta zona como Antofagasta e Iquique y en Punta Arenas destaca la presencia de migrantes de Austria-Hungría/Yugoslavia. En estas ciudades, la migración croata (como parte del Imperio Austro-Húngaro y posteriormente Yugoslavia) se hizo presente en las últimas décadas del siglo XIX, pero posterior a la Primera Guerra Mundial arriba otra oleada numerosa cuyo peak se observa en el 1925 y declina luego de 1930 (Zlatar, 2002).

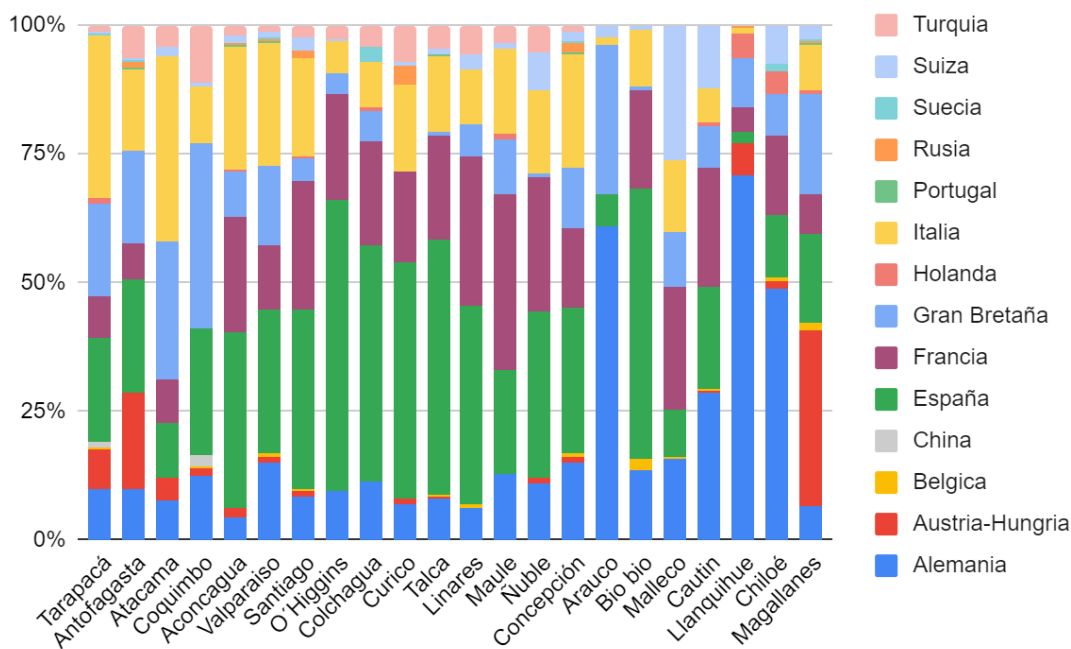


Figura 3: Proporción de migrantes mujeres de distintas nacionalidades en Chile en el año 1907. Elaboración propia a partir del censo (INE, 1908).

Para el sur de país, el Estado de Chile impulsó la ocupación de estos territorios siguiendo intereses económicos y de soberanía en dicha zona a partir de 1850. De esta forma inició un proyecto de colonización con pobladores proveniente de Europa y que tuvo dos etapas. La primera ocurrió entre 1846 y 1875 con la llegada de inmigrantes alemanes en las ciudades de Valdivia y Puerto Montt (Tillería y Vela, 2017). Esta ocupación temprana promovió la expansión a nuevos lugares como las regiones de los Lagos y la Araucanía (Marínez, 2020). En ambas figuras 3 y 4 se observa la presencia mayoritaria de alemanes en las provincias de Arauco, Cauquín, Llanquihue y Chiloé. Luego, con la creación de agencias oficiales como privadas se inicia la segunda etapa donde se reclutaron inmigrantes para los distintos proyectos de colonización en los territorios de la Araucanía, Chiloé, Aysén y Magallanes, hasta la isla Navarino. En esta etapa se captó a

colonos de diversas nacionalidades europeas, entre ellas, principalmente inmigrantes españoles, franceses, italianos, suizos, ingleses y alemanes (Tillería y Vela, 2017). En el caso de los franceses, además del sur, también se concentraron en Santiago, Valparaíso y Concepción (Fernández, 2006). Los británicos se asentaron en Santiago, Punta Arenas, Valparaíso y también fueron prioridad para la colonización de Chiloé junto con colonos alemanes (Agar, 2015; Marínez, 2020).

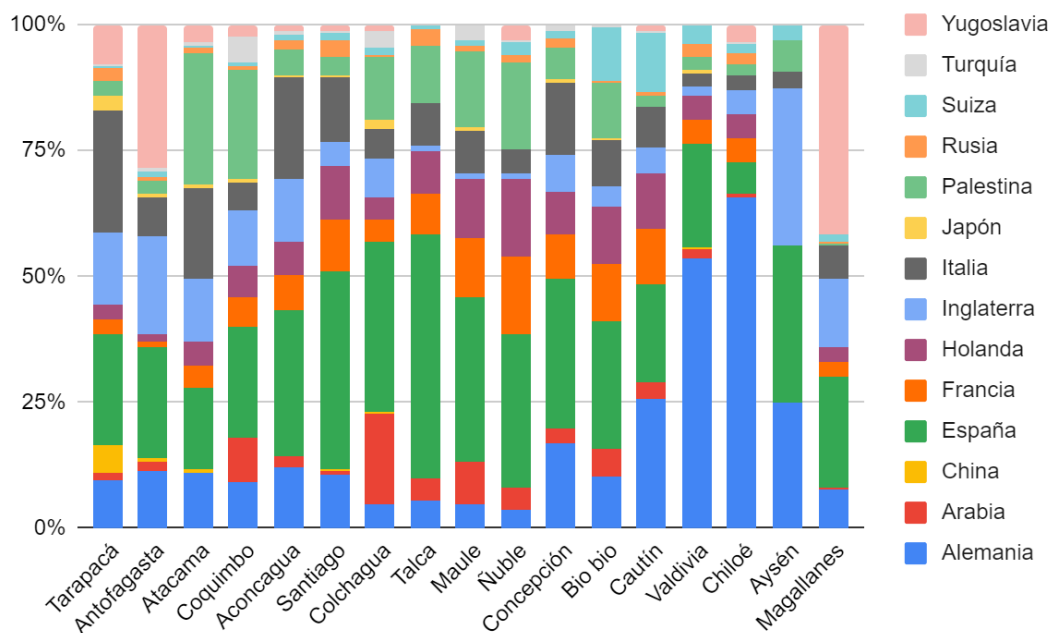


Figura 4: Proporción de migrantes mujeres de distintas nacionalidades en Chile en el año 1930. Elaboración propia a partir del censo (INE, 1931).

1.1.2. Antecedentes genéticos

1.1.2.1. Composición genética de las poblaciones sudamericanas

A nivel genético, las poblaciones latinoamericanas reflejan esta mezcla asimétrica entre tres grupos ancestrales: nativo americano, europeo y africano, este último con un

menor aporte en general. Para analizar las diferencias en las proporciones de cada uno a nivel de genoma completo, la principal estrategia ha correspondido al uso de AIMs (*ancestry informative marker*), polimorfismos que presentan variación en la frecuencia de sus alelos entre las distintas poblaciones que permite inferir origen geográfico (Galanter y col., 2012). En Sudamérica, los valores promedio para cada componente varían entre las distintas poblaciones, encontrando un mayor aporte africano en poblaciones de Brasil y Colombia que lo observado en Argentina o Chile. En poblaciones de Argentina se observó un mayor aporte europeo al igual que en México, Costa Rica, mientras que las poblaciones rurales de Colombia y Chile estudiadas muestran un mayor componente amerindio (Adhikari y col., 2016).

En Chile, se ha estimado en promedio un 44,3% de componente amerindio, 51,8% de europeo y 3,8% de africano (Fuentes y col., 2014). Un estudio contemplando un análisis de GWAS, entregó cifras similares con un 44,7% de contribución amerindia, 52,7% europea y 3,01% africana (Eyheramendy y col., 2015) y recientemente, un panel de SNPs aplicado en individuos de distintas poblaciones a lo largo del país sigue estimando en promedio una proporción de ancestría europea de 53% (± 14), 42% (± 14) de ancestría amerindia y un 4% (± 4) de ancestría africana (Verdugo y col., 2020). Los tres estudios concuerdan con que existen algunas diferencias entre las distintas zonas del país, donde el centro de Chile concentra los valores más altos de ancestría europea. Por su parte, la ancestría africana mostró ser mayor en el norte del país y ambas observaciones pueden ser explicadas a partir de los antecedentes históricos mencionados previamente (Fuentes y col., 2014; Eyheramendy y col., 2015).

Las poblaciones de Latinoamérica también exhiben lo que se conoce como sesgo

por sexo en el mestizaje, manifestado en que la mayor parte de los linajes maternos son de origen amerindio, mientras que los linajes paternos son en su mayoría de origen europeo, debido al cruce asimétrico entre mujer indígena y hombre europeo (Ruiz-Linares, 2015). En este sentido, los marcadores uniparentales (ADN mitocondrial y Cromosoma Y) han mostrado ser una herramienta fundamental para el estudio de la genética de poblaciones, ya que la estructura genética de las diferentes poblaciones puede ser estudiada para inferir procesos migratorios, identificar el origen étnico o ancestría y con ello indagar en la historia de las mismas.

1.1.2.2. ADN mitocondrial como marcador poblacional

El ADN mitocondrial (Figura 5) posee una estructura circular de doble hebra con una longitud de 16.569 pares de bases según lo descrito en la primera secuencia completa obtenida por Anderson y col. (1981). Se encuentra en múltiples copias al interior de la mitocondria, organelos citoplasmáticos donde ocurre la producción de energía y el proceso de fosforilación oxidativa de la célula (Jobling y col. 2013). Esta molécula de ADN contiene 37 genes que codifican para 13 proteínas, 22 ARNs de transferencia y 2 ARN ribosomales (Anderson y col., 1981). La región codificante abarca cerca del 90% de la molécula, mientras que el resto corresponde a la región control, denominada así por su participación en el control de la replicación y transcripción del genoma. También conocida como D-loop, esta región se extiende desde las posiciones 16024 a 16569 pb y de 1 a 576 pb, comprendiendo cerca de 1100 pb. Para esta región se han descrito 3 regiones hipervariables, siendo HVSI (16024-16365 pb) y HVSII (73-340 pb) las más variables (Chinnery, 2006).

El ADN mitocondrial se caracteriza por poseer una herencia vía materna, ser multicopia, no sufrir recombinación genética y tener una alta tasa de mutaciones (las cuales se van acumulando a lo largo del tiempo), siendo estas características las que despertaron el interés por este marcador en estudios genéticos evolutivos (Kivisild, 2015).

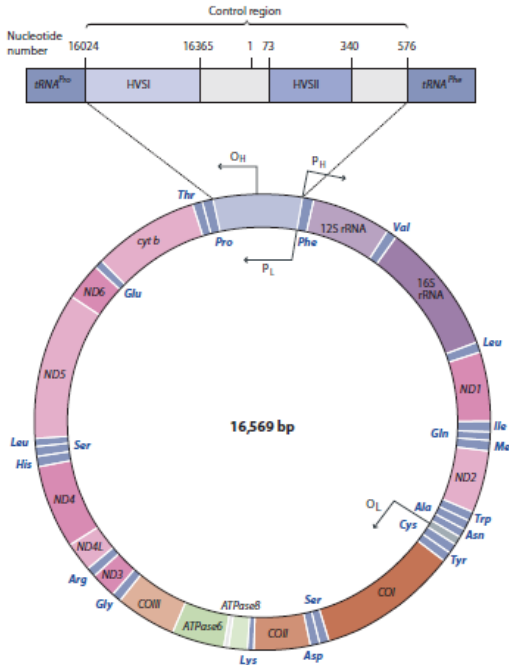


Figura 5: Estructura de la molécula de ADN mitocondrial. Se indican los genes, los orígenes de replicación para la cadena pesada y liviana y hacia superior se destaca la región control o d-loop (Fuente: Jobling y col., 2013).

La elevada tasa de mutación descrita para el genoma mitocondrial se ha calculado entre 5 a 10 veces mayor que el ADN nuclear (Brown y col., 1979). A su vez, la tasa de cambio dentro del ADN mitocondrial no es homogénea, siendo la región control la que evoluciona a una tasa mayor que la región codificante. Es más, la región control no es heterogénea y posee regiones más variables que otras (Castro y col., 1998).

La determinación de haplogrupos y su concepto surge en la década de los 90 y se

utiliza para referirse a secuencias que están relacionadas filogenéticamente, al poseer mutaciones compartidas y presentar patrones de distribución geográfica (Pakendorf y Stoneking, 2005). Los primeros haplogrupos se describieron en población nativa americana, los cuales llevan las letras A-D (Torroni y col., 1992). Estudios posteriores han definido los demás haplogrupos en los distintos continentes, encontrando al haplogrupo L en la raíz de la filogenia del ADN mitocondrial humano y presente en población africana. A partir del linaje L3 surgen dos grandes grupos, los haplogrupos M y N que dan cuenta del cuello de botella experimentado en la diversidad del ADN mitocondrial durante la dispersión fuera de África del *Homo sapiens* anatómicamente moderno. El tiempo en que ocurrió la separación de estos dos linajes se ha estimado hace 62 a 95 miles de años atrás (Kivisild, 2015). Los macrohaplogrupos M y N estuvieron involucrados en el poblamiento del resto del mundo, con M distribuyéndose principalmente por Asia y N en el continente europeo. Del haplogrupo N derivó el linaje R, donde 3 de sus ramas principales (haplogrupos H, JT y U) corresponden aproximadamente al 80-90% de la diversidad mitocondrial de Europa (Pereira y col. 2005).

Los linajes mitocondriales que presentan una distribución en toda Sudamérica son los considerados fundadores donde encontramos a A2, B2, C1, D1 y D4h3a (Achilli y col., 2008; Tamm y col., 2007). Sus distribuciones no son homogéneas y en el caso chileno, los macrohaplogrupos presentan una distribución de gradiente latitudinal, donde las frecuencias más altas de A2 y B2 se concentran en el norte del país, las que disminuyen hacia el sur, donde se vuelven más frecuentes los haplogrupos C1 y D1/D4h3a (Gómez-Carballa y col., 2016).

En poblaciones mestizas de Chile, cerca del 88% corresponde a haplogrupos amerindios (Gómez-Carballa, y col., 2016), mientras que el resto se caracteriza por linajes europeos, asiáticos y africanos. Dentro de los linajes europeos, el más frecuente es el haplogrupo H y representa cerca del 40-50% de la variación del ADN mitocondrial en ese continente (Achilli y col., 2004; Roostalu y col., 2007; Badro y col., 2013). Este haplogrupo tiene altas frecuencias en Europa occidental, zona donde es posible constatar una subestructura geográfica en la Península Ibérica. Mientras el haplogrupo H está altamente representado en el noroeste de la península (con las frecuencias más altas en Galicia, 52% en la costa de Cantabria y 45-49% en los territorios de Portugal), estas cifras comienzan a disminuir hacia el mediterráneo, con frecuencias de 37-39% en el Mediterráneo y Andalucía y 32% en la meseta Ibérica (Barral-Arca y col., 2016). Dentro de la Península Ibérica también se observan otros patrones de distribución geográfica con la prevalencia del haplogrupo J hacia el noroeste y el país vasco, mientras que el haplogrupo T es el más frecuente en la región Mediterránea (Barral-Arca y col., 2016). El haplogrupo U es el segundo linaje más frecuente en Iberia con una distribución principalmente homogénea, aunque alcanza la frecuencia más alta en el norte de Portugal con un 26% (Barral-Arca y col., 2016).

Además de los haplogrupos ya mencionados, en poblaciones europeas también podemos encontrar en menor frecuencia los linajes HV, K, W, I, V y X. En el centro de Europa, las frecuencias de los haplogrupos T y U5, le siguen a las del haplogrupo H (Cocoş y col., 2017). Es importante mencionar que, dentro de H, sus sublinajes también presentan variación geográfica siendo H1 frecuente en el oeste de Europa especialmente la región escandinava y el sur de la Península Ibérica, H3 es el segundo más común con

una prevalencia en el sur de Europa, principalmente Francia y España, mientras que los subclados H2 y H6 se han asociado con poblaciones del este de Europa y Asia (Achilli y col. 2004; Pereira y col., 2005; Roostalu y col., 2007; Chinnery y Gómez-Durán, 2018).

En cuanto a los linajes mitocondriales de origen africano, estos corresponden principalmente al haplogrupo L y sus sublinajes que se han adscrito a distintas regiones de África. Estos presentan un patrón de distribución característico, ya que en África subsahara estos linajes representan el 92% de la diversidad mitocondrial, mientras que en el norte solo corresponden al 17% (Cerezo y col., 2016). En el norte de África podemos encontrar los haplogrupos L3e5 y L3f3 que también son los linajes L3 más frecuentes en el centro oeste del continente, además de los linajes M1 y U6 particulares al norte de África que se asocia a migraciones desde el este de África y Medio Oriente (Cerezo y col., 2016). El haplogrupo L0 es común en el este, centro y sureste de África (Gonder y col., 2007), siendo muy diverso en el este y sureste y disminuyendo su diversidad en el centro de África occidental. El linaje L1, en cambio, presenta sus frecuencias más altas en el sur y centro oeste africano (Cerezo y col., 2016). Éste posee dos linajes principales, L1b y L1c, el primero con una distribución asociada a la costa occidental de África y el segundo con una distribución más amplia por el continente. El haplogrupo L2 posee una alta diversidad en el centro de África occidental (Cerezo y col., 2016) y el subclado L2c se ha encontrado en poblaciones mestizas de Argentina y Brasil (Catelli y col., 2011). Por su parte, el haplogrupo L3 posee frecuencias de 45% en el suroeste de África, 43% en la región centro occidental y un 34% hacia el este (Cerezo y col., 2016). En poblaciones brasileñas afrodescendientes, el haplogrupo más frecuente resultó ser el L3 con un 38,2%, en su mayoría pertenecientes a L3e (Schaan y col., 2017).

La proporción de linajes no amerindios en Chile varía a lo largo del país. En promedio, el estudio de Gómez-Carballa y col. (2016) encontró un 11,9%, siendo caracterizados por los haplogrupos H, U, T, T2, K, K1 y L, donde los linajes africanos fueron contabilizados en un 1,1%. Según Vieira-Machado y col. (2016), valores más altos de haplogrupos no amerindios se registraron en la comuna de Independencia (20%) y Santiago (25%), mientras que el resto de las poblaciones analizadas en el sur de Chile no supera el 10% y no se describe el detalle de los linajes encontrados. Recientemente, se constató en la región de Antofagasta la presencia de linajes africanos en una frecuencia mayor a lo descrito hasta el momento, con un 5,7% en la costa norte de la región y un 4,5% en Calama y Chuquicamata, representados principalmente por los haplogrupos L2 y L3 (de Saint Pierre, Faure y Morales, 2020).

1.2. Problematización

Considerando lo expuesto en antecedentes y el contexto actual del país, ha quedado en evidencia cómo la sociedad ha invisibilizado y negado la participación de minorías como las mujeres y personas negras en la conformación de la población mestiza chilena actual. Estos últimos años se ha intentado reescribir la historia desde una perspectiva feminista dando a conocer el rol de las mujeres indígenas, europeas y africanas en este proceso y a partir de la genética es posible aportar con evidencia. Por otro lado, en Chile los estudios con ADN mitocondrial se han enfocado en el componente amerindio, mientras que el aporte foráneo en la composición genética materna de nuestro país no se ha explorado.

Teniendo en cuenta lo anterior, caracterizar y analizar la diversidad de linajes mitocondriales no amerindios en poblaciones mestizas a lo largo del país, permitirá aportar en la reconstrucción de la historia materna a nivel país, por medio de una mayor comprensión de las dinámicas de mestizaje y los movimientos migratorios de mujeres europeas, africanas y asiáticas que aportaron al pool genético de Chile. Para ello se planteó la siguiente pregunta de investigación: ¿Cómo se relaciona la distribución de linajes mitocondriales no amerindios en Chile con los procesos migratorios descritos desde la historiografía?

1.3. Hipótesis

Considerando: a) la presencia de mujeres españolas en el proceso de conquista y colonización de Chile y los lugares de origen de los conquistadores que se embarcaron junto a sus mujeres y familiares; b) la presencia de mujeres esclavas en Chile colonial y tanto las políticas gubernamentales como las rutas de tráfico de esclavos establecidas en la época; y c) la procedencia de las oleadas migratorias recientes y su asentamiento en las distintas zonas del país a finales de siglo XIX e inicios del XX. Se propone la siguiente hipótesis:

Las mujeres que llegaron al territorio chileno durante los procesos migratorios descritos desde la historiografía se establecieron en las distintas ciudades del país y fueron parte de las dinámicas poblacionales que dieron origen a las poblaciones mestizas chilenas, dejando una impronta a nivel genético caracterizada por una distribución geográfica diferencial de linajes mitocondriales no amerindios a lo largo de Chile.

Además, se proponen las siguientes predicciones:

1. Las poblaciones mestizas del centro de Chile poseen una mayor frecuencia de linajes mitocondriales provenientes de la Península Ibérica debido a los procesos iniciales del mestizaje.
2. Las poblaciones mestizas del norte de Chile poseen una mayor cantidad de linajes mitocondriales de origen africano producto del tráfico de esclavas.
3. Las poblaciones mestizas de Chile presentan una distribución geográfica diferencial de linajes mitocondriales europeos no ibéricos que da cuenta de los procesos migratorios post-mestizaje ocurridos durante los siglos XIX y XX.

1.4. Objetivos

1.4.1. Objetivo General

Analizar los distintos patrones migratorios durante el proceso de mestizaje y post-mestizaje en Chile por parte de mujeres no nativas.

1.4.2. Objetivos Específicos

1. Caracterizar la diversidad de linajes mitocondriales no amerindios en diversas poblaciones mestizas de Chile.
2. Determinar la distribución geográfica que poseen los linajes mitocondriales restringidos a regiones específicas de Europa, África y Asia en diversas poblaciones mestizas de Chile.
3. Inferir patrones migratorios a partir de los linajes mitocondriales no amerindios presentes en diversas poblaciones de Chile y lo descrito desde la historiografía.

4. Examinar la filogenia de los linajes mitocondriales no amerindios con mayor representación en la muestra mediante la secuenciación de mitogenomas.

2. MATERIALES Y MÉTODOS

2.1. Muestra

Las muestras analizadas corresponden a un total de 520 individuos portadores de haplogrupos mitocondriales no amerindios residentes de distintas ciudades a lo largo de Chile. En la tabla 1 se detalla el tamaño muestral para cada población y el porcentaje de linajes no amerindios con respecto a la muestra total. La mayor parte de las muestras fueron obtenidas por el Proyecto Chile Genómico y caracterizadas dentro del proyecto Fondecyt #1181889, por lo que se accedió directamente a las secuencias de D-loop completo en formato fasta. De igual forma, se accedió a las secuencias en archivo fasta de aquellas muestras caracterizadas por tesis de estudiantes del Laboratorio de Genética de Poblaciones y Evolución Humana, Facultad de Medicina, U. de Chile. En el caso de San Felipe-Los Andes y Concepción no se contaba con las secuencias, ya que las muestras fueron caracterizadas a través de PCR-RFLP para los macrohaplogrupos americanos. Por ello se procedió a amplificar y secuenciar la región control del ADN mitocondrial de aquellas muestras que fueron identificadas como no amerindias. En el próximo apartado se describe el protocolo utilizado.

Adicionalmente se incluyeron datos obtenidos de la literatura como aquellas de Faure (2018) y Gómez-Carballa y col. (2016). En ambos casos se contaba con los sitios variantes a la secuencia de referencia de Cambridge (rCRS, Andrews y col., 1999) para las regiones hipervariables I y II, por lo que se convirtieron a secuencia fasta a través de la herramienta *mtSNP conversion tool* de mtDNAprofiler (Yang y col. 2013, <http://mtprofiler.yonsei.ac.kr/>). Dado que las poblaciones reportadas por Gómez-Carballa

(2016) ya estaban contempladas en este estudio, se decidió unificar las muestras para los análisis posteriores, constituyendo las poblaciones de Iquique (n = 28), Santiago (n = 126), Concepción (n = 43), Temuco (n = 82) y Punta Arenas (n = 46).

Tabla 1: Poblaciones chilenas incluidas en este estudio.

Población	N no amerindio	N total	% no amerindio	Referencia
Arica	25	234	10,68%	Proyecto Chile Genómico
Azapa	1	45	2,22%	Apata, 2017
Iquique	17	210	8,09%	Proyecto Chile Genómico
Antofagasta	18	253	7,11%	Faure, 2018
Calama	6	171	3,5%	Faure, 2018
La Serena	12	191	6,28%	Proyecto Chile Genómico
Coquimbo	11	177	6,21%	Proyecto Chile Genómico
Aconcagua	7	100	7%	Orellana, 2014
San Felipe-Los Andes	18	204	8,82%	Leiva, 2010
Santiago	90	679	13,25%	Proyecto Chile Genómico
Curicó	5	43	11,26%	Silva, 2021
San Carlos	9	143	6,29%	Castillo, 2021
Chillán	60	486	12,34%	Proyecto Chile Genómico
Concepción	19	231	8,22%	Pezo, 2010
Temuco	58	549	10,56%	Proyecto Chile Genómico
Osorno	7	119	5,88%	Guerrero, 2019
Puerto Montt	16	126	12,69%	Proyecto Chile Genómico
Punta Arenas	19	123	15,44%	Flores, 2016
<i>Iquique</i>	11	200	5,5%	Gómez-Carballa, 2016
<i>Santiago</i>	36	203	17,7%	Gómez-Carballa, 2016
<i>Concepción</i>	24	201	11,94%	Gómez-Carballa, 2016
<i>Temuco</i>	24	194	12,37%	Gómez-Carballa, 2016
<i>Punta Arenas</i>	27	197	13,7%	Gómez-Carballa, 2016

Además de los datos mitocondriales, para 296 muestras del Proyecto Chile Genómico se contó con las proporciones de ancestría europea, africana y nativa, ésta

última distinguiéndose entre un componente norte y componente centro-sur de Chile. Estos datos para 9 poblaciones chilenas fueron reportados por Verdugo y col. (2020) y Flores-Alvarado y col. (2021). Los valores de ancestría individual se estimaron a partir de la genotipificación de CLG, un panel de 150 marcadores informativos de ancestría (AIMs) diseñado por el Proyecto Chile Genómico y el posterior análisis en el programa ADMIXTURE (Verdugo y col., 2020).

Las muestras del Proyecto Chile Genómico y del Laboratorio de Genética de Poblaciones y Evolución Humana cuentan con el consentimiento informado de los participantes, que fue aprobado por el Comité de Ética de Investigación en Seres Humanos, Facultad de Medicina, Universidad de Chile.

2.2. Técnicas de recolección y procesamiento de datos

2.2.1. Secuenciación Sanger de D-loop completo

A partir de las muestras de ADN ya extraído previamente para las muestras de San Felipe - Los Andes y Concepción, se amplificó por PCR la región control (D-loop) del ADN mitocondrial que contempla las posiciones 16024 - 576 pb. Cada reacción de PCR para un volumen de 25 μ l incluyó 2 μ l de ADN, 0,2 μ l de GoTaq G2 Flexi polimerasa (Promega), 5 μ l de buffer 5x de la enzima, 2 μ l del mix de dNTPs (2,5 mM c/u), 1,5 μ l de MgCl₂ a una concentración final de 1,5 mM, 1,25 μ l de cada partidor (Tabla 2) a una concentración de 10 μ M y 11,8 μ l de agua libre de nucleasas. Una vez en el termociclador, se utilizó un programa de PCR con las siguientes condiciones: desnaturalización inicial a

94°C por 3 minutos, 35 ciclos de desnaturalización a 94°C, alineamiento a 60°C y extensión a 72°C por 45 segundos cada ciclo, más una extensión final a 72°C por 5 min.

Tabla 2: Partidores utilizados para la amplificación del D-loop del ADN mitocondrial con sus respectivas temperaturas de alineamiento (T°a).

Marcador	Partidores	Posición	Secuencia (5' - 3')	T° a	Referencia
Región control (D-loop)	F15792 (F)	15792-15811	tcattggacaagtagcatcc	60°C	Moraga y col. (2000)
	M14 (R)	708-727	agggtgaactcactggaacg	60°C	

Los productos de PCR se visualizaron a través de electroforesis en un gel de agarosa al 1,5%. Posteriormente, la purificación y secuenciación fue realizada por Macrogen Inc, Corea del Sur, utilizando los partidores que se describen en la Tabla 3.

Tabla 3: Partidores utilizados para la secuenciación del D-loop de ADNmt.

Partidor	Posición	Secuencia (5' - 3')	Referencia
M1	15978-15997	caccattagcacccaaagct	Laboratorio Uchile
R698	677-698	gcatgtgtaacttactaagag	Handt y col. 1996
F16475	16475-16495	tagctaaagtgaactgtatcc	Bailliet y col. 1994
M3	8-30	ggctatcacctattaaccact	Laboratorio Uchile
M3rev	8-29	gtggtaaatagggtgatagacc	Laboratorio Uchile
F314	314-333	ccgcttctggccacagcact	Brandstätter y col. 2004
MM6	16303-16322	tgcttatgtactatgtac	Laboratorio Uchile

Una vez obtenidas las secuencias, se revisaron manualmente en el programa Geneious (Kearse y col., 2012, <http://www.geneious.com>) para el proceso de edición donde se eliminaron las zonas de baja calidad. Las secuencias para cada muestra se alinearon a la región control de la secuencia de referencia de Cambridge (Andrews y col.,

1999) formando un contig. Luego se verificaron los sitios polimórficos y se editó cuando fue necesario, para finalmente exportar la secuencia consenso en formato FASTA.

2.2.2. Secuenciación de Nueva Generación (NGS) de mitogenomas

Con la finalidad de caracterizar los linajes mitocondriales no amerindios a una mayor resolución, se secuenció el genoma mitocondrial completo de 45 individuos. Para la selección de las muestras se consideraron aquellos haplogrupos más frecuentes y/o que presentaran una distribución restringida geográficamente. Se priorizaron linajes africanos y europeos y en algunos casos se incluyó más de una muestra por haplogrupo, pero con distinto haplotipo.

Los haplogrupos no amerindios pesquisados para la secuenciación de mitogenomas fueron los siguientes: D4p, H1c, H1cf, H1e1a1, H2a2a, H3p, H5a3a1, HV0, J1c2e2, K1a4a1a+195, K1b1a2, L0a1b1, L0d1c, L1b1a+189, L1c1'2'4'6, L2a1+143, L2c3, L2d+16129, L3d2b, L3e1a, L3e1d, L3e1f*, N1b1, R0a+60.1T, T2b3+151, T2b7a3, U4c1, U5a1b1, U5b2a2, U8b1a2b, W1+119, W3a1+199, X2d. Debido a la calidad de algunas muestras, la elección se fue ajustando a medida que se realizaban los trabajos de laboratorio y no fue posible incluir algunos linajes contemplados inicialmente como D4b, D4h, H4a1a+195, L1c3a1b y U6.

La obtención de los mitogenomas por NGS involucró distintas etapas de preparación de las muestras como la construcción de librerías, y posteriormente el procesamiento de los datos resultantes de la secuenciación. La figura 6 resume el flujo de trabajo seguido para este propósito y pretende guiar la lectura de los procedimientos que se detallan a continuación.

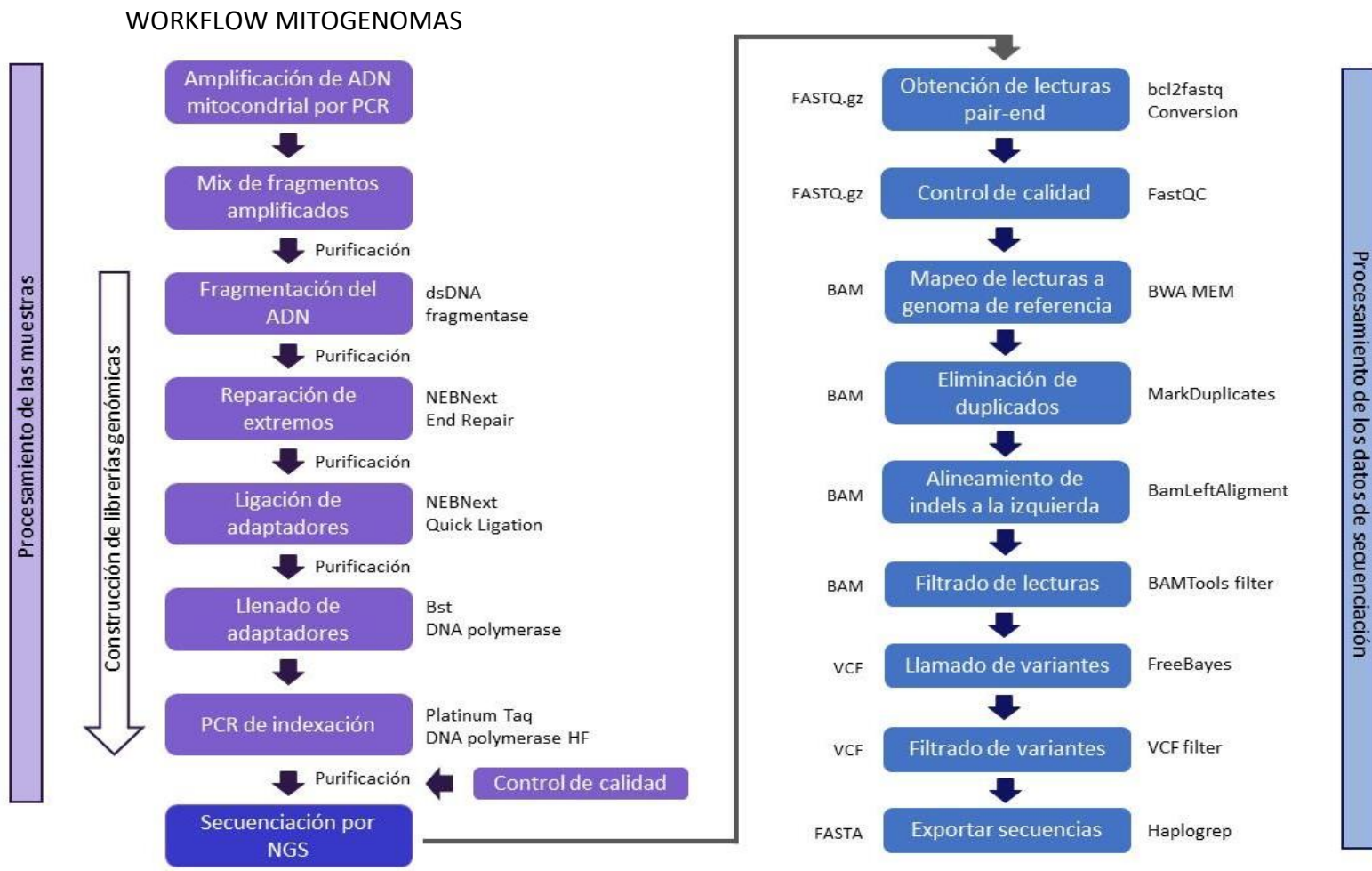


Figura 6: Flujo de trabajo de la secuenciación NGS de mitogenomas. A la izquierda se presentan los procedimientos de laboratorio para la preparación de las muestras, mientras que a la derecha se señalan los análisis bioinformáticos para el procesamiento de los datos. Además, al lado derecho de cada recuadro se indica el reactivo o herramienta utilizada en esa etapa y al lado izquierdo el formato del output generado.

2.2.2.1. Evaluación y diseño de protocolo de amplificación por PCR

Actualmente existe un amplio repertorio de técnicas para la obtención del genoma mitocondrial. Entre ellas, la secuenciación de nueva generación (NGS) de amplicones de ADN mitocondrial generados por PCR es considerada una de las más robustas (Legati y col., 2021; Zhang, Cui y Wong, 2012). Se contempló la evaluación de tres protocolos de amplificación por PCR de largo alcance que generan fragmentos sobrelapados de tamaños superiores a los 5.000 pares de bases. Se utilizó la enzima TaKaRa LA Taq DNA polimerasa (TaKaRa) que ha sido optimizada para la amplificación de fragmentos largos. En la tabla S1 de Anexos se detallan los sets de partidores y el tamaño de los fragmentos para los tres protocolos que han sido reportados por otros estudios (Gonder y col. 2007 en Fendt y col. 2009; Illumina Sequencing Platform, 2016).

Debido a que estos protocolos fueron puestos a prueba en el laboratorio previo al desarrollo de esta tesis, se procedió evaluando aquellos que entregaron mejores resultados inicialmente (Protocolos 1 y 2 de la tabla S1 en Anexos). En ambos casos no se observó una uniformidad en la amplificación, es decir, que uno o más set de partidores en cada protocolo no resultó adecuadamente. Se decidió diseñar un nuevo protocolo que permitiera generar más fragmentos de menor tamaño y que pudieran ser amplificados con la enzima GoTaq G2 Flexi polimerasa (Promega). Para ello se revisaron los partidores de ADN mitocondrial disponibles en el laboratorio y se escogieron 7 pares de partidores que amplificaran para 7 fragmentos de tamaño similar y que se sobrelaparan en al menos 200 pb, permitiendo cubrir el genoma mitocondrial completo. En la tabla 4 y figura 7 se detallan los partidores que dieron mejores resultados y que fueron finalmente

seleccionados. Se indica su secuencia y posición, junto al tamaño de los fragmentos obtenidos que van desde los 3030 a 3520 pares.

Se realizaron 7 reacciones de PCR para cada muestra y el protocolo fue el mismo descrito para la amplificación del D-loop, a excepción de que se agregaron 3 µl de ADN. En el termociclador se utilizó un programa con las siguientes condiciones: desnaturalización inicial a 94°C por 3 minutos, 35 ciclos de desnaturalización a 94°C por 45 segundos, alineamiento (entre 55° y 59°C, ver tabla 4) por 45 segundos y extensión a 72°C por 4 minutos, más una extensión final a 72°C por 7 min.

Tabla 4: Partidores utilizados en la amplificación de fragmentos para la secuenciación de mitogenomas con sus respectivas temperaturas de alineamiento (T° a).

Fragmento	Partidor	Secuencia (5' - 3')	Posición	Referencia	Tamaño fragmento	T° a
1	M13	gtagcttacctcctcaaagcaa	583-604	Moraga y col. 2000	3030 pb	59°
	3547R	aggcctaggtgaggtgac	3594-3613	Laboratorio Uchile		
2	H2698	ttgacctgccctgaagagg	2679-2698	Laboratorio Uchile	3520 pb	58°
	6119R	tttatgcggggaaacccat	6180-6199	Laboratorio Uchile		
3	M11	cctaactactaccgcattccta	5099-5120	Moraga y col. 2000	3275 pb	59°
	L8351	ttggggcatttactgtaaagagg	8351-8374	Laboratorio Uchile		
4	H7801	ctatctgcccgccatcatc	7782-7801	Laboratorio Uchile	3429 pb	58°
	11177R	aggaagtatgtgctgcgtt	11192-11211	Laboratorio Uchile		
5	FampB	gccatactagtctttgccgc	10653-10672	Gonder y col. 2007	3433 pb	59°
	14049R	tgcccttttgggtgaggtga	14066-14086	Laboratorio Uchile		
6	M7	cgcccttacacaaatgacatcaa	13209-13232	Moraga y col. 2000	3389 pb	59°
	M3rev	gtggttaatagggtgatagacc	8-29	Laboratorio Uchile		
7	MTL-F2	tatccgccatcccatacatt	15195-15214	Illumina Platform	3266 pb	55°
	MTL-R1	ttgctctccttgcaaagtt	1873-1892	Illumina Platform		

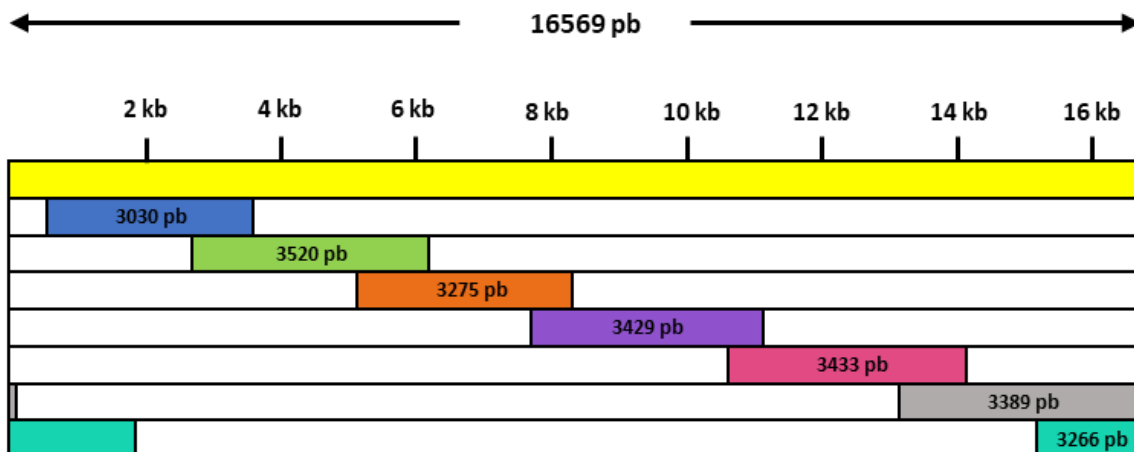


Figura 7: Ubicación de los siete fragmentos amplificados por PCR para la secuenciación del genoma mitocondrial completo. Se posicionan siguiendo el orden en que aparecen en la tabla XX y se indica el tamaño de cada uno.

2.2.2.2. Cuantificación y purificación de ADN

Dada la gran cantidad de reacciones, no fue posible cuantificar ni purificar todos los productos de PCR, así que se optó por escoger alrededor de 45 productos contemplando los 7 fragmentos para cuantificar la concentración de ADN a través de Qubit Fluorometric Quantification Broad Range (ThermoFisher). Para el resto se estimó comparando la intensidad de las bandas en el gel con el estándar de peso molecular de 1Kb (ThermoFisher) y las bandas de los productos PCR cuantificados. Con estas estimaciones, se calculó el volumen en μ l necesario de cada fragmento por muestra para obtener un mix lo más equimolar posible.

Se procedió a purificar cada mix de 7 fragmentos, con la finalidad de eliminar las bandas inespecíficas generadas en algunas reacciones de PCR. Para esto, se empleó el kit ProNex® Size-Selective Purification System (Promega), sistema de purificación basado en resina magnética que permite seleccionar el tamaño de dsDNA a retener o descartar.

En este caso se eliminaron los fragmentos menores a 1000 pb equivalente a una proporción de ProNex® Chemistry (v/v) del 1X, por lo que se aplicó el protocolo disponible en el Manual Técnico (Promega Corporation, 2017-2018). Para ello, se pipetearon 50 µl de muestra (mix) en un tubo de 0,2 ml y se mezcló con 50 µl de ProNex® Chemistry y al momento de eluir el ADN, se agregó 50 µl de Elution Buffer a cada muestra. Luego de seguir todos los pasos indicados, la eficiencia de la purificación se verificó a través de un gel de agarosa al 1,5%.

2.2.2.3. Construcción de librerías genómicas

Fragmentación de ADN:

Los análisis de secuenciación de nueva generación en plataforma Illumina requieren de un input pequeño, por lo que el protocolo de preparación de librerías contempla la fragmentación del ADN. Para ello se utilizó la NEBNext dsDNA Fragmentase (New England Biolabs), un reactivo en base a enzima que rompe el ADN para producir fragmentos de tamaño deseado dependiente del tiempo. Esta enzima genera fragmentos que son cohesivos de 1 a 5 pares de bases en los extremos 5' y una mezcla de cohesivos y romos en 3'.

Se realizaron pruebas para estimar el tiempo de acción de la fragmentasa y se hicieron ajustes a las condiciones de laboratorio para minimizar el tiempo empleado en accionar y detener la enzima. Se definió un tiempo de 5,5 minutos que permitía obtener fragmentos que se concentraban en los 500 pb. Se realizó una primera tanda de 34 muestras purificadas, siguiendo el protocolo de la enzima que se describe a continuación:

Primero se debe vortexear la dsDNA Fragmentase por 3 segundos, realizar un spin rápido y poner en hielo. Luego, para cada reacción se mezclan los siguientes componentes en un tubo PCR y vortexear: 16 μ l de ADN (se agregó el volumen máximo de ADN propuesto por el protocolo) y 2 μ l de Buffer 10X Fragmentase Reaction, completando un volumen de 18 μ l. Se agregaron 2 μ l de dsDNA Fragmentase a cada muestra y se vortexea por 3 segundos. Las muestras se incubaron a 37°C por 5,5 minutos y se agregaron 5 μ l de 0.5M EDTA para detener la reacción. Finalmente, se recomienda purificar el ADN fragmentado usando SPRI beads. Para el último paso, se purificó nuevamente con el kit ProNex® Size-Selective Purification System (Promega), pero esta vez se realizó una purificación doble que produce una población de fragmentos de ADN centrado en un tamaño deseado, eliminando fragmentos por sobre y debajo del rango escogido. Se siguió el protocolo con una proporción de 1,05/0,35X de ProNex® Chemistry con el objetivo de obtener fragmentos centrados en 450-500 pb y eliminar los fragmentos de tamaño superior a las 1000 pb e inferior a 200 pb (Promega Corporation, 2017-2018). Para ello, los volúmenes utilizados fueron 20 μ l de cada muestra en un tubo de 0,2 ml y luego se mezclaron 21 μ l de ProNex® Chemistry (es decir: 1,05:1 v/v de ProNex® Chemistry a la muestra). Siguiendo el protocolo de purificación doble, se mezclaron 7 μ l adicionales de ProNex® Chemistry en el sobrenadante de cada muestra (es decir: 0,35:1 v/v de ProNex® Chemistry a la muestra). Finalmente, el ADN se eluyó en 20 μ l de Elution Buffer.

Se evaluó el resultado de la purificación doble en 10 muestras a través de un gel de agarosa al 1,5%, cargando 15 μ l de cada una. También se cuantificó la concentración de ADN con Qubit Fluorometric Quantification Broad Range (ThermoFisher) en 6 muestras adicionales.

Reparación de los extremos:

Los fragmentos de ADN generados se repararon en sus extremos para convertirlos en fragmentos de extremos romos, utilizando el módulo NEBNext End Repair (New England BioLabs). La baja concentración de ADN obtenida en las muestras, posterior a la fragmentación y purificación impidió que se aplicara el protocolo recomendado. Se procedió a utilizar un protocolo aplicado previamente en el laboratorio que dio buenos resultados para bajas cantidades de ADN como es el caso de muestras de ADN antiguo:

Se preparó un mix con 2,5 µl de NEBNext 10x End Repair Buffer, 1,25 µl de NEBNext End Repair Enzyme Mix y 3,7 µl de agua libre de nucleasas. Se agregaron 7,5 µl de mix por muestra con 17,5 µl de ADN para completar un volumen final de 25 µl. Las muestras se incubaron por 20 minutos a 12°C seguido de 15 minutos a 37°C.

Las muestras se purificaron con el mismo kit ProNex® Size-Selective Purification System (Promega) siguiendo el protocolo para purificación simple descrito anteriormente, pero con la intención de eliminar los fragmentos de ADN menores a 100 pb. Para ello se consideró una razón de 3X, que implicó agregar 75 µl de ProNex Chemistry a los 25 µl de muestra. Finalmente, el ADN se eluyó en 18 µl de Elution Buffer.

Ligación de adaptadores:

El proceso de ligación de los adaptadores compatibles con la plataforma de secuenciación Illumina a los fragmentos de ADN con los extremos reparados se realizó por medio del módulo NEBNext Quick Ligation (New England BioLabs). El protocolo adaptado a bajas cantidades de ADN fue el siguiente: Se mezclaron los siguientes reactivos: 5 µl de NEBNext Quick Ligation 5x Buffer, 0,5 µl de mix adaptadores de ADN

(25uM), más 2 µl de agua libre de nucleasas. Se agregaron 7,5 µl del mix a cada muestra con 15 µl de ADN. Al final se agregaron 2,5 µl de la enzima NEBNext Quick T4 ligase separadamente a cada tubo, para completar un volumen final de 25 µl. Las muestras se incubaron en el termociclador por 20 minutos a 20°C. Se siguió el mismo protocolo de purificación descrito para el paso anterior, pero se eluyó en 26 µl.

Llenado de adaptadores:

Para esta reacción se utilizó la enzima Bst DNA polymerase (New England BioLabs) que es capaz de moverse entre hebras, permitiendo llenar los nicks y completar la secuencia de doble hebra de los adaptadores. El protocolo utilizado fue el siguiente: Se preparó un mix con 3 µl de Buffer de la enzima Bst, 0,5 µl de dNTPs (10 mM c/u), 1,5 µl de agua libre de nucleasas y 1,5 µl de la enzima Bst DNA polymerase. De éste, se repartieron 6,5 µl a cada muestra con 23,5 µl de ADN para completar un volumen de 30 µl por reacción. Las muestras se incubaron por 20 minutos a 65°C, seguido por 20 minutos a 80°C. En esta oportunidad no se requirió un paso de purificación, así que las librerías fueron almacenadas directamente a -20°C.

PCR de indexación:

Para finalizar el proceso de construcción de librerías, fue necesario multiplexar las muestras a través de la indexación por PCR. En este paso se agrega un par de índices único a cada muestra, los que actúan como códigos de barras y permiten correr paralelamente más de una librería en el flowcell de secuenciación y luego es posible identificarlas en los softwares de análisis de datos de secuenciación. Para esta reacción se utilizó la enzima Platinum Taq DNA Polymerase High Fidelity (Invitrogen, Thermo Fisher Scientific) y se

incorporó una combinación única de índices para cada muestra. Para esto, se contó con los primeros 6 índices (i5) y 8 índices (i7) Truseq DNA (Illumina), los que generan 48 combinaciones posibles que fueron suficientes para cubrir las 45 muestras analizadas. Cada reacción de PCR para un volumen de 50 µl incluyó los siguientes componentes: 10 µl del templado de los fragmentos más adaptadores, 0,2 µl de la enzima Platinum Taq High Fidelity, 5 µl de Buffer 10x de la enzima, 2 µl de MgCl₂ a una concentración final de 2 mM, 1 µl del mix de dNTPs (2,5 mM c/u), 2 µl de cada índice a una concentración de 10 µM y 27,8 µl de agua libre de nucleasas. En el caso de los índices, estos se agregaron primero de manera separada a cada muestra. Una vez en el termociclador, se utilizó un programa de PCR con las siguientes condiciones: desnaturalización inicial a 94°C por 2 minutos, 9 ciclos de desnaturalización a 94°C, alineamiento a 60°C y extensión a 68°C por 30 segundos en cada ciclo, más una extensión final a 72°C por 5 min. Por último, se realizó una purificación simple siguiendo el protocolo descrito previamente, pero en esta ocasión se decidió eliminar los fragmentos que fueran de tamaños inferiores a 250 pb (proporción entre ProNex Chemistry y muestra de 1,5X). Para ello, se mezclaron 75 µl de ProNex Chemistry en el volumen total obtenido en el paso anterior y al final se eluyó en 30 µl de Elution Buffer.

2.2.2.4. Control de calidad y secuenciación de las librerías genómicas

Las librerías genómicas se cuantificaron por medio de Qubit Fluorometric Quantification High Sensitivity (ThermoFisher) y posteriormente se evaluó la calidad de la construcción de librerías a través del análisis High Sensitivity D1000 ScreenTape Assay

en el equipo de TapeStation 2200 (Agilent Technologies). Los resultados se visualizaron en el programa de TapeStation Analysis (A.01.04).

Las librerías genómicas construidas presentaron las condiciones necesarias para ser secuenciadas. Este proceso se llevó a cabo en la Unidad de Secuenciación y Bioinformática de la Universidad de Santiago con el secuenciador NextSeq 550 System Mid-Output. Se secuenció con la técnica *paired end sequencing* y dadas las características de este equipo Illumina se generaron pares de lecturas de una longitud máxima de 150 pb. El proceso de demultiplexación se realizó en la misma institución a través del programa bcl2fastq Conversion Software (Illumina), lo que permitió separar las lecturas de cada muestra en dos archivos FASTQ, R1 y R2. Además, en esta etapa se eliminaron los index y adaptadores utilizados en la construcción de las librerías.

2.2.2.5. Procesamiento de lecturas obtenidas por secuenciación NGS

Los datos obtenidos a partir de NGS requieren de distintas etapas de procesamiento que se describen a continuación. Para ello se trabajó en la plataforma web Galaxy a través del servidor público usegalaxy.org (The Galaxy Community, 2022) y se utilizó como guía el tutorial para llamado de variantes en genomas haploides recomendado por la comunidad Galaxy (<https://galaxyproject.org/tutorials/var-hap/>).

Control de calidad de lecturas de secuenciación:

Como primer paso se aplicó un análisis de control de calidad de las lecturas de secuenciación que permite conocer la calidad de las secuencias y detectar si existen problemas asociados a la secuenciación o a la construcción de librerías que se deben

considerar antes de proseguir con los análisis. Se utilizó el software FastQC versión 0.11.9 (Andrews, 2010, <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) y se generaron reportes HTML para cada muestra. Con el fin de evaluar las estadísticas generales de la secuenciación, los resultados obtenidos por FastQC se ingresaron en MultiQC v. 1.11 (Ewels y col., 2016, <https://multiqc.info/>) para generar un reporte único resumen de los parámetros de calidad para todas las muestras.

Alineamiento de lecturas a genoma de referencia:

Una vez realizado y revisado el control de calidad, se procedió al alineamiento de las lecturas contra el genoma de referencia. Para este paso se utilizó el alineador BWA (Burrows Wheeler Aligner) v. 0.7.17 (Li y Durbin, 2009), uno de los más utilizados en datos obtenidos con Illumina y en específico, el algoritmo MEM (Li, 2013) que ha mostrado mayor precisión para lecturas entre los 70 y 1000 pb. BWA-MEM funciona maximizando las coincidencias exactas en el alineamiento, conservando las ventajas del algoritmo Smith y Waterman (SW) y la velocidad de análisis de BWA con una estrategia conocida como “*seeding and reseeding*”, donde se buscan fragmentos de la lectura que coincidan exactamente entre ésta y la referencia denominados como ‘semillas’. Luego, estas semillas se extienden en ambas direcciones y se selecciona la que tenga mejor puntaje de alineamiento (Li, 2013). Para el proceso de mapeo se ingresó la secuencia corregida de Cambridge como genoma de referencia (NCBI: NC_012920.1) y se le indicó al programa que construyera el index asociado. Se usaron los parámetros entregados por defecto y se estableció la opción *Set read groups (SAM/BAM specification)*. A partir de

este paso, para cada muestra se obtuvieron archivos BAM, un formato binario del alineamiento que es utilizado para los análisis posteriores.

Procesamiento del alineamiento en formato BAM:

Para explorar la calidad del alineamiento, el archivo BAM obtenido en el paso anterior se ingresó en la herramienta Samtools stats de SAMtools v. 1.13 (Li y col. 2009, <https://samtools.sourceforge.net/>). Con el objetivo de generar estadísticas generales, los resultados se resumieron en un reporte único a través de MultiQC. También se evaluó la cobertura de las secuencias a lo largo de los mitogenomas y se visualizó a través del programa BAM Coverage Plotter que genera un histograma en formato PNG.

Al conocer la forma en que fueron preparadas las librerías y que implicó la amplificación por PCR, es necesario eliminar los duplicados. Para ello se empleó la función *MarkDuplicates* de Picard v.2.18.2 (<http://broadinstitute.github.io/picard/>) que permite identificar lecturas duplicadas originadas de un mismo fragmento de ADN y eliminarlas. Se asumió que el archivo input ya estaba ordenado o *sorted*. Luego, se realizó un alineamiento hacia la izquierda de los indels, procesamiento que homogeniza la distribución de las inserciones y deleciones posicionándolas lo más a la izquierda posible. Este paso se realizó con la función *BamLeftAlign* (Garrison y Marth, 2012) y es requerido para el llamado de variantes con FreeBayes. Se aplicó la cantidad máxima de iteraciones correspondiente a 5. Previo al llamado de variantes, se aplicó un filtrado de lecturas a través de la función *filter* de BAMTools v.2.5.1 (Barnett y col. 2011), donde se filtró según la calidad de mapeo seleccionando aquellas que tengan una calidad de al menos 30 ($Q \geq 30$) en la escala phred (probabilidad de error 1 en 1000).

Llamado de variantes para genoma mitocondrial:

Para la etapa de llamado de variantes se utilizó el programa FreeBayes v. 1.3.6, detector bayesiano de variantes genéticas diseñado para polimorfismos pequeños como SNPs e indels (Garrison y Marth, 2012). Se consideró a la mayoría de los parámetros entregados por defecto, pero se modificaron algunos donde se estableció la ploidía en 1 y se configuraron los filtros de entrada para excluir del análisis las lecturas con una calidad de mapeo inferior a 20 y los alelos con una calidad de base inferior a 30 (escala phred), de modo que FreeBayes considere lecturas alineadas de manera confiable y bases de alta calidad. Por defecto, este llamador de variantes considera las variantes respaldadas por al menos 2 observaciones y también por al menos el 20% de las lecturas de una sola muestra. Este análisis generó un archivo VCF único, el que fue sometido a una etapa de filtrado de variantes con la herramienta VCFfilter de vcflib v.1.0.0_rc3 (Garrison, 2015). Se aplicaron los filtros de profundidad de cobertura ≥ 20 y calidad ≥ 30 . El archivo VCF obtenido se ingresó en Haplogrep v2 (Weissensteiner y col., 2016, <https://haplogrep.i-med.ac.at/app/index.html>), donde se exportó en formato FASTA la secuencia de cada muestra.

Confirmación de las variantes identificadas:

Para revisar el resultado del llamado de variantes, se realizó un alineamiento múltiple de las secuencias en el programa MAFFT 7 (Katoh, Rozewicki y Yamada, 2019), alineador online que permite procesar una gran cantidad de información en poco tiempo. Posteriormente, el archivo FASTA generado se abrió en el programa MEGA 11 (Tamura, Stecher y Kumar, 2021) con el fin de editar las secuencias y confirmar las variantes,

realizando una revisión de cada una de ellas. Este paso se llevó a cabo considerando distintas aproximaciones. Por un lado, el mapeo de las lecturas en formato BAM se visualizó en el programa IGV (*Integrative Genomics Viewer*) para ver cada diferencia entre las lecturas y el genoma de referencia junto a la cobertura de estas. También se tuvo en cuenta la información entregada por Haplogrep v2 asociada a la inferencia de haplogrupos y los sitios polimórficos observados. Esto permitió corregir los indels que debido a los requerimientos del llamado de variantes fueron alineados hacia la izquierda, pero que por convención del ADN mitocondrial se ubican hacia la derecha. Por último, es importante considerar la estructura circular del genoma mitocondrial, por lo que los alineadores suelen mostrar una disminución en la cobertura de la región control (16,024–16,569; 1–576) debido al quiebre artificial de esa zona para obtener un genoma lineal (Laricchia y col. 2022). Para enfrentar este problema, a través del programa Geneious se revisaron las secuencias de D-loop obtenidas por secuenciación Sanger que estaban disponibles para todas las muestras analizadas aquí. De esta forma, se corrigieron las variantes en caso de que fuese necesario y se obtuvo la versión final de las secuencias FASTA para cada mitogenoma.

2.3. Métodos de análisis de datos

2.3.1. Determinación de haplogrupos a partir de D-loop completo

Para la determinación de los haplogrupos mitocondriales, las secuencias FASTA se ingresaron a la aplicación *mtDNA nomenclature tool* de mtDNA profiler, donde se obtuvo el perfil de sitios polimórficos de cada muestra y se exportó en formato hsd. Luego,

dicho archivo se ingresó al programa Haplogrep v2 (Weissensteiner y col., 2016, <https://haplogrep.i-med.ac.at/app/index.html>) que reportó la asignación de haplogrupos para cada muestra, en base a PhyloTree 17 (van Oven y Kayser, 2009, <http://www.phylotree.org>). Todas las secuencias fueron alineadas contra la secuencia de referencia de Cambridge rCRS en el programa MEGA 11 (Tamura y col., 2021), aplicando el algoritmo MUSCLE (Edgar, 2004). En el alineamiento múltiple se eliminaron los sitios 16182, 16183, 16193.1C, 16519 y la variación de citosinas en los sitios 309, 315 y de indels en 513-524. Estos sitios corresponden a regiones del ADN mitocondriales de alta variabilidad o inestabilidad, por lo que no se consideraron en los análisis.

Una vez obtenidos los haplogrupos mitocondriales, se calcularon las frecuencias de haplogrupos no amerindios para cada población por medio de conteo directo. En el caso de Azapa solo se contaba con un individuo, así que no fue incluido en la mayoría de los análisis estadísticos comparativo entre poblaciones, salvo en los índices de diversidad de molecular y las redes de haplotipos que se describen más adelante.

2.3.2. Análisis de diversidad genética a partir de D-loop completo

A partir de las secuencias de D-loop completo se calcularon los índices de diversidad genética a nivel país y poblacional, a través del programa DnaSP v.6.12.03 (Rozas y col., 2017, <http://www.ub.edu/dnasp/>), contemplando el número de haplotipos (h); el número de sitios polimórficos (S); la diversidad genética o haplotípica (H_d) (Nei, 1978); el número promedio de diferencias pareadas (K) (Tajima, 1983); y la diversidad nucleotídica (π). También se realizaron las pruebas de neutralidad (Tajima's D y Fu's FS), con la intención de evaluar la desviación del equilibrio Wright-Fisher.

2.3.3. Análisis de estructura poblacional a partir de D-loop completo

Para evaluar la diferenciación genética entre pares de poblaciones, se calculó el índice de FST en el programa Arlequin 3.5.2.2 (Excoffier y Lischer, 2010, <http://cmpg.unibe.ch/software/arlequin35/>). La comparación de poblaciones chilenas se realizó de dos formas, primero considerando los haplotipos estimados por el programa DNAsp a partir de las secuencias D-loop y luego, en base a las frecuencias de los siguientes haplogrupos: L0, L1b, L1c, L2, L3, L3d, L3e, N, D4, W1, W3, X2, HV, H, H1, H1c, H1e, H2a, H3, V, J, T, K, U, U5, U8, Otros. Las matrices de distancias genéticas pareadas se utilizaron para construir un dendrograma Neighbor-Joining en el software MEGA 11, con el fin de representar gráficamente las distancias genéticas entre las poblaciones comparadas. Los valores de FST negativos en la matriz se cambiaron a 0 para ser trabajada en función del dendrograma.

A partir de las frecuencias de haplogrupos se realizó un Análisis de Componentes Principales (PCA) con el paquete *FactoMineR* del software R (R Core Team, 2018). Este análisis se caracteriza por reducir la multidimensionalidad de los datos al generar nuevas variables o componentes a través de combinaciones lineares que capturan la mayor parte de varianza de los datos originales (Kassambara, 2017). De esta forma, es posible integrar una gran cantidad de información en un mismo análisis y, además, permite visualizar las variables que contribuyen a los respectivos componentes. En este sentido, con este análisis se busca identificar qué variables explican las relaciones observadas entre las poblaciones. Adicionalmente, se realizó un PCA sin considerar el haplogrupo L lo que redujo el tamaño muestral para algunas poblaciones, por lo que se agruparon de acuerdo a cercanía

geográfica: Antofagasta-Calama, La Serena-Coquimbo, Aconcagua-San Felipe-Los Andes y Chillán-San Carlos. Se omitieron los individuos de la población de Curicó puesto que el tamaño de la muestra disminuyó a 3 y no se encuentra cercana a ninguna otra población.

A partir de las secuencias de la región control de ADN mitocondrial de las poblaciones chilenas se construyeron redes de haplotipos mediante el programa Network 10.2.0.0 (Bandelt y col., 1999, fluxus-engineering.com). Se realizó una red de haplotipos por origen geográfico de linajes no amerindios: euroasiático de occidente o europeo (haplogrupos HV, JT, K, U, W, X, I), africano (haplogrupo L), y asiático (haplogrupos D, G, M). Este análisis se llevó a cabo con el objetivo de identificar haplotipos compartidos entre distintas poblaciones e inferir la distribución geográfica de los linajes. Para ello, se empleó el algoritmo *Median Joining* (Bandelt y col., 1999), que fue creado para datos genéticos haploides (ADNmt y cromosoma Y) donde hay ausencia de recombinación. Este método combina dos algoritmos (Kruskal y Farris) y calcula la mínima distancia entre dos haplotipos dependiendo del número de mutaciones. Posteriormente se aplicó el método de Máxima Parsimonia como opción de post-procesamiento. Como se mencionó anteriormente, existen sitios *hotspots* y regiones inestables que mutan de manera recurrente a lo largo de los linajes y confunden las relaciones filogenéticas entre los individuos. Por ello, se eliminaron los sitios 16182, 16183, 16519, la variación de citosinas en los sitios 309, 315, 16193 y de indels AC en 513-524. A los sitios homoplásicos 146, 150, 152, 195, 16093, 16129, 16189, 16311 y 16362 (Soares y col. 2009) se les asignó un peso molecular bajo (1).

2.3.4. Análisis multivariados con información de ancestría genómica

Con la intención de comparar los datos de ADN mitocondrial con los datos nucleares disponibles para 296 individuos de la muestra total, se realizó un análisis factorial de datos mixtos (FAMD por sus siglas en inglés). Este análisis permite comparar una base de datos con variables cualitativas, en este caso el origen del linaje mitocondrial (europeo, africano y asiático) y variables cuantitativas que aquí corresponden a las proporciones de ancestría continental obtenidas a partir de un panel de 150 AIMs (YRI = africano; CEU = europeo; NAT_NORTE = amerindio del norte; NAT_CSUR = nativo del centro sur de Chile). A partir del análisis es posible explorar asociaciones entre ambos tipos de variables (Kassambara, 2017), por lo que se buscó observar la información de ancestría según el origen del linaje mitocondrial. Se utilizó el paquete *FactoMineR* para el análisis y *factoextra* para la visualización de los datos en el programa R (R Core Team, 2018).

2.3.5. Comparación con poblaciones europeas y africanas

Para comparar la composición de haplogrupos mitocondriales no amerindios en Chile con países de referencia, se realizaron dos Análisis de Componentes Principales a partir de frecuencias de haplogrupos caracterizados por HVSI y HVSII, uno con población euroasiática y otro con población africana. En el primer caso se consideraron las nacionalidades con mayor frecuencia en Chile en tiempos históricos (Figura 3 y 4) y en el caso de África, se seleccionaron las regiones históricas relacionadas con la trata de esclavos (Figura 1) para caracterizar cada país africano. En la tabla S4 de Anexos se

describen las poblaciones de referencia de África que fueron tomadas desde la literatura. En el caso de las poblaciones euroasiáticas, se utilizó la base de datos de Karachanak y col. (2011) y se agregó una muestra de Holanda (Chaitanya y col. 2016). Para el PCA de haplogrupos africanos L-U6 se agruparon las muestras chilenas de menor tamaño y que fueran cercanas geográficamente entre sí, para no generar un sesgo en el análisis. De esta forma se consideraron los grupos La Serena-Coquimbo y Chillán-San Carlos-Concepción. Curicó se descartó en este caso porque no se pudo agrupar por cercanía geográfica. Adicionalmente, se llevó a cabo un PCA entre poblaciones chilenas y de España para comparar la diversidad genética mitocondrial de linajes euroasiáticos en busca de posibles relaciones entre el país y regiones españolas durante el contacto inicial. Para ello se consideraron 4 regiones en las que se divide España según Barral-Arca y col. (2016): (1) Costa Cantábrica: Galicia, Asturias, Cantabria, País Vasco, Navarra; (2) Plateau Ibérico: Castilla y León, Castilla La Mancha, La Rioja, Extremadura, Madrid, Aragón; (3) Costa Mediterránea: Murcia, Valencia, Catalonia; (4) Andalucía.

Para comparar la estructura genética a nivel de haplotipo y en busca de haplotipos compartidos con poblaciones de referencia, se realizaron redes de haplotipos para los haplogrupos HV y L considerando las medidas metodológicas que se describen en la sección 2.3.3. En Anexos, las tablas S5 y S6 detalla las poblaciones utilizadas para este análisis.

Con el fin de conocer visualmente la distribución de los linajes mitocondriales de interés, se obtuvo una representación espacial de las frecuencias de estos haplogrupos en Chile y en sus continentes de origen a través del programa ArcGIS Pro 2.9 (Esri Inc., 2021, <https://www.esri.com/es-es/arcgis/products/arcgis-pro>). Los mapas se construyeron

utilizando el método de interpolación *Kriging* de la extensión *Spatial Analyst*. Las herramientas de interpolación permiten predecir una superficie continua a partir de valores de punto conocidos, en este caso se cuenta con las frecuencias de haplogrupos para cada población o ciudad que es georreferenciada en el mapa y se utilizan para estimar las frecuencias en los sitios donde no se tiene información. Entre los distintos métodos de interpolación, *Kriging* se caracteriza por ser un procedimiento de estadística geográfica, es decir, que considera la relación estadística entre los puntos medidos, otorgándole una medida de precisión a las predicciones y una mayor robustez al método (Esri Inc., 2021). Para estos análisis se consideraron las frecuencias de los haplogrupos calculadas a partir del N total de la población.

2.3.6. Análisis filogenéticos a partir de genoma mitocondrial completo

El alineamiento múltiple de los mitogenomas en formato FASTA se editó en el programa MEGA 11 con el objetivo de generar una matriz de alineamiento sin los sitios que no son informativos filogenéticamente. Para ello, se eliminaron los sitios 16182, 16183, 16519, la variación de citosinas en los sitios 309, 315, 16193 y de indels AC en 513-524. Al igual que con el marcador D-loop, la determinación de haplogrupos se realizó a través de Haplogrep 2 y en este caso se accedió a la opción de *Graphical Phylogenetic Tree* para visualizar cómo los linajes encontrados se ajustan al árbol de ADN mitocondrial humano descrito por Phylotree que a la fecha incorpora un total de 5.437 haplogrupos (Van Oven y Kayser, 2009; Weissensteiner y col. 2016). De esa forma, a partir de las muestras se obtuvo un árbol enraizado que incluye los polimorfismos encontrados con respecto a la secuencia de referencia rCRS.

Para los análisis filogenéticos se agregó como outgroup la secuencia reconstruida “*Reconstructed Sapiens Reference Sequence*” (RSRS) considerada como el linaje más antiguo del árbol filogenético del ADN mitocondrial (Behar y col. 2012a). Luego de descartar una muestra porque la mayor resolución de los datos permitió caracterizarla como haplogrupo amerindio, la matriz de alineamiento quedó constituida por un total 45 secuencias de un largo de 16567 pb. A partir de estos datos moleculares, se procedió a realizar reconstrucciones filogenéticas de los linajes de ADN mitocondrial empleando tres métodos distintos (Parsimonia, Máxima Verosimilitud e Inferencia Bayesiana). En todos los casos se definió una partición de los caracteres, distinguiendo la región control o d-loop (delimitada entre 1-575, 16025-16567) y la región codificante (576-16024). Se realizó una prueba de Homogeneidad de Partición con el programa PAUP* v.4.0a169 (Swofford, 2003), para evaluar cómo se comportan las dos regiones analizadas y si existen incongruencias entre estas. Se consideró un número de réplicas igual a 100 en una búsqueda heurística. Luego, se procedió con la primera aproximación filogenética que fue de Máxima Parsimonia a través del mismo programa PAUP. Se definió la posición de las particiones, se excluyeron los caracteres no informativos, se definió el outgroup (RSRS) y se realizó una búsqueda heurística usando el método de optimización por defecto de ACCTRAN. Luego, se aplicó un análisis de bootstrap con 1.000 réplicas.

El segundo método de reconstrucción filogenética fue de Máxima Verosimilitud o *Maximum Likelihood* que se realizó usando la interfaz gráfica del programa raxML2.0 (Edler y col., 2020). Se ingresó el alineamiento en formato FASTA y un archivo de texto de las particiones donde se indican las posiciones que ocupan cada una de las regiones. Se

seleccionó la opción de *High Performance Computing* (HPC) como modelo de optimización que es recomendado para obtener un análisis de bootstrap robusto contando con datos reducidos. Entonces, se corrió un análisis de bootstrap rápido con 1.000 réplicas, se definió a RSRS como outgroup y se optó por el modelo de sustitución más complejo que corresponde al GTRGAMMAI. Este modelo contempla a GTR (permite que todas las tasas de sustitución nucleotídica puedan ser diferentes; Tavaré, 1986) y agrega el parámetro GAMMA (Yang, 1994) que indica la tasa de cambio a través de los sitios, e I que corresponde a la proporción de sitios invariantes.

El último método de reconstrucción filogenética aplicado fue de inferencia bayesiana mediante el programa MrBayes 3.2.7 (Ronquist y col. 2012). Para este análisis fue necesario determinar ciertos parámetros con respecto al modelo de evolución que mejor se ajusta a nuestros datos. La búsqueda del mejor modelo para cada una de las particiones consideradas previamente se realizó a través del programa ModelFinder (Kalyaanamoorthy y col. 2017; Chernomor, Von Haeseler y Minh, 2016) implementado en IQ-TREE, software para la inferencia filogenética que posee un servidor web de acceso libre (Trifinopoulos y col. 2016, <http://iqtree.cibiv.univie.ac.at/>). Se utilizó el criterio de información Akaike corregido (AICc) para estimar los modelos con mayor *likelihood* y de acuerdo con lo obtenido, se ingresaron los priors de los parámetros, los parámetros del modelo de *likelihood* y los de la cadena de Markov Monte Carlo. Se indicó un total de 10 millones de generaciones, donde se muestreó cada 1.000 árboles, se hicieron 4 corridas independientes con 4 cadenas cada una y se consideró un 25% de burning.

Todos los árboles filogenéticos obtenidos fueron revisados con el programa FigTree v1.4.4 (Rambaut, 2018, <http://tree.bio.ed.ac.uk/software/figtree/>). Se comparó la

topología de los árboles y al ser similares en los tres análisis se decidió reportar solo uno, indicando los valores de soporte de nodo. Para ello, se editó de manera manual en el editor de imágenes Paint. Los árboles obtenidos a partir de los otros métodos se presentan en Anexos.

3. RESULTADOS

3.1. Resultados del trabajo de laboratorio y bioinformático

3.1.1. Secuenciación de D-loop completo

A partir del procedimiento descrito en Materiales y Métodos, 19 muestras de San Felipe - Los Andes y 19 de Concepción fueron amplificadas por PCR y secuenciadas para la región control de ADN mitocondrial. En la figura 8 se observa el resultado de la amplificación por PCR para 11 muestras de San Felipe - Los Andes a través de un gel de agarosa al 1,5%, donde las bandas obtenidas son de 1504 pb.

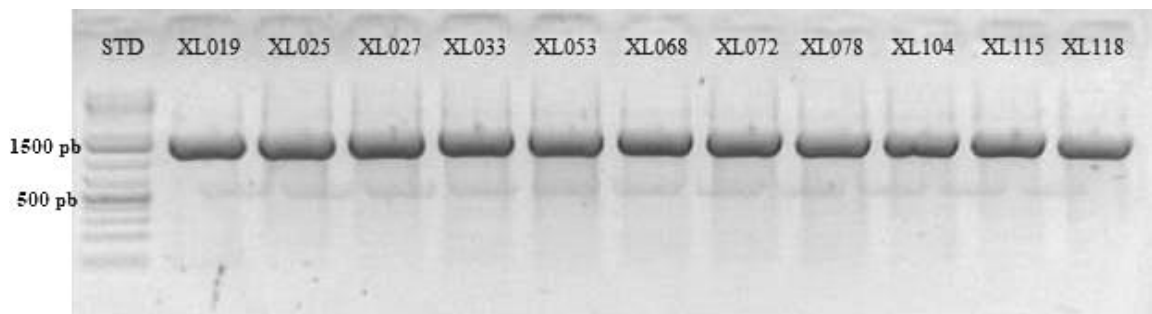


Figura 8: Gel de agarosa al 1,5% luego de electroforesis para 11 productos de PCR de D-loop. STD: estándar de peso molecular (en este caso de 1kb).

Todas las muestras amplificaron correctamente, pero en el proceso de tipificación se descartaron algunas. En el caso de San Felipe - Los Andes, una muestra presentó dos señales en más de un sitio polimórfico que definía para haplogrupo (Figura 9), lo que impidió caracterizar dicha muestra.

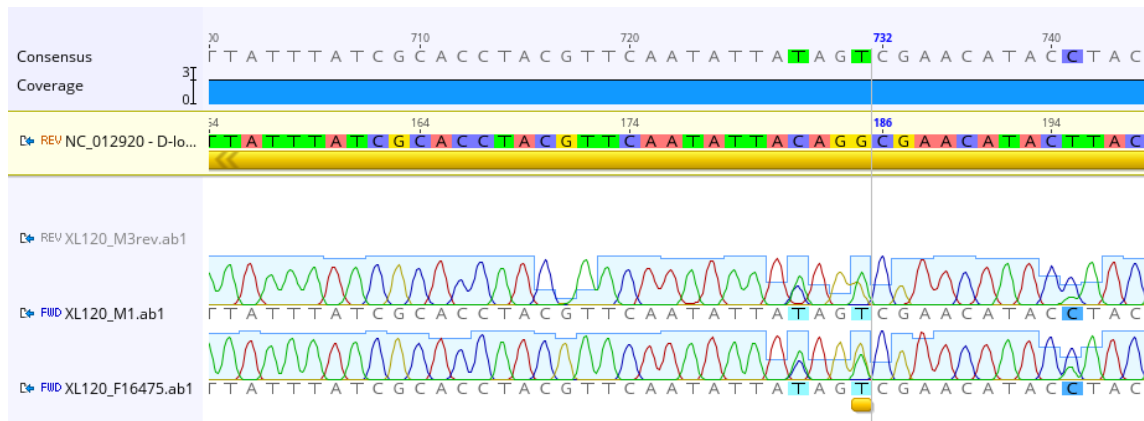


Figura 9: Visor de programa Geneious con los cromatogramas de la muestra XL120 alineados a la secuencia de referencia. Aparecen destacados sitios donde se observa una doble señal de bases nitrogenadas.

3.1.2. Secuenciación NGS de mitogenomas

Posterior a la amplificación por PCR de los siete fragmentos y las estimaciones del volumen a agregar por cada uno, se obtuvieron mezclas que fueran lo más equimolar posible por muestra. Sin embargo, las concentraciones de ADN entre muestras varían debido a la calidad original de éstas, siendo la más baja estimada de 38,03 ng/μl y la más alta de 106,86 ng/μl, con un promedio de 79,02 ng/μl.

En algunos casos las reacciones de PCR generaron bandas inespecíficas, como ocurrió al amplificar el fragmento 6 donde también se amplificaron fragmentos menores a 250 pb y en el caso del fragmento 2 se observaron bandas con tamaños alrededor de los 500 y 750 pb. Con el fin de eliminar estas bandas, se purificaron las muestras y se verificó la eficiencia de este proceso a través de un gel de agarosa al 1,5% (Figura 10) donde se comparó una misma muestra antes y después de la purificación, comprobándose la eliminación de las bandas inespecíficas.

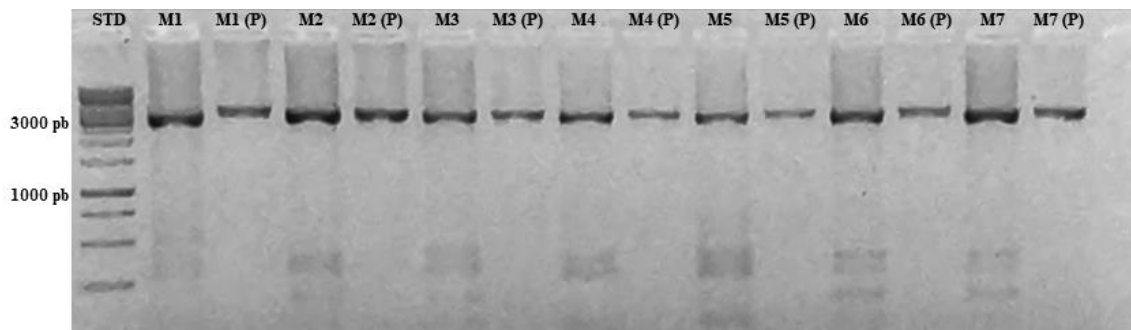


Figura 10: Gel de agarosa al 1,5% del mix de 7 fragmentos antes y después de la purificación de ADN. Las muestras purificadas se indican con una (P). STD: estándar de peso molecular (1kb).

Siguiendo con la secuenciación NGS, la construcción de librerías genómicas implicó la fragmentación del ADN cuyo resultado fue visualizado en un gel de agarosa (ver figura 11). Se obtuvieron digestiones con fragmentos de tamaño superior a las 1000 pb e inferior a 200 pb, los cuales se decidieron eliminar.

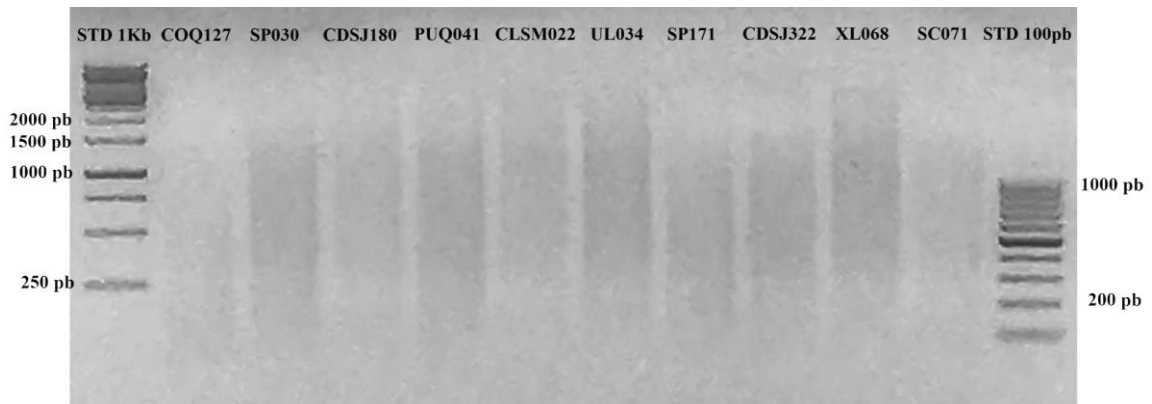


Figura 11: Gel de agarosa al 1,5% posterior a la fragmentación de ADN por la acción de la dsDNA Fragmentase durante 5,5 minutos. STD: estándar de peso molecular (1 kb en extremo izquierdo y 100 pb en extremo derecho).

Luego de la purificación doble, se cuantificaron las muestras lo que permitió constatar la pérdida de ADN durante el proceso, ya que las concentraciones en ng/μl fluctuaron entre 0,108 y 8,77, con un promedio de 3,78 ng/μl. Dada la baja cantidad de

ADN, se cargaron 15 μ l de muestra para que pudieran observarse en un gel de agarosa. En la figura 12 se comprobó que la purificación dio resultado, ya que se obtuvo el tamaño deseado.

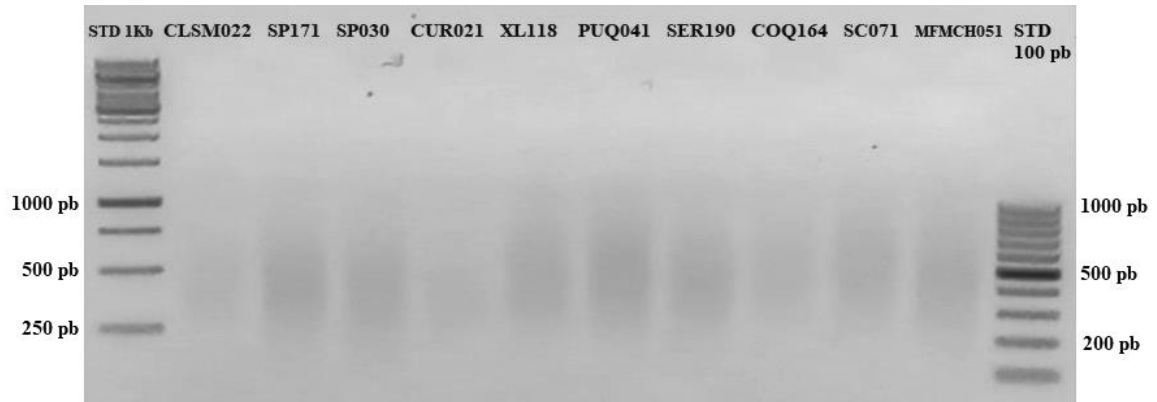


Figura 12: Gel de agarosa al 1,5% posterior a la purificación doble del ADN fragmentado. STD: estándar de peso molecular (1kb en extremo izquierdo y 100 pb en extremo derecho).

Una vez agregado los adaptadores e índices, las librerías construidas fueron cuantificadas obteniendo concentraciones de ADN entre 2,81 ng/ μ l y 18 ng/ μ l con un promedio de 10,94 ng/ μ l. Se realizó un control de calidad a través del equipo TapeStation, evaluando la distribución de tamaños de las librerías. En la figura 13 se presentan 3 ejemplos donde se comprobó una distribución de tamaño de librerías entre 200 pb y 1000 pb con una moda que varía entre 276 pb y 352 pb. La mayoría de las muestras resultaron como el primer ejemplo, con una distribución normal y un peak alrededor de los 300 pb. También se observaron distribuciones sesgadas hacia tamaños más grandes, como el segundo ejemplo, pero no suponen un problema para el proceso de secuenciación. En el tercer caso, la distribución principal presenta pequeños peaks en tamaños largos y resultó ser un caso aislado, por lo que se contempló igualmente para la secuenciación.

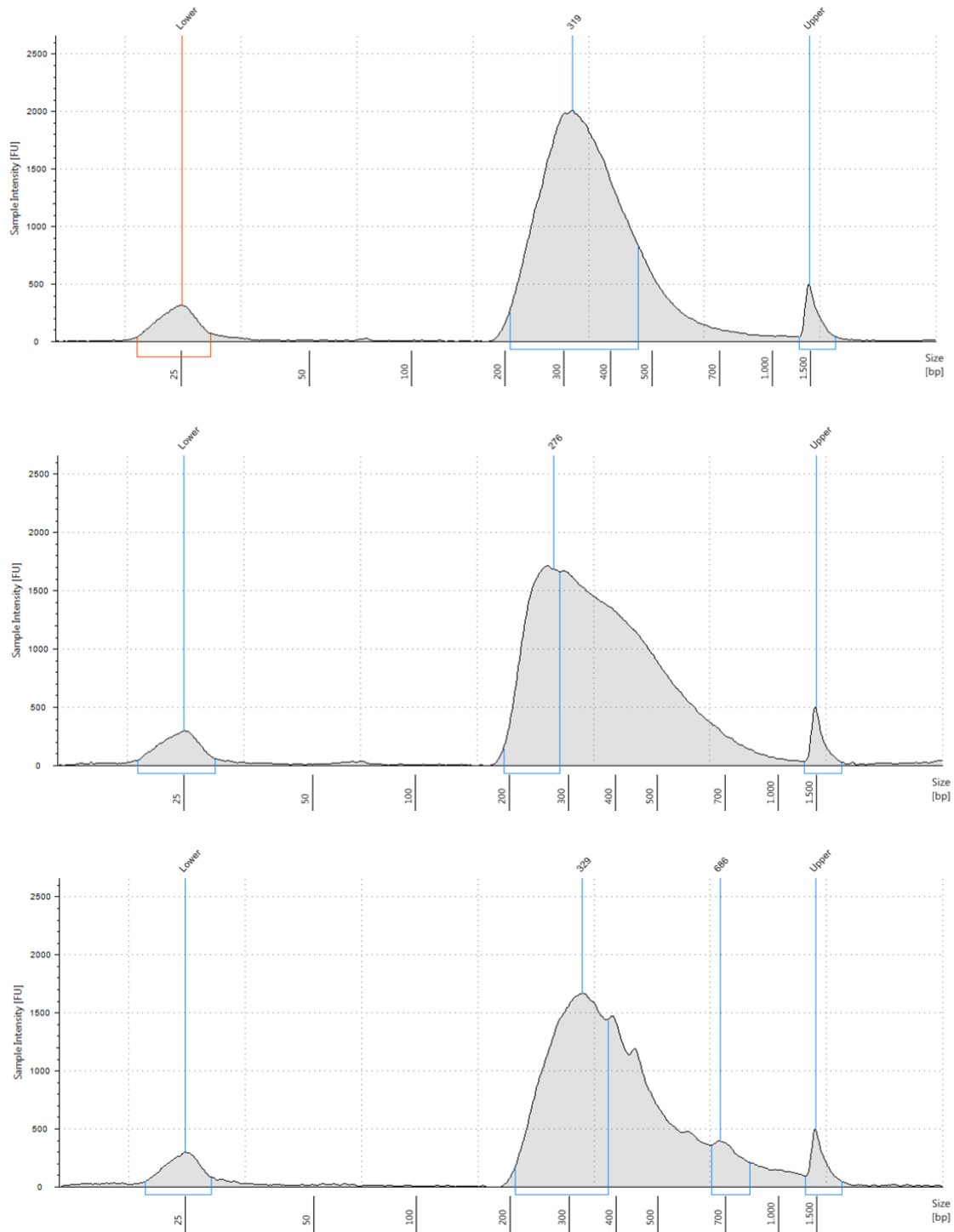


Figura 13: Electroferograma de TapeStation para tres librerías genómicas de ADN mitocondrial con la distribución de tamaños de fragmentos. Se destaca el peak de la distribución, mientras que los peaks en 25 pb y 1500 pb corresponden a estándares internos.

Posterior al proceso de secuenciación fue necesario realizar un control de calidad para conocer la calidad de las lecturas obtenidas. Dentro de los distintos gráficos entregados por el reporte de FastQC y luego resumido por MultiQC, la figura 14 nos muestra la calidad promedio de las lecturas en cada una de las bases que conforman la secuencia, es decir, la certeza con que se está incorporando un nucleótido en esa posición. En el eje Y vemos la calidad representada como el valor de Q (phred) que va de 0 a 40, mientras que en el eje X encontramos las posiciones de la lectura en pb. En general todas las muestras presentaron un perfil de calidad similar (en verde), donde se observa una buena calidad de las bases con un promedio superior a Q30 en gran parte de la secuencia. En color naranja se indican cinco muestras que presentaron una calidad promedio menor con valores entre Q20 y Q30. En todos los casos se observó que las últimas posiciones obtuvieron valores de calidad más bajos con una mayor dispersión de los datos que es visible al revisar los reportes individuales entregados por FastQC. Este descenso en la calidad se ha reportado como un efecto común producto de factores técnicos como la efectividad de los reactivos utilizados a medida que transcurren los ciclos y este fenómeno se acentúa en el caso de las R2.

A partir de este análisis se pudo constatar que los resultados obtenidos por FastQC además de reportar la calidad de secuenciación, también nos entregan información sobre la forma en que fue construida la librería para estos datos. La figura 15 muestra que todas las librerías secuenciadas contienen un alto nivel de fragmentos duplicados, indicando que cerca del 40% de la librería presenta un nivel de duplicación mayor a 10. Dada la técnica de PCR utilizada para amplificar el genoma mitocondrial, es esperable que existan múltiples copias del mismo fragmento.

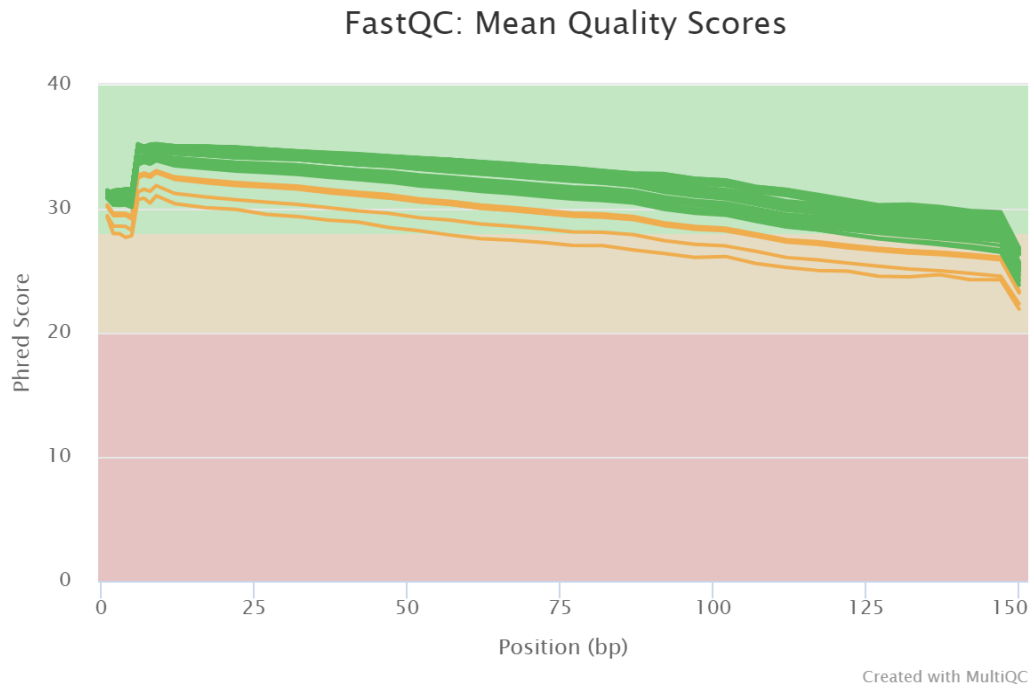


Figura 14: Histograma de la calidad promedio de las secuencias.

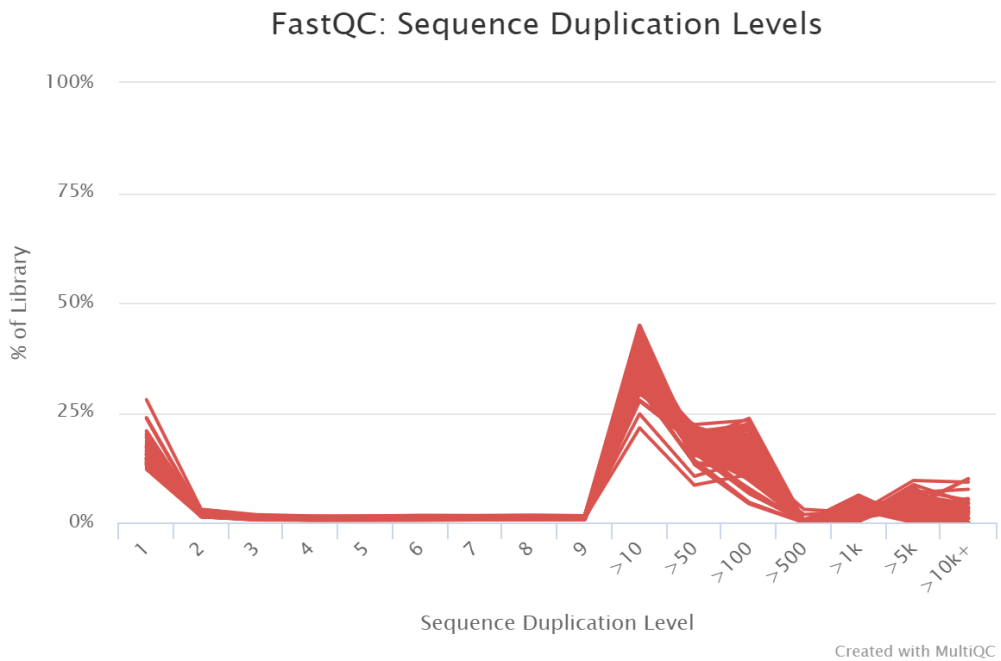


Figura 15: Nivel de duplicación de las secuencias.

Luego del mapeo de las lecturas contra el genoma mitocondrial de referencia se visualizaron las variantes detectadas a través del programa IGV. En la figura 16 se observa una captura del programa para una muestra, donde se distinguen 4 sitios polimórficos en una región del mitogenoma específica (7111-7544 pb).

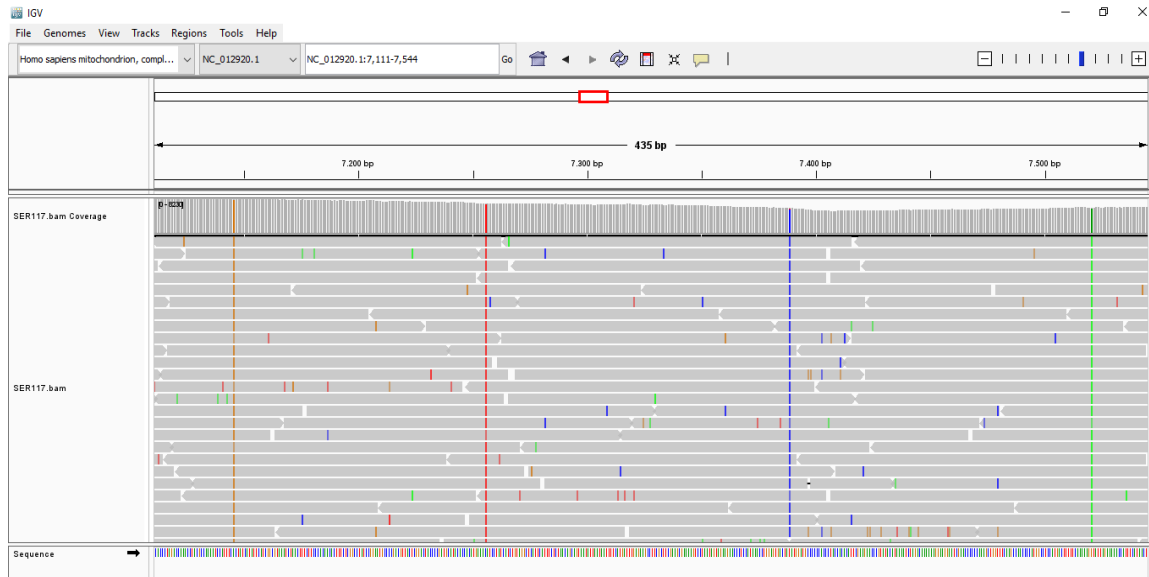


Figura 16: Visor del programa IGV con variantes en región del mitogenoma de la muestra SER117. Se incluye el track del alineamiento (BAM), la secuencia de referencia (NC_012920_1) y se indica la profundidad de cobertura para cada sitio nucleotídico.

Al realizar esta revisión del archivo BAM se pudo constatar un ruido en el alineamiento producto de la presencia de falsas variantes en una o pocas lecturas a la vez. Esto se puede explicar por la tasa de error de la Taq polimerasa utilizada (GoTaq G2 Flexi) al momento de incorporar un nucleótido. A pesar de ello, no interfieren en la identificación de las variantes reales.

En cuanto a las estadísticas generales del alineamiento inicial, la figura 17 indica la cantidad de lecturas de secuenciación obtenidas por muestra, comparando además aquellas que se mapearon o no al genoma de referencia. El número de lecturas varía entre

muestras, pero todas superan el millón de lecturas. Al revisar en detalle la proporción de lecturas mapeadas y no mapeadas, se observó un alto porcentaje de secuencias mapeadas que en la mayoría de los casos es superior al 95%. En cinco muestras el porcentaje fue menor a 95%, pero mayor al 80%.

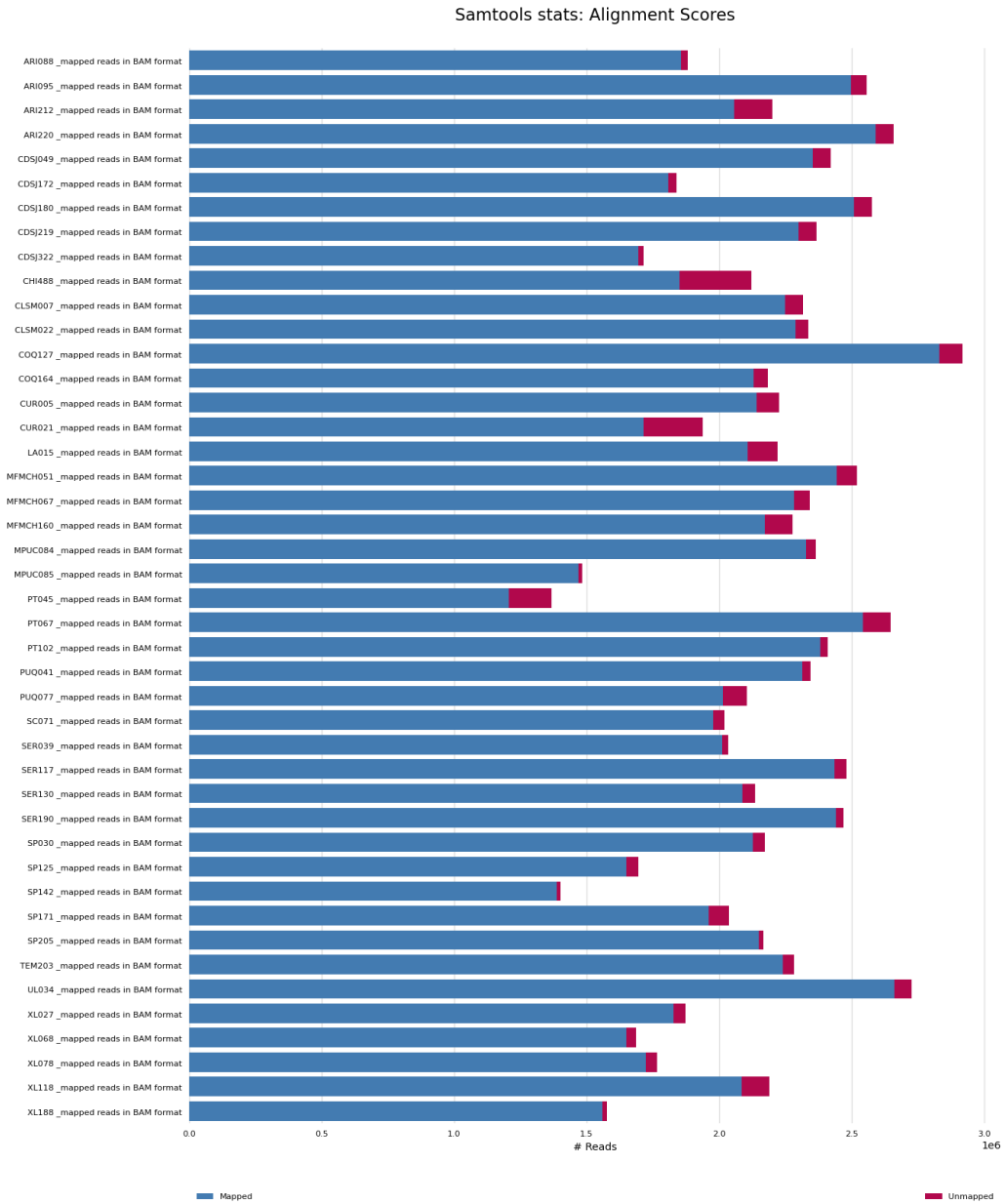
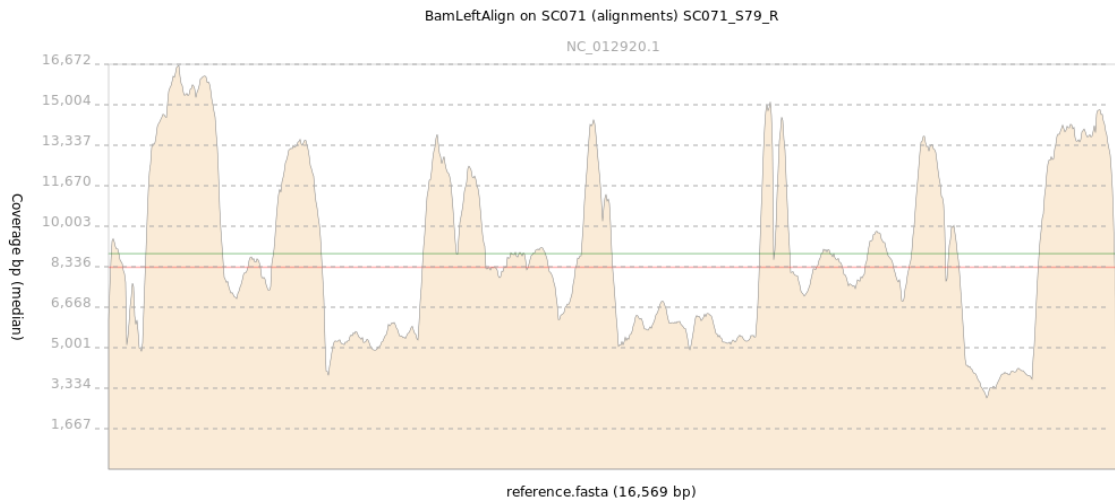


Figura 17: Número de lecturas mapeadas y no mapeadas para cada muestra.

Si evaluamos la cobertura de la secuenciación a lo largo del mitogenoma de referencia, las muestras mostraron una buena cobertura tanto en extensión como profundidad. En el caso de la figura 18 y como ocurre con el resto de las muestras, se observa una cobertura total de la longitud del mitogenoma. En cuanto a la profundidad, la cobertura varía y podemos ver alineamientos como el de la figura 17a donde la profundidad mínima es de aproximadamente 3 kb y otros donde la calidad de la secuenciación fue menor y la cobertura mínima disminuye a menos de 1 kb (figura 18b). Todas las muestras presentaron siete peaks de cobertura a lo largo del mitogenoma donde en casos como la muestra SC071 (figura 18a) una posición está soportada por 16.672 pb. Estos peaks se asocian a las regiones donde se sobreponen los fragmentos utilizados para amplificar el genoma mitocondrial completo.

a)



b)

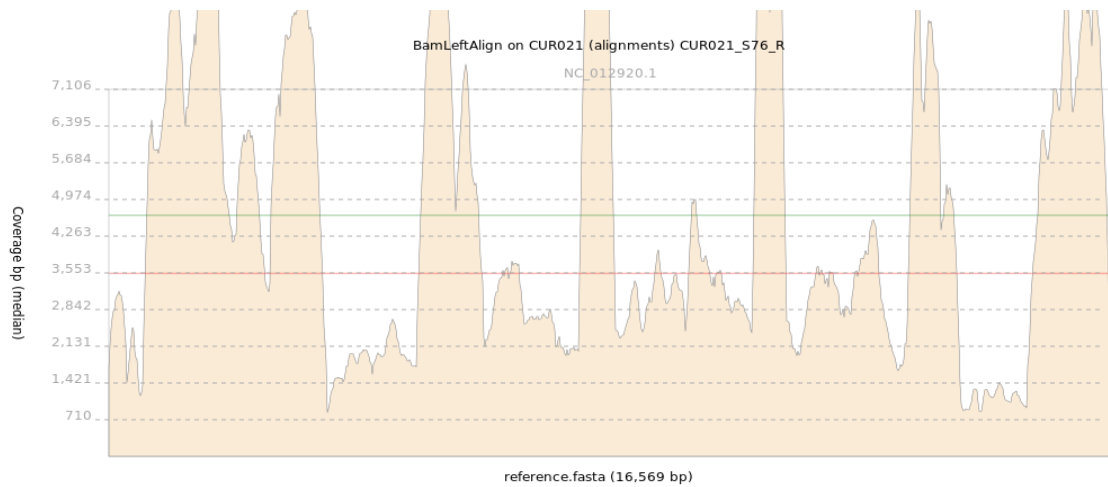


Figura 18: Gráfico de la cobertura en pb de las lecturas de secuenciación en genoma mitocondrial de referencia para dos muestras a) y b).

3.2. Resultados del análisis de datos

3.2.1. Caracterización de la diversidad de linajes mitocondriales no amerindios en Chile

3.2.1.1. Frecuencias de haplogrupos mitocondriales no amerindios

Los haplogrupos mitocondriales no amerindios en las poblaciones chilenas poseen una baja frecuencia y no superan el 15%. De acuerdo con la tabla 1, la proporción más baja de linajes no amerindios se reporta en la población de Azapa con un solo individuo (2,22%) y le sigue otra población del norte de Chile correspondiente a Calama con 3,5%. Por otro lado, las ciudades de Santiago y Punta Arenas presentan las frecuencias más altas para linajes no amerindios con un 14,26% y 14,38% respectivamente.

Para facilitar la comparación entre las poblaciones, los distintos haplogrupos mitocondriales analizados en este estudio se agruparon en los haplogrupos principales encontrados en la muestra. Además, las frecuencias relativas respectivas se estimaron considerando solo el total de linajes no amerindios (Tabla 5), por lo que no representan su proporción real en las poblaciones muestreadas. Con la intención de visualizar de manera gráfica la proporción de haplogrupos euroasiáticos de occidente y su distribución a lo largo de las distintas poblaciones chilenas se generó el gráfico de la figura 19.

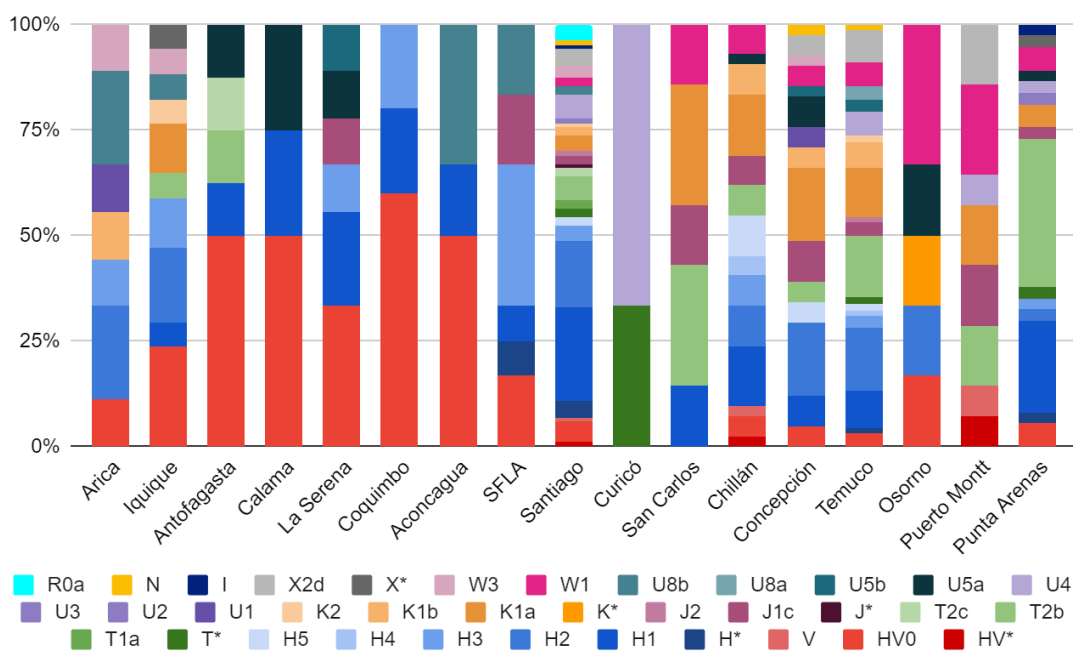


Figura 19: Gráfico de barras apiladas con la proporción de haplogrupos mitocondriales de origen euroasiático en las poblaciones chilenas.

Entre los linajes de origen euroasiático que poseen una distribución en Europa, destaca el macrohaplogrupo HV incluido el haplogrupo H. En la tabla 5 como en la figura 19, el haplogrupo H se presenta de forma separada para mostrar que es el linaje que se encuentra mayormente distribuido a lo largo de Chile, con las frecuencias más altas en el

centro del país con las poblaciones de Santiago y San Felipe-Los Andes (38,89% en ambos casos). Además, corresponde a uno de los más diversos con un total de 41 linajes en toda la muestra. Dentro de este macrohaplogrupo, el H1 es a su vez el más diverso con 20 subhaplogrupos en la totalidad de la muestra. En el caso del haplogrupo H2, el haplotipo H2a2a es el más recurrente (n = 40) encontrado principalmente en el centro y sur de Chile. Otros linajes que destacan son H3p (n = 12), H1c (n = 12) y H5a3a1 (n = 8), al ser más frecuentes en el país. La distribución de los linajes H varía en algunos casos, donde los haplogrupos H4 y H5 se encuentran solo en Santiago, Chillán, Concepción y Temuco. Por otra parte, el haplogrupo HV(xH), representado principalmente por HV0 también se encuentra en gran parte de las poblaciones chilenas, predominando en el norte y centro del país (33,3% en Calama y 42,8% en el valle de Aconcagua).

Otro linaje que está presente en la mayoría de las poblaciones chilenas es el haplogrupo U, con mayor frecuencia en la población de Curicó (40%) y Aconcagua (28%), aunque ambas poblaciones poseen un bajo tamaño muestral. De este haplogrupo, los linajes U1, U2 y U3 son poco frecuentes, representados por un individuo en los últimos dos casos. Los haplogrupos U4, U5 y U8 resultaron ser los más frecuentes, en el primer caso destaca el haplogrupo U4c1 (n = 7) que concentra a individuos del centro y sur de Chile. U5 presenta el haplogrupo derivado U5a1b1 (n = 7) que tiene una distribución variada por el país. En el caso de U8, el haplogrupo más frecuente corresponde a U8b1a2b (n = 8) reportado en poblaciones del norte y centro de Chile.

El haplogrupo K presenta una distribución que se concentra principalmente en el centro y sur de Chile, con las frecuencias más altas en la población de San Carlos (22,2%) y Concepción (20,9%). El subhaplogrupo K1a4a1a+195 (n = 30) es el más frecuente en

toda la muestra, especialmente en las poblaciones de Chillán, Concepción y Temuco. Le sigue el linaje K1b1a2 (n = 11) y otros menos frecuentes.

El macrohaplogrupo JT también está presente mayoritariamente en poblaciones del centro y sur del país. En el caso del haplogrupo J, la frecuencia más alta es reportada en la población de Puerto Montt (14,2%). Este linaje está representado principalmente por J1c2e2 (n = 12) y el resto de los linajes corresponden a haplotipos únicos. Para el haplogrupo T, su mayor frecuencia se encuentra en Punta Arenas con un 34,1%, siendo además el linaje europeo más frecuente para dicha población. En este haplogrupo, destacan T2b3+151 (n = 24) y T2b7a3 (n = 14) como los linajes más recurrentes.

Otro linaje mitocondrial encontrado en gran parte del centro y sur del país es el haplogrupo W que muestra su frecuencia más alta en la población de Osorno (28,5%). Este linaje se encuentra representado por solo dos haplogrupos, W1+119 (n =19) y W3a1+199 (n =6). Por otra parte, el linaje X no amerindio está presente con una distribución más reducida y en menores frecuencias, alcanzando la frecuencia más alta en la población de Puerto Montt con un 14,2%.

Se reportaron otros linajes que poseen una distribución en Europa que fueron poco frecuentes, como el haplogrupo I, encontrando el linaje I1 en Santiago y I1a1 en Punta Arenas. También se encontró el haplogrupo R0a+60.1T en 4 individuos de Santiago y los linajes N1a y N1b en 3 individuos de Santiago, Concepción y Temuco.

Tabla 5: Frecuencias relativas de los macrohaplogrupos mitocondriales no amerindios por población chilena.

Población	N	Hg D	Hg HV	Hg H	Hg J	Hg K	Hg L	Hg T	Hg U	Hg W	Hg X	Otros*
Arica	25	0	0,0400	0,1200	0	0,0400	0,6400	0	0,1200	0,0400	0	0
Iquique	28	0	0,1429	0,2143	0	0,1071	0,2857	0,0357	0,0714	0,0357	0,0357	0,0714
Antofagasta	18	0	0,2222	0,0556	0	0	0,5556	0,1111	0,0556	0	0	0
Calama	6	0	0,3333	0,1667	0	0	0,3333	0	0,1667	0	0	0
La Serena	12	0	0,2500	0,2500	0,0833	0	0,2500	0	0,1667	0	0	0
Coquimbo	11	0	0,2727	0,1818	0	0	0,5455	0	0	0	0	0
Aconcagua	7	0	0,4286	0,1429	0	0	0,1429	0	0,2857	0	0	0
San Felipe-Los Andes	18	0	0,1111	0,3889	0,1111	0,0556	0,2222	0	0,1111	0	0	0
Santiago	126	0	0,0556	0,3889	0,0317	0,0556	0,1587	0,0952	0,0794	0,0397	0,0317	0,0635
Curicó	5	0	0	0	0	0	0,4000	0,2000	0,4000	0	0	0
San Carlos	9	0	0	0,1111	0,1111	0,2222	0,2222	0,2222	0	0,1111	0	0
Chillán	60	0,0667	0,0500	0,3167	0,0500	0,1500	0,2333	0,0500	0,0333	0,0500	0	0
Concepción	43	0,0233	0,0465	0,2791	0,0930	0,2093	0,0233	0,0465	0,1395	0,0698	0,0465	0,0233
Temuco	82	0,0366	0,0244	0,2683	0,0366	0,1585	0,1220	0,1341	0,0976	0,0488	0,0610	0,0122
Osorno	7	0,1429	0,1429	0,1429	0	0,1429	0	0	0,1429	0,2857	0	0
Puerto Montt	14	0	0,0714	0	0,1429	0,1429	0	0,1429	0,0714	0,2143	0,1429	0,0714
Punta Arenas	41	0	0,0488	0,2683	0,0244	0,0488	0,0488	0,3415	0,0976	0,0488	0,0244	0,0488

* En 'Otros' se incluyeron los haplogrupos menos frecuentes encontrados: C, G, I, M, N, R0a.

En cuanto al principal linaje mitocondrial africano, el haplogrupo L alcanza frecuencias que van desde un 2,3% en Concepción hasta un 64% en Arica. Junto a esta última, las poblaciones del norte del país concentran las frecuencias más altas, con un 55,6% en Antofagasta y 54,5% en Coquimbo. En general, el resto de las poblaciones presentan cifras superiores al 10%, destacando Chillán en el sur de Chile con un 23,3%. El haplogrupo L corresponde al más diverso con un total de 45 haplogrupos distintos presentes en la muestra. La figura 20 presenta la proporción de linajes L en las distintas poblaciones a una mayor resolución. En ella se observa que las poblaciones del norte grande del país presentan una mayor diversidad de estos linajes, la mayoría con al menos un sublinaje de los haplogrupos mayores subsaharianos L0, L1, L2 y L3.

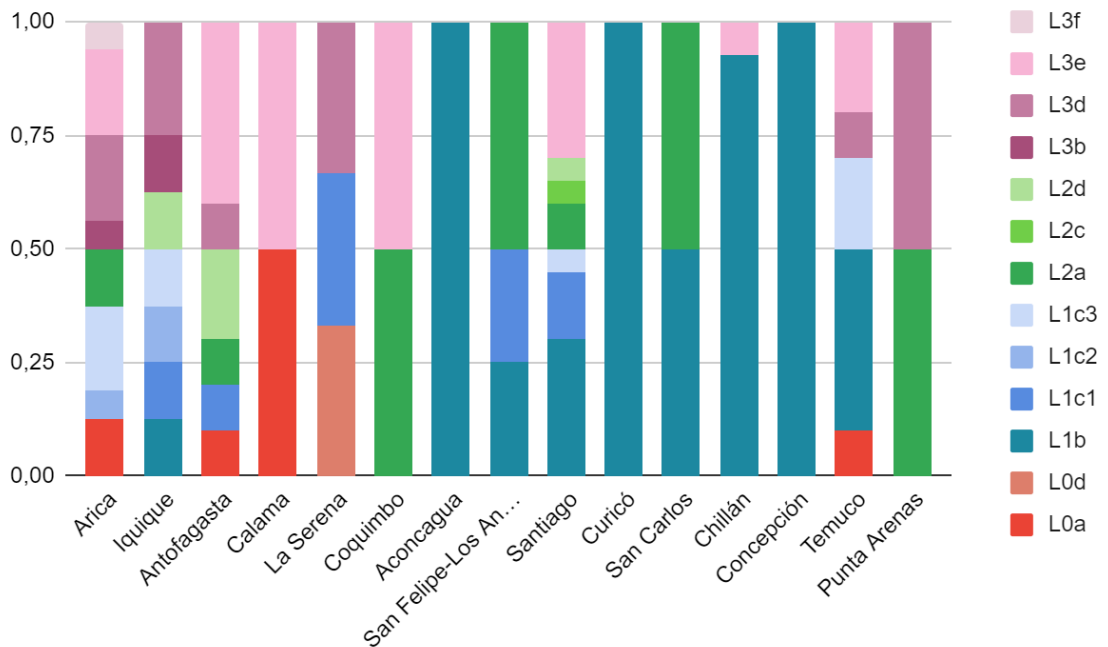


Figura 20: Gráfico de barras apiladas con la proporción de haplogrupos mitocondriales de origen africano en las poblaciones chilenas.

En el caso de L0, con sus linajes L0a y L0d, de los 6 individuos se encontraron 5 haplotipos distribuidos principalmente en el norte de Chile. Para el haplogrupo L1, el linaje L1b (n = 30) está representado en su mayoría por L1b1a+189 (n = 27) que se concentra en la población de Chillán y ciudades aledañas, mientras que para L1c (n = 16) se observa una mayor diversidad con 9 linajes. Por otro lado, el haplogrupo L2 se encuentra caracterizado por linajes de L2a (n = 9), L2c (n = 4) y L2d (n = 4) los que presentan una distribución mayoritaria en el norte y centro de Chile. En este haplogrupo, destacan L2a1+143 y L2d+16129 al ser linajes compartidos por 4 individuos en cada caso, el primero con una distribución amplia, mientras que el segundo está presente en individuos de Iquique, Antofagasta y Santiago. Por último, el haplogrupo L3 es el más diverso de la muestra, agrupando individuos portadores de linajes L3b (n = 2), L3d (n = 9), L3e (n = 20) y L3f (n = 1). De estos, el haplogrupo L3e es el más frecuente y diverso con 10 linajes distintos distribuidos a lo largo de Chile y las frecuencias más altas en el norte y centro del país. L3f está representado por un linaje derivado único L3f1b1a1 que se reportó en la población de Arica. En cuanto al haplogrupo U6 considerado como linaje africano, se encontró en baja frecuencia, con 4 haplotipos distribuidos en Iquique, Chillán, Santiago y Punta Arenas.

Con respecto a los haplogrupos de origen asiático, el haplogrupo D se restringe a 4 poblaciones del sur del país, Osorno (14,2%), Chillán (6,6%), Temuco (3,6%) y Concepción (2,3%). Inicialmente, los resultados obtenidos a partir de D-loop sobreestimaron la frecuencia del haplogrupo D no amerindio por la presencia del linaje D4p. Sin embargo, a partir del análisis de mitogenomas, una muestra tipificada originalmente como D4p (PUQ081) fue caracterizada por Haplogrep como D1g con una

calidad del 97,7%, siendo éste un haplogrupo amerindio del sur de Chile y Argentina (Bodner y col., 2012). Este linaje es definido por el patrón mutacional A8116G–C16187T, sin embargo, en esta muestra no se encontró el polimorfismo 16187, pero sí 2092T, 8116G y 16325C que distingue a D1 de D4. De esta forma, se descartaron de los análisis a aquellos individuos caracterizados por D-loop como D4p (2 de Puerto Montt y 5 de Punta Arenas). En cuanto a otros linajes asiáticos, se identificaron linajes del haplogrupo M presentes en 3 individuos de Iquique y Santiago. El haplogrupo G1a1a3 se identificó en 2 individuos de Iquique y Punta Arenas y finalmente, se reportó el linaje C en un individuo de Puerto Montt.

3.2.1.2. Índices de diversidad molecular

En la tabla 6 se describen los índices de diversidad genética por país y población. Es importante mencionar que los resultados obtenidos aquí corresponden a una submuestra que solo incluye a linajes no amerindios, por lo que las interpretaciones se deben realizar considerando aquello. Al estimar la diversidad genética o haplotípica (H_d), es decir, la probabilidad de tomar dos individuos al azar y que presenten haplotipos distintos, se obtuvieron valores altos para todas las poblaciones (en su mayoría $H_d > 0,9$), con un H_d para Chile de $0,982 \pm 0,0019$. En el rango de 0 a 1, San Carlos presenta la cifra más alta ($H_d = 1$), siendo una población de bajo tamaño muestral donde los 9 individuos tienen haplotipos distintos. Le siguen con los valores más altos, las poblaciones de Arica ($H_d = 0,987 \pm 0,015$) y Santiago ($H_d = 0,986 \pm 0,005$), mientras que Curicó tiene el índice de diversidad más bajo ($H_d = 0,800 \pm 0,164$). La diversidad nucleotídica (π) estimada como la probabilidad de que dos sitios homólogos seleccionados aleatoriamente sean

diferentes, muestra valores que van de 0,00725 en Aconcagua, un máximo de 0,01315 en Arica y un promedio para la población chilena de 0,01022.

El número promedio de diferencias a nivel de nucleótidos entre par de secuencias (K) fue de 11,346 en el total de la muestra, con el promedio más bajo en Aconcagua (K = 8,095) y alcanza el valor más alto en la población de Arica (K = 14,673). Este índice incorpora la dimensión temporal, puesto que las mutaciones se acumulan en los linajes de manera proporcional al tiempo. Estos valores concuerdan con la diversidad de linajes no amerindios presentes en la muestra, siendo Arica la población que concentra una alta frecuencia de linajes africanos que poseen una profundidad temporal mayor.

En cuanto a los tests de neutralidad que ponen a prueba la desviación del equilibrio de Wright-Fisher, la mayoría de las poblaciones obtuvo un valor de Tajima's D negativo, a excepción de Coquimbo y Curicó (ver tabla 6). Considerando la totalidad de Chile, el valor es negativo y significativo (-1,98001; $p < 0,05$), indicando que la población se encuentra fuera del equilibrio W-F o mutación-deriva, lo mismo que ocurre con Santiago y Temuco. El test Fu's Fs también mostró valores negativos para la mayoría de las poblaciones, excepto aquellas que tienen un pequeño tamaño muestral. Con este análisis, además, se obtuvieron más valores significativos, adicionando a Arica, Iquique, Antofagasta, Chillán, Concepción y Punta Arenas. Si se consideran ambos índices, de igual forma se rechaza el equilibrio de W-F para estas poblaciones, ya que el test de Fu's Fs es más potente para detectar desviación del equilibrio. Los valores de D y Fu menores a 0 en ambos casos estaría indicando un patrón que se caracteriza por haplogrupos raros.

Tabla 6: Índices de diversidad molecular y pruebas de neutralidad a nivel país y poblacional.

Población	n	h	S	Hd	π	K	Tajima's D	Fu's Fs
CHILE	520	212	215	0,982 ± 0,0019	0,01022	11,346	-1,98001 (p < 0,05)*	-32,342 (p = 0,000)*
Arica	25	21	79	0,987 ± 0,015	0,01315	14,673	-1,20393 (p > 0,10)	-5,164 (p = 0,004)*
Iquique	28	23	78	0,981 ± 0,016	0,01105	12,317	-1,53927 (p > 0,10)	-7,183 (p = 0,001)*
Antofagasta	18	16	66	0,980 ± 0,028	0,01228	13,699	-1,23417 (p > 0,10)	-3,906 (p = 0,016)*
Calama	6	5	31	0,933 ± 0,122	0,01033	11,533	-0,95426 (p > 0,10)	1,054 (p = 0,419)
La Serena	12	10	53	0,970 ± 0,044	0,01060	11,818	-1,50157 (p > 0,10)	-1,015 (p = 0,180)
Coquimbo	11	6	25	0,873 ± 0,071	0,00896	10,000	0,78792 (p > 0,10)	2,657 (p = 0,135)
Aconcagua	7	5	20	0,905 ± 0,103	0,00725	8,095	-0,04679 (p > 0,10)	1,238 (p = 0,330)
San Felipe-Los Andes	18	13	50	0,961 ± 0,030	0,00970	10,830	-1,16992 (p > 0,10)	-1,341 (p = 0,119)
Santiago	126	90	150	0,986 ± 0,005	0,00936	10,434	-2,08426 (p < 0,05)*	-88,057 (p = 0,000)*
Curicó	5	3	22	0,800 ± 0,164	0,01111	12,400	1,29205 (p > 0,10)	4,313 (p = 0,110)
San Carlos	9	9	45	1,000 ± 0,052	0,01232	13,750	-0,95434 (p > 0,10)	-2,084 (p = 0,111)
Chillán	60	27	74	0,940 ± 0,018	0,01124	12,544	-0,75720 (p > 0,10)	-2,026 (p = 0,047)*
Concepción	43	24	80	0,952 ± 0,018	0,00931	10,394	-1,60155 (0,10 > p > 0,05)	-3,890 (p = 0,011)*
Temuco	82	47	116	0,969 ± 0,008	0,00973	10,852	-1,84424 (p < 0,05)*	-19,305 (p = 0,000)*
Osorno	7	6	27	0,952 ± 0,096	0,00913	10,190	-0,42812 (p > 0,10)	0,184 (p = 0,367)
Puerto Montt	14	10	42	0,945 ± 0,045	0,01007	11,231	-1,48948 (p > 0,10)	-0,085 (p = 0,227)
Punta Arenas	41	30	72	0,973 ± 0,015	0,00848	9,457	-1,63727 (0,10 > p > 0,05)	-12,854 (p = 0,000)*

n: n° de secuencias; **h:** n° de haplotipos; **S:** n° de sitios polimórficos; **Hd:** diversidad genética o haplotípica; **K:** n° promedio de diferencias por pares; **π :** diversidad nucleotídica. * Valor significativo estadísticamente (p < 0,05).

3.2.1.3. Estructura poblacional: Análisis de FST y Dendrogramas

Con el objetivo de evaluar la estructura poblacional, se calculó el índice de FST a partir de dos tipos de datos. Primero, en base a las frecuencias de 211 haplotipos estimados por DNAsp y en el segundo caso, en base a las frecuencias de 27 haplogrupos. Las matrices de distancia obtenidas (ver tablas S2 y S3 en Anexos) permitieron la construcción de dendrogramas Neighbor-Joining para visualizar las distancias genéticas entre las poblaciones analizadas (ver Figura 21). En la figura 21a se observa que la mayoría de las poblaciones se agrupan en una rama del árbol, mientras que el grupo conformado por Curicó y Chillán ($F_{ST} = 0,0101$, $p = 0,342$) junto a San Carlos (F_{ST} con respecto a Chillán = $-0,01897$, $p = 0,027$) se separa genéticamente del resto de poblaciones chilenas, posicionándose en una rama aparte. El conglomerado de poblaciones se divide a su vez en dos ramas, una que concentra a poblaciones del norte y centro del país, mientras que en la otra se agrupan las poblaciones del centro-sur y sur de Chile. En el primer caso destaca la agrupación que contiene a las poblaciones nortinas correspondientes a Arica, Calama, Antofagasta y Coquimbo, estas últimas formando un clúster ($F_{ST} = 0,0178$, $p = 0,288$). Con una baja diferenciación, le sigue la población de Iquique, también del norte de Chile, luego La Serena, y las poblaciones de la zona centro, Aconcagua, San Felipe – Los Andes y Santiago. La población de Santiago se diferencia significativamente de las poblaciones que se encuentran en la otra rama del centro-sur y sur de Chile, como es Concepción ($F_{ST} = 0,0145$, $p = 0,018$), Puerto Montt ($F_{ST} = 0,0442$, $p = 0,009$) y Punta Arenas ($F_{ST} = 0,030$, $p = 0,000$), pero no así de Temuco ($F_{ST} = 0,0038$, $p = 0,108$) y Osorno ($F_{ST} = 0,0215$, $p = 0,1351$).

En la figura 21b las poblaciones de Chile se diferencian en dos grupos principales, uno conformado por las ciudades del norte y centro del país, mientras que en la otra rama se encuentran las poblaciones del centro-sur y sur de Chile. En el primer caso, se observa una agrupación dada entre la mayoría de las poblaciones del norte de Chile, a excepción de Iquique que se diferencia de éstas con una moderada distancia y no significativa. En este dendrograma también se observa la relación entre Antofagasta y Coquimbo ($F_{ST} = -0,0608$, $p = 0,9369$) y se generó un grupo entre La Serena y Aconcagua ($F_{ST} = -0.0016$, $p = 0,3783$).

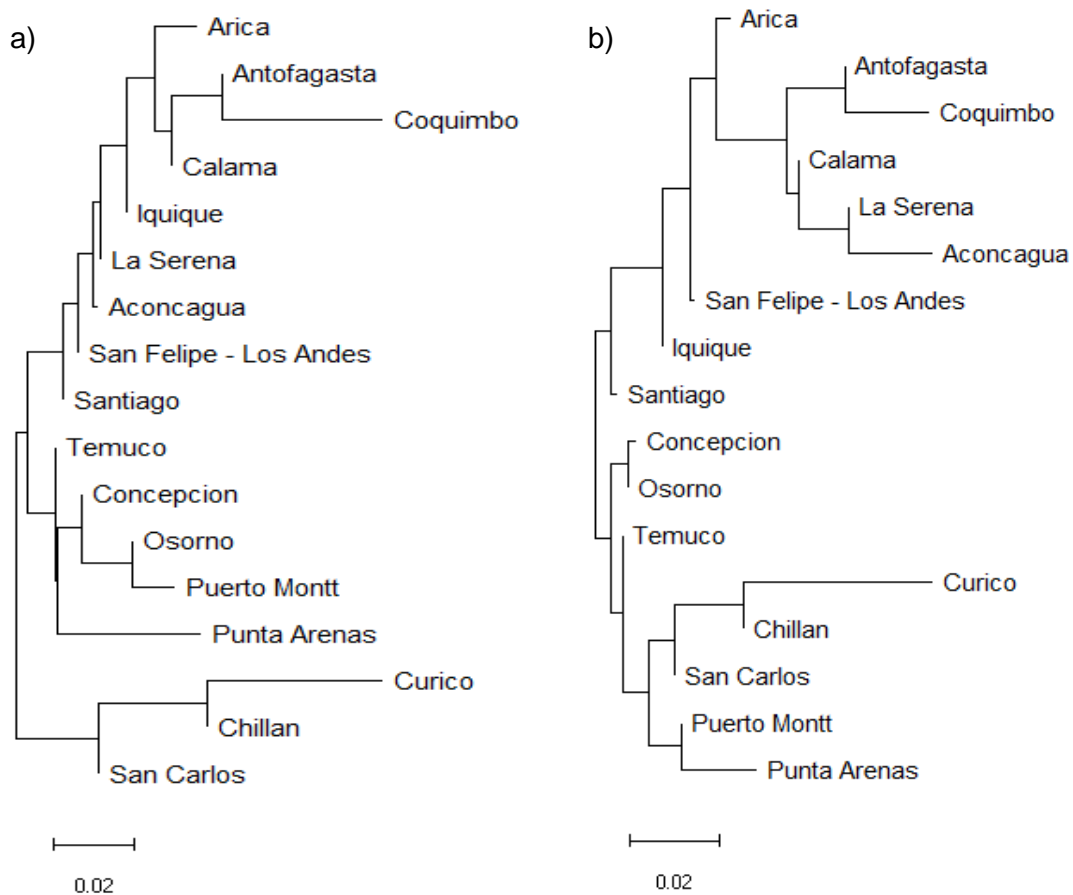


Figura 21: Dendrogramas Neighbor-Joining construidos a partir de matrices de distancias de F_{ST} pareadas considerando las frecuencias de 211 haplotipos estimados a partir de las secuencias d-loop (a) y las frecuencias para 27 haplogrupos (b).

En la rama inferior vemos un grupo formado por las poblaciones de Concepción y Osorno (-0.0240 , $p = 0,6396$), también se repite la agrupación entre Curicó y Chillán, junto a San Carlos, pero esta vez como cluster hermano al grupo de Puerto Montt y Punta Arenas. Tanto la relación Antofagasta-Coquimbo como Curicó-Chillán-San Carlos destacan por mantenerse en ambas aproximaciones, donde se calculó el índice de FST entre pares de poblaciones a partir de la diversidad haplotípica y de la diversidad basada en 27 haplogrupos escogidos como representativos.

3.2.1.4. Estructura poblacional: Análisis de Componentes Principales

El análisis de componentes principales realizado a partir de la frecuencia de 27 haplogrupos muestra que las poblaciones también se diferencian geográficamente. En la figura 22, el componente 1 que explica el 22,2% de la varianza separa hacia el lado negativo del eje a aquellas poblaciones del norte y centro norte de Chile, mientras que en el lado positivo del componente 1 se ubican las poblaciones del centro y sur del país, incluyendo a Santiago. En la figura 22b se observa que todos los linajes africanos L tiene mayor peso en los cuadrantes donde se encuentran las poblaciones del norte, a excepción del haplogrupo L1b, siendo el único asociado a poblaciones del centro sur de Chile hacia el lado positivo del componente 1. Otros haplogrupos como HV, U5, U8, H3 también contribuyen a que las poblaciones del norte y centro norte se diferencien del resto, donde el haplogrupo HV es la segunda variable que contribuye en mayor porcentaje al PC1 (ver Figura S1 en Anexos). La variable con la contribución más alta en dicho componente corresponde al haplogrupo K, el que junto a otros haplogrupos europeos y euroasiáticos

como D4, X2, W1, J, T poseen una mayor frecuencia en poblaciones del centro y sur del país, diferenciándolas a un lado del componente 1.

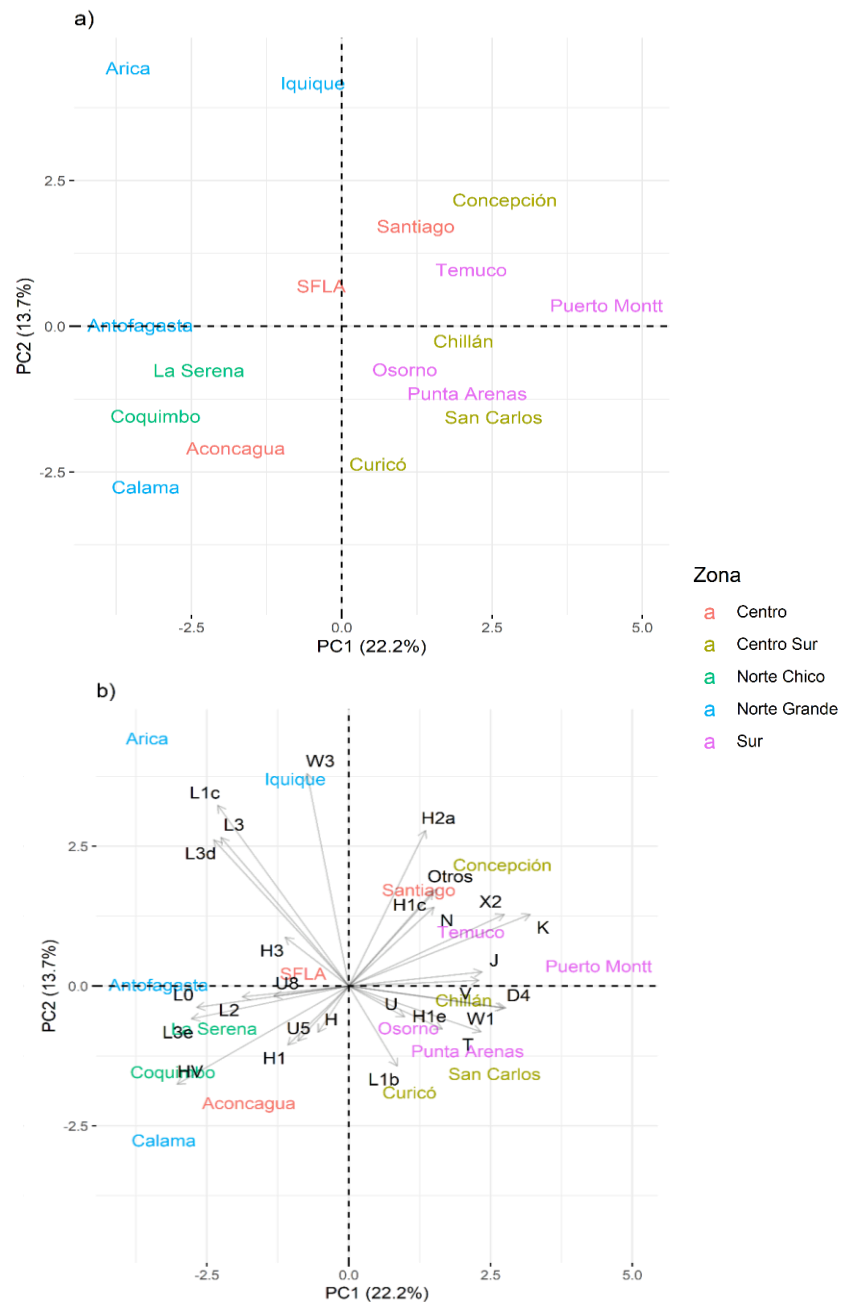


Figura 22: Análisis de Componentes Principales de las poblaciones chilenas (a). En (b) se representan las variables sobre los componentes. Este análisis se realizó en base a frecuencias de hablogrupos.

En cuanto al componente 2, éste captura la variación que permite separar a las poblaciones del extremo norte del país, Arica e Iquique, de todas las demás. En estas poblaciones predomina una mayor frecuencia de los haplogrupos W3, L1c, L3(xL3d, L3e), L3d, los que a su vez mostraron una contribución más alta para el componente 2 (ver Figura S2 en Anexos). En las otras poblaciones del norte grande, Antofagasta y Calama, y del norte chico, La Serena y Coquimbo, tienen mayor peso los haplogrupos africanos L0, L2 y L3e (figura 22b). Por otro lado, el haplogrupo H2a es el segundo en porcentaje que contribuye al componente 2, y junto a los linajes H1c, X2, K, J separa a Santiago, Concepción, Temuco y Puerto Montt del resto de poblaciones del centro y sur del país donde priman los haplogrupos H1e, T, W1, U(xU5, U8), L1b, D4.

Si se elimina del análisis el componente africano representado por el haplogrupo L, las poblaciones chilenas se estructuran de manera similar a la observada en la figura 22. La figura 23 nos muestra los componentes 1 y 2 que en conjunto explican el 43,8% de la varianza. En el lado negativo del componente 1 se encuentran las poblaciones del norte (Antofagasta-Calama, La Serena-Coquimbo) y del centro del país (Aconcagua-San Felipe-Los Andes). En el caso de Arica e Iquique, éstas se diferencian ligeramente del centro del eje hacia este grupo. De manera similar, pero hacia el otro lado del eje, se ubican las poblaciones de Santiago, Chillán-San Carlos y Punta Arenas. El resto de las poblaciones del centro-sur y sur de Chile, correspondientes a Concepción, Temuco, Osorno y Puerto Montt se disponen en el lado positivo del componente 1. La figura 23b muestra que las variables con mayor peso a cada lado del componente 1 se mantienen al eliminar los haplogrupos africanos L. Las tres variables que contribuyen en mayor porcentaje a la diferenciación observada en el este componente son el haplogrupo HV para

las poblaciones del norte y los haplogrupos K y X2 para las poblaciones del centro y sur del país.

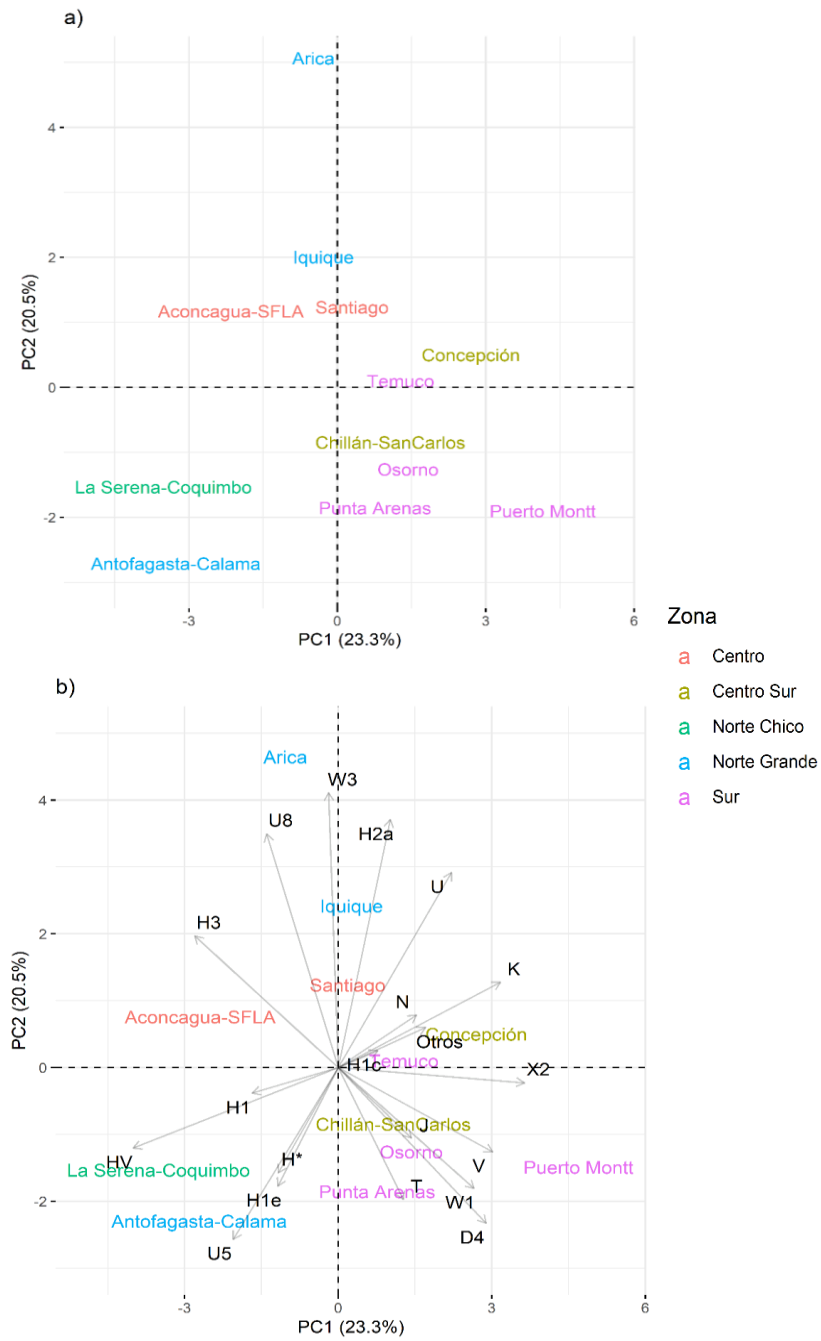


Figura 23: Análisis de Componentes Principales de las poblaciones chilenas sin considerar el haplogrupo L (a). En (b) se representan las variables sobre los componentes. Este análisis se realizó en base a frecuencias de haplogrupos.

3.2.1.5. Redes de haplotipos

Para los haplogrupos de origen euroasiático que poseen una distribución en Europa, se creó una *network* que se observa en la figura 24, constituida por 388 individuos. El macrohaplogrupo HV destaca como el más diverso y que concentra a gran parte de los individuos. El haplotipo principal corresponde al haplogrupo H2a2a, cuyos individuos portan la mutación 263G. Este haplotipo es el más frecuente y es compartido principalmente por individuos del centro y sur del país, pero también se encuentra en ciudades del norte como Arica e Iquique y del extremo sur en Punta Arenas. El haplogrupo H3p también es compartido por individuos provenientes de distintas ciudades a lo largo del país. Para el haplogrupo H1e se observan haplotipos únicos principalmente y algunos compartidos por 2 o 3 individuos. Estos linajes están presentes mayoritariamente en poblaciones de Antofagasta y Calama, Santiago y Punta Arenas.

Por otro lado, el haplogrupo HV0 concentra a la mayoría de los individuos del norte de Chile con representación de todas las poblaciones de esa zona, incluido la muestra proveniente del valle de Azapa. Estos individuos están distribuidos en dos haplotipos principales, donde uno se diferencia del otro por el polimorfismo 16172C. También comparten estos haplotipos con individuos del centro del país, San Felipe- Los Andes y Santiago. Dentro del haplogrupo H, también destacan algunos linajes compartidos solo por poblaciones del centro y sur del país como aquellos pertenecientes al haplogrupo H1c y H5a3a1.

Siguiendo con otros haplotipos de la red, la rama formada por el macrohaplogrupo JT se encuentra representada por individuos del centro y sur de Chile, como ocurre con

los haplotipos J1c2e2 y T2b3+151, éste último característico de Punta Arenas. El haplotipo T2b37a3 también es frecuente en poblaciones del sur del país, pero compartido por individuos de Iquique y Antofagasta. Los haplogrupos W1 y X2d se restringen a la zona centro y sur de Chile, destacando la presencia de estos linajes en ciudades como Osorno y Puerto Montt. En cuanto al haplogrupo K, tanto los linajes K1b1a2 y K1a4a1a+195 son compartidos mayoritariamente por individuos de Concepción, Chillán, Temuco y Santiago.

Por último, el macrohaplogrupo U muestra una distribución geográfica que varía según linaje. En el caso de los haplotipos U5a, estos están presentes en individuos del norte y sur del país, como Antofagasta, La Serena, Concepción y Chillán. Para el haplogrupo U8b se observa que es compartido por individuos del norte y centro como Arica, Iquique, San Felipe-Los Andes y Santiago. Por otro lado, los portadores de linajes U4 residen en ciudades del centro y sur de Chile, destacando Santiago, Curicó y Punta Arenas.

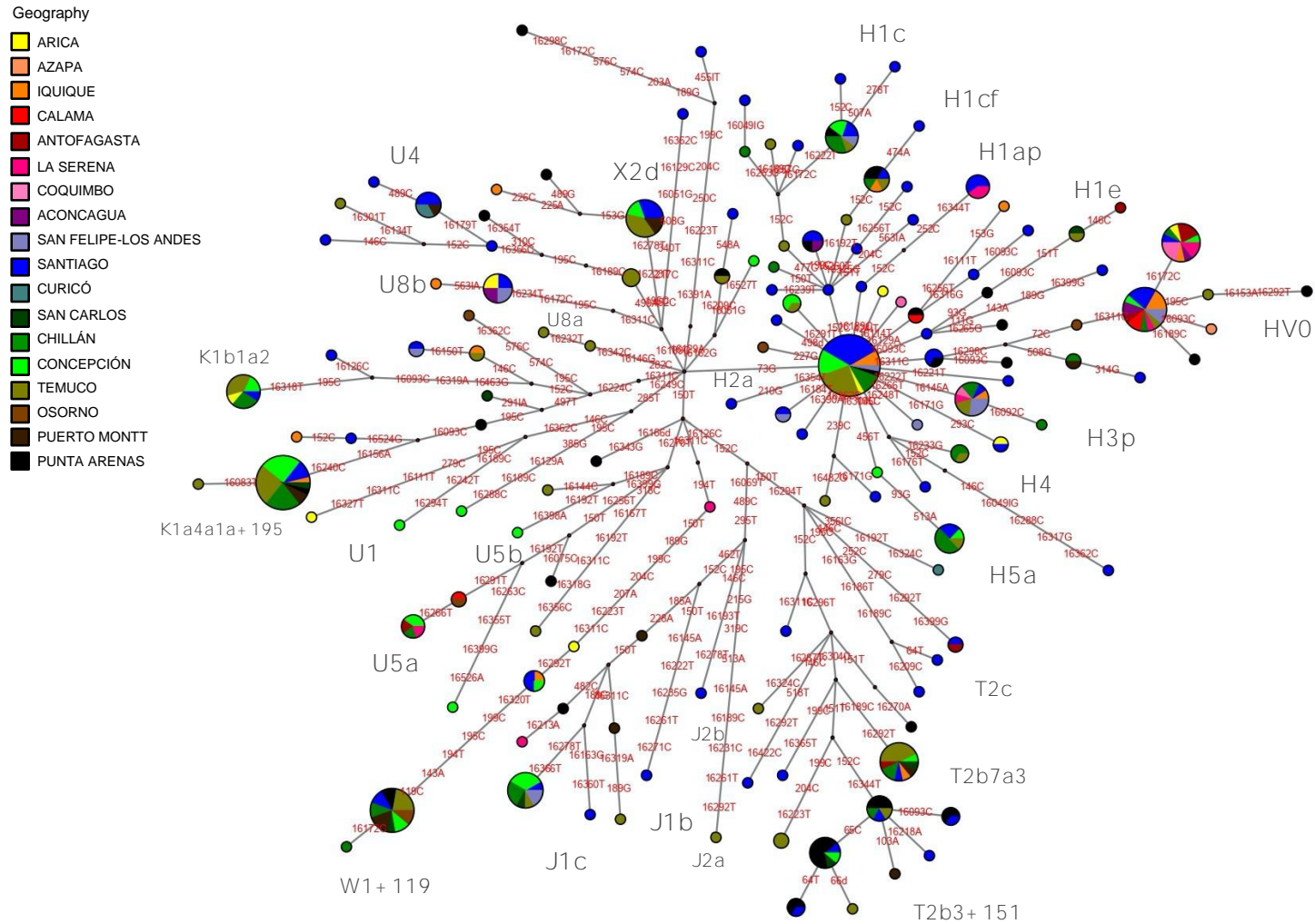


Figura 24: Red de haplotipos de haplogrupos euroasiáticos de occidente en poblaciones chilenas.

La figura 25 muestra la red haplotípica del macrohaplogrupo L construida a partir de 101 individuos, sin existir representación en las poblaciones de Azapa, Osorno y Puerto Montt. En esta red destaca el haplotipo L1b1a+189 como el más frecuente y que concentra a la mayoría de los individuos del centro y sur del país, principalmente de Chillán. Existe un individuo de Iquique portador de este haplotipo, cuyo lugar de nacimiento de la madre es San Carlos. El resto de la red se caracteriza por la presencia de varios haplotipos únicos y con una gran cantidad de pasos mutacionales. De acuerdo con la leyenda, estos haplotipos están simbolizados en su mayoría con colores cálidos correspondientes a poblaciones norteñas. También se observan individuos de Santiago distribuidos en distintos linajes del haplogrupo L. Siguiendo con los haplotipos compartidos, en la rama del haplogrupo L3d se observa el haplotipo L3d2b compartido solo por individuos del norte de Chile, Iquique, Antofagasta y La Serena. Para el haplogrupo L2a se identifica un haplotipo L2a1+143 compartido por individuos de distintas zonas del país. De las poblaciones del sur, Temuco es aquella que presenta más haplotipos compartidos o únicos, además de L1b, encontrándose en linajes L1c, L3e, L3d y L0a. Para este último haplogrupo es el único de la zona, ya que el resto corresponde a individuos del norte del país (Arica, Antofagasta y Calama).

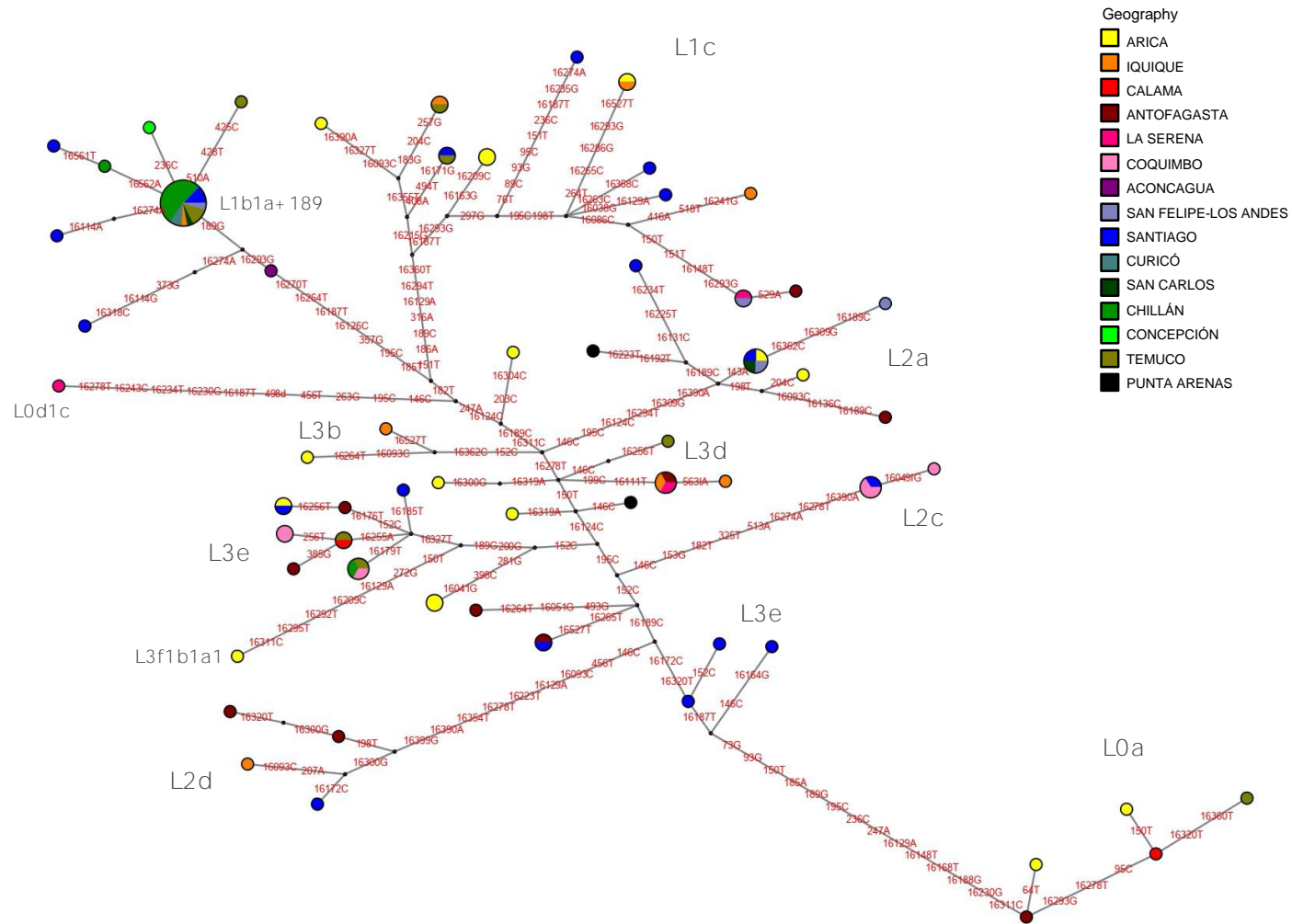


Figura 25: Red de haplotipos del macrohaplogrupo L en poblaciones chilenas.

Se construyó una red de haplotipos para 15 individuos que poseen linajes de origen asiático. En la figura 26 se observa que estos haplotipos son en su mayoría únicos y distantes entre sí. Además, de acuerdo con la simbología, estos haplotipos se restringen solo a poblaciones del centro y sur de Chile, con la excepción de dos individuos de Iquique. Dentro del haplogrupo D no amerindio, destacan los linajes compartidos D4b y D4e que están representados por individuos de Temuco y Concepción, y Temuco y Chillán respectivamente.

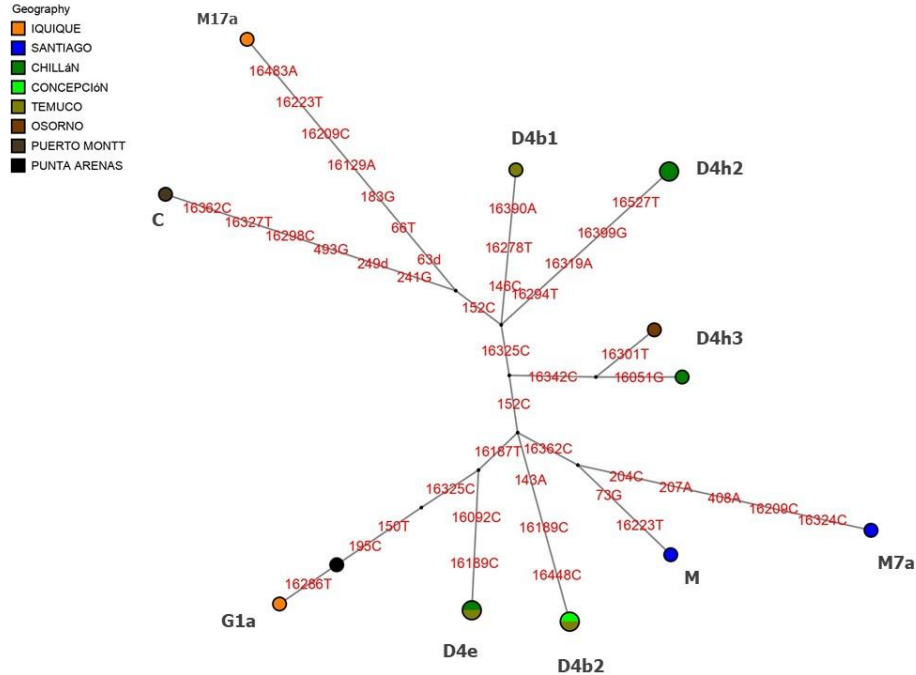


Figura 26: Red de haplotipos de haplogrupos asiáticos en poblaciones chilenas.

3.2.1.6. Comparación con marcadores informativos de ancestría (AIMs)

Las proporciones de ancestría genética individual de cada población chilena se resumen en la tabla 7 donde se disponen los valores promedio para cada ancestría

estimada. Al observar el componente africano (YRI), las poblaciones nortinas Coquimbo, Arica e Iquique son las que presentan las medias más altas en Chile, le siguen otras poblaciones del norte y centro del país, mientras que los valores disminuyen hacia el sur. En cuanto al componente europeo (CEU), la mayoría de las poblaciones poseen una contribución cercana o superior al 50%, siendo Punta Arenas y Santiago aquellas con los valores promedio más altos. La ancestría amerindia para Chile se encuentra dividida en un componente norte y uno centro sur. En general, las poblaciones tienen un mayor componente nativo del centro sur del país, a excepción de Arica, donde la contribución de una ancestría nativa del norte es mayor. Las poblaciones sureñas de Temuco y Puerto Montt poseen los valores promedio más altos para este componente nativo del sur de Chile.

Tabla 7: Proporciones de ancestría promedio para cada población chilena considerando solo individuos con linajes mitocondriales no amerindios.

Población	N	YRI	CEU	NAT_NORTE	NAT_CSUR
Arica	25	0,0665668	0,53298968	0,23948948	0,160954
Iquique	16	0,0576003	0,48321744	0,24098413	0,21819825
La Serena	7	0,051564	0,519561429	0,124263714	0,304610714
Coquimbo	9	0,07419344	0,503827	0,20400322	0,20095056
Santiago	89	0,0465391	0,59786017	0,13003339	0,21653375
Chillán	60	0,04189528	0,55207387	0,157274	0,24278072
Temuco	55	0,0322622	0,51527784	0,14426731	0,30350364
Puerto Montt	16	0,03631025	0,48537044	0,12326531	0,35505419
Punta Arenas	19	0,03286984	0,68655084	0,07012342	0,21045595

Ancestrías: YRI = Africana; CEU = Europea; NAT_NORTE = Amerindia del norte; NAT_CSUR = Amerindia del Centro Sur.

El análisis factorial de datos mixtos permitió comparar las proporciones de ancestría individual con el origen de ADN mitocondrial, evaluando la relación entre linajes africanos, europeos y asiáticos. En la figura 27, la dimensión 1 explica el 29,3% de la inercia y separa a la variable de ancestría europea (CEU) del resto de grupos ancestrales que se agrupan en el lado derecho. La dimensión 2 que explica el 25,6% de la inercia permite diferenciar los componentes amerindios del norte de Chile (NAT_NORTE) junto al componente africano (YRI) hacia el lado positivo del eje Y, mientras que el componente amerindio del centro sur (NAT_CSUR) tiene mayor peso en el lado negativo de la dimensión 2. Se evaluaron las cinco primeras dimensiones con el fin de diferenciar a la ancestría africana del resto, pero en todas se encontró asociado a otra ancestría.

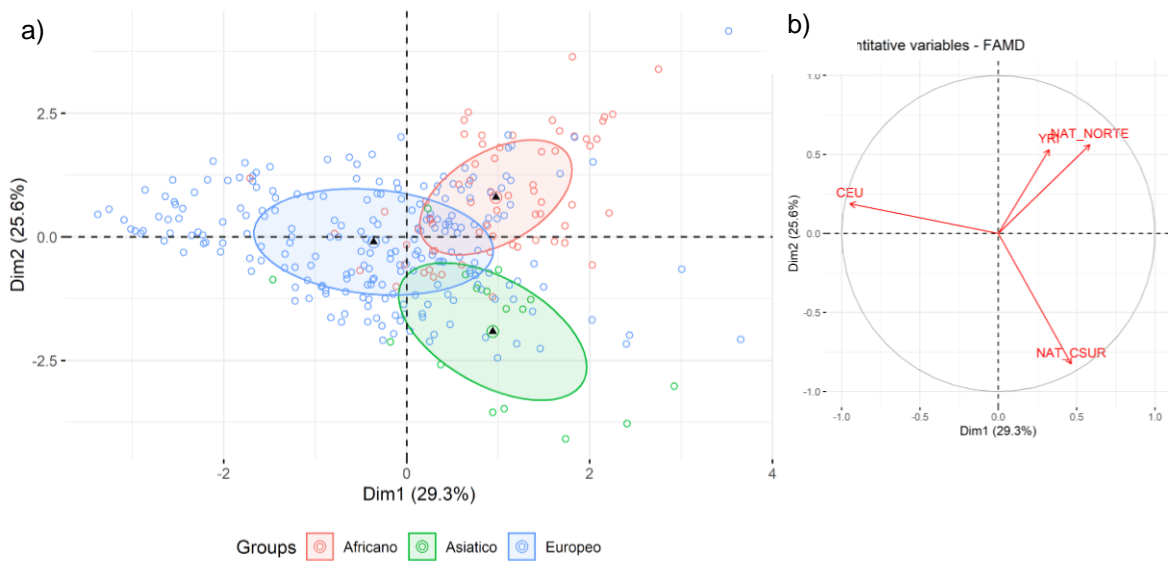


Figura 27: Análisis factorial de datos mixtos para los grupos de linajes mitocondriales africanos, europeos y asiáticos (a). Se indican las variables cuantitativas sobre las dimensiones 1 y 2 (b).

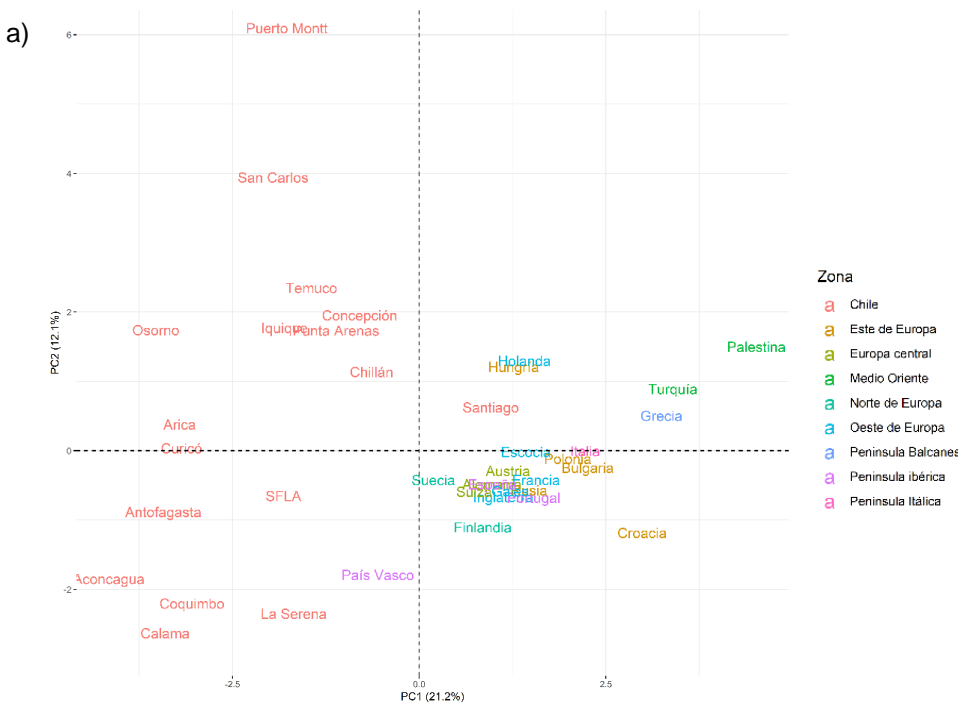
Los individuos portadores de linajes mitocondriales europeos se posicionan mayoritariamente en el lado izquierdo del eje X, asociado con una ancestría europea, mientras que los linajes africanos se agrupan en el lado contrario, asociado a las ancestrías africana y nativa del norte. Por otra parte, los linajes asiáticos se diferencian en la dimensión 2 y se asocian al componente nativo del sur de Chile. Se identificaron individuos que se alejan del resto de su grupo, como en el caso de los linajes europeos donde algunos se ubican al lado contrario de la ancestría europea, es decir, donde tienen mayor peso el componente africano y amerindio del norte. Al revisar a qué linajes mitocondriales corresponden se encontró a los haplogrupos H1c, H1e1a1, H3z, X2d, T2b3+151, T2b7a3, K1a4a1a+195, entre otros. Dentro de los linajes europeos también destacan aquellos que poseen una proporción de ancestría europea mayor al promedio, encontrando que la mayoría provienen de las ciudades de Punta Arenas y Santiago y poseen haplogrupos derivados de H. Por otro lado, se observaron individuos con linajes mitocondriales africanos asociados a un mayor grado de ancestría europea, que son portadores de los haplogrupos L1b1a1'4, L1b1a+189, L1c1d, L1c3a.

3.2.2. Distribución geográfica y origen de linajes mitocondriales no amerindios en Chile

3.2.2.1. Linajes euroasiáticos en Chile y Europa

Luego de caracterizar la diversidad de haplogrupos mitocondriales no amerindios en Chile y las distintas poblaciones analizadas, en la presente sección se evalúa esta diversidad en contexto con otras poblaciones fuera de Chile con la finalidad de responder

a los objetivos 2 y 3 planteados. Estos hacen referencia a la distribución espacial de los linajes y el posible origen de la diversidad mitocondrial no amerindia observada a lo largo del país. Se realizó un Análisis de Componentes Principales con poblaciones europeas ilustrado en la figura 28. En este análisis se decidió incluir dos poblaciones del Medio Oriente, Turquía y Palestina, por los antecedentes de inmigración al país en siglos recientes. En la figura 28a, la variación contenida en PC1 separa a las poblaciones chilenas de las europeas, a excepción de Santiago que muestra una diversidad mitocondrial similar a la observada en países de Europa. Por su parte, el país vasco se agrupó junto a poblaciones chilenas hacia el lado izquierdo del componente 1. En cuanto al PC2, la mayoría de las poblaciones chilenas se estructuran geográficamente, diferenciándose las del norte y centro norte de aquellas del centro y sur del país. Por otra parte, gran parte de las poblaciones de Europa se agrupan juntas, mientras que los países de Medio Oriente junto a Grecia, Holanda y Hungría se separan en el componente 2.



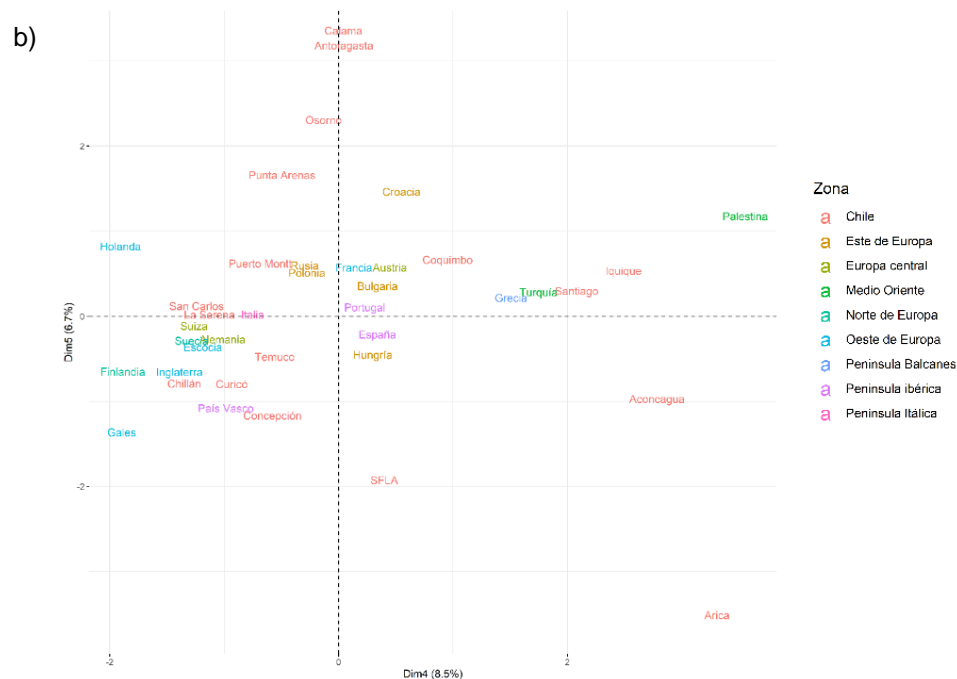


Figura 28: Análisis de Componentes Principales de poblaciones chilenas y europeas, incluidas dos poblaciones de Medio Oriente, a partir de frecuencias de haplogrupos mitocondriales euroasiáticos. En a) se grafican los componentes 1 y 2 y en b) los componentes 4 y 5.

En los PC4 y PC5 que en conjunto explican un 15,2% de la varianza (Figura 28b) se encontró una mayor diferenciación de las poblaciones en cada caso. Sin embargo, debido el bajo porcentaje de variación explicado, las relaciones están dadas por la presencia de linajes raros. En el componente 4 se observa que las poblaciones de Medio Oriente se ubican cerca de Santiago e Iquique por la presencia de linajes R0a y M. También se rescata la relación observada entre las poblaciones de Calama, Antofagasta y Punta Arenas con Croacia hacia el lado positivo del eje Y.

Adicionalmente se realizó un PCA solo con poblaciones de España para evaluar cómo se relaciona la diversidad mitocondrial en base a frecuencias observada en Chile con 4 regiones españolas descritas por Barral-Arca y col (2016): Costa Cantábrica, Plateu Ibérico, Costa Mediterránea y Andalucía. Se presenta en Anexos la figura S3 con los componentes 4 y 5 que explican el 8,5% y 6,8% de la varianza respectivamente, combinación donde se diferenciaron las poblaciones de España. En el PC5 se observa que Andalucía se diferencia de las demás regiones del país y se agrupa junto a la mayoría de las poblaciones chilenas, cerca de Santiago, Antofagasta, Calama y Aconcagua. Hacia el otro lado del eje, se encuentran las tres regiones de España y las poblaciones chilenas de Arica, Curicó, Osorno y Puerto Montt, donde tiene mayor peso los haplogrupos U4, T*, H1 y K*.

Se creó una red de haplotipos para el haplogrupo HV que incluye a H, HV0 y V para poblaciones chilenas y europeas. La figura 29 muestra una red en forma de estrella con un haplotipo principal que de acuerdo con el D-loop pertenece al haplogrupo H2a2a y coincide con la secuencia rCRS. Este linaje es compartido por individuos de España, Italia, Alemania e Inglaterra y de él deriva el resto de los haplotipos presentes. En general, se encontraron haplotipos compartidos por estos países europeos. Destacan dos linajes que se han descrito como autóctonos del país vasco, H2a5a1 y H1e1a1 (Behar y col., 2012b) y en el primer caso, se encontró una secuencia de dicho país coincidente con el haplotipo encontrado en Concepción y Temuco, y también en España (Andalucía). Sobre el segundo haplogrupo, se encontraron haplotipos distintos para dicho linaje y aquellos compartidos con población europea son de España.

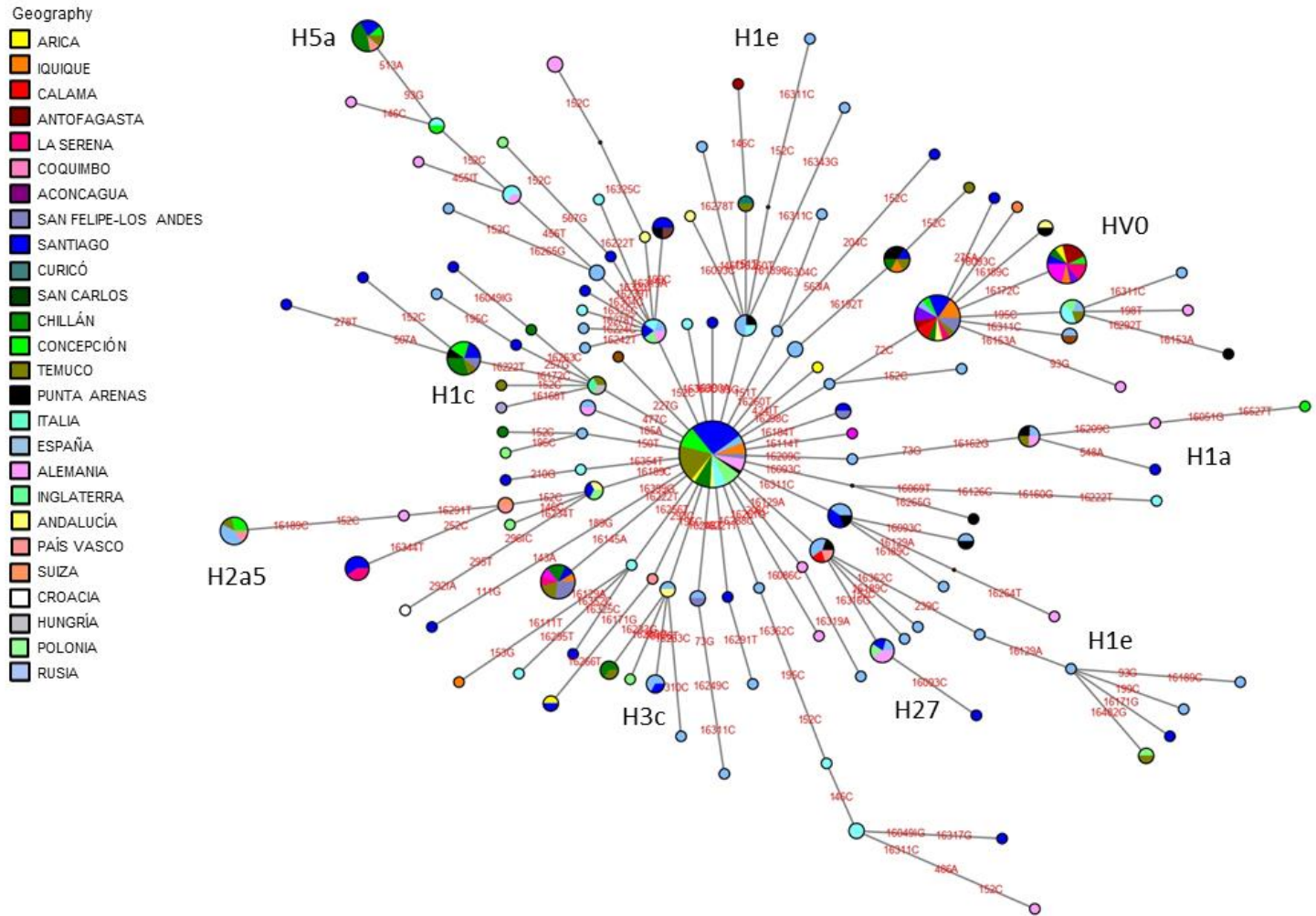


Figura 29: Red de haplotipos del macrohaplogrupo HV en poblaciones chilenas y europeas. Ver Anexo Tabla S5 para detalle de muestras europeas consideradas.

3.2.2.2. Linajes africanos en Chile y África

Se realizó un PCA a partir de frecuencias incluyendo a poblaciones del continente africano (Figura 30). Se etiquetaron según región de acuerdo con la denominación que recibieron durante la trata de esclavos (Figura 1). Se decidió incluir una muestra de Andalucía, ya que en el sur de España se han reportado linajes del norte de África producto del flujo migratorio a través del mediterráneo (Hernández y col. 2015), lo que convierte a España en un posible origen de los linajes que arribaron a Chile.

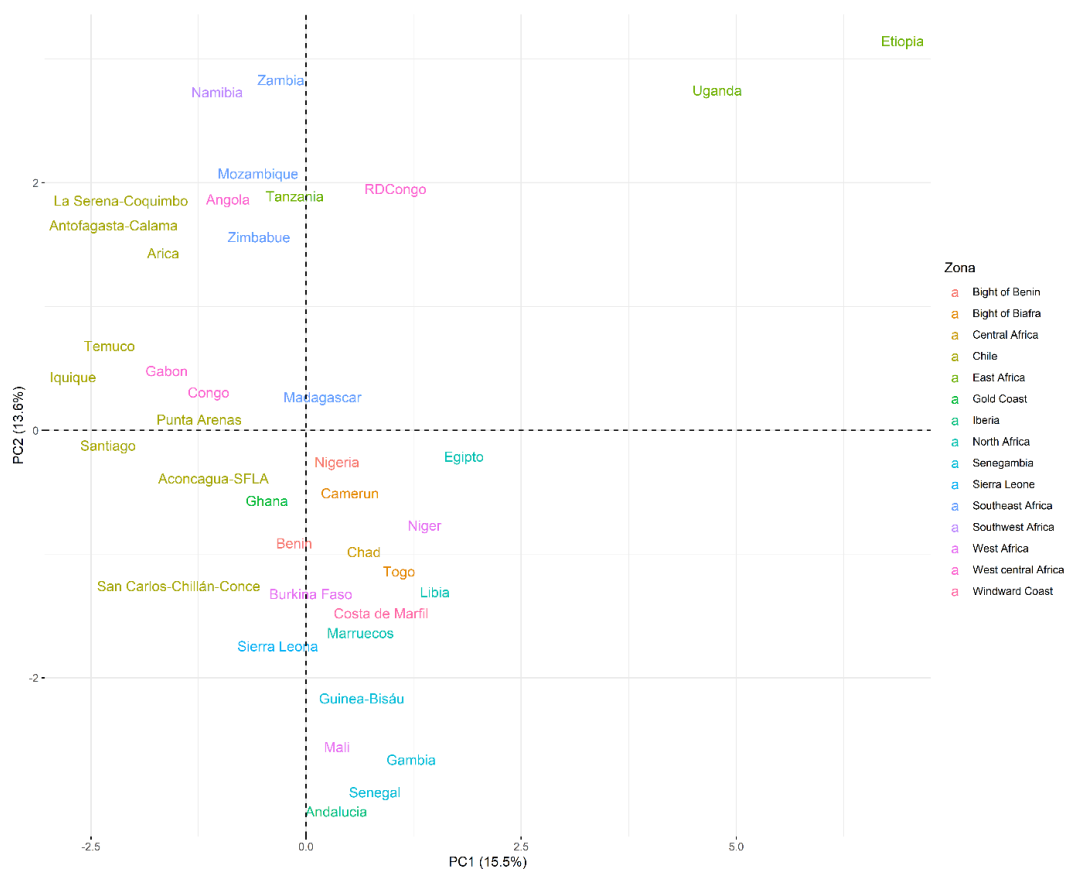


Figura 30: Análisis de Componentes Principales de poblaciones chilenas y africanas, incluida la población europea de Andalucía, a partir de frecuencias de haplogrupos mitocondriales L-U6. Más detalle sobre las regiones a la que pertenecen los países en Figura 1 y Anexo S4. Se utilizó la función *Repel* para facilitar la lectura de las poblaciones en el plano.

En la figura 30, el componente 1 que explica el 15,5% de la varianza separa principalmente a las poblaciones del este de África, Etiopía y Uganda, del resto. En este componente también se observa una diferenciación entre las poblaciones chilenas y africanas, pero menos pronunciada que lo ocurrido con las poblaciones europeas en la figura 28. En el PC2 (13,6% de la varianza) las poblaciones nortinas de Arica, Antofagasta-Calama y La Serena-Coquimbo están cercanas a Angola, de la costa centro occidental de África y también de Mozambique, Zimbabue y Tanzania al este del país africano. Hacia el otro lado del PC2 se ubican las poblaciones de Chillán-San Carlos-Concepción, San Felipe-Los Andes-Aconcagua y Santiago, donde también se encuentran los países de Senegambia, Sierra Leona, Burkina Faso, Costa de Marfil y Marruecos donde tienen mayor peso los haplogrupos L1b, U6 y L2c.

Para establecer posibles relaciones sobre el origen de los linajes africanos encontrados en Chile a nivel de haplotipo, se creó una red de haplotipos considerando gran parte de las poblaciones consideradas en el análisis anterior. Dada la cantidad de poblaciones y de mutaciones presentes en la red que dificultan la visualización del resultado, se decidió presentar la *network* sin los polimorfismos y con las poblaciones africanas simbolizadas de un mismo color, cian (Figura 31). Sin embargo, se indican los haplotipos compartidos a continuación. Al observar la red de la figura 31 destaca el haplotipo más frecuente que corresponde al linaje L1b1a+19. Éste es compartido por 4 individuos africanos de Ghana, Camerún y Angola. Este linaje deriva de L1b encontrado en un individuo de Aconcagua, 3 de Ghana y 1 de Costa de Marfil. En cuanto al haplogrupo L1c, vemos un individuo de Arica cuyo d-loop coincide con 2 muestras de Angola, una de Gabón y otra de Mozambique, y corresponde al subhaplogrupo L1c3b1a

que se ha descrito previamente en estos países (Pereira y col. 2001; Quintana-Murci y col. 2008; Coelho y col. 2009). En esta rama también se observa otro haplotipo compartido por un individuo de Iquique y Angola que pertenecen al haplogrupo L1c1b.

En la rama del haplogrupo L2, vemos un haplotipo perteneciente al linaje L2a1+143 compartido por las poblaciones de Arica, San Felipe-Los Andes, Santiago y San Carlos, y de la costa africana en las regiones de Senegambia (Guinea Bisáu), Costa de oro (Ghana), Ensenada de Benin (Togo) y también del sureste africano (Mozambique). L2a es un haplogrupo con una amplia distribución en África y presenta sus frecuencias más altas en los países de Ghana, Sudán y Mozambique. En cuanto a L2a1+143 estaría asociado a poblaciones del este de África (Silva y col. 2015). También se encontró un linaje compartido entre una muestra de San Felipe-Los Andes y Guinea Bisáu que corresponde al haplogrupo L2a1i, descrito con un origen y distribución en el centro y oeste de África (Silva y col. 2015).

El haplogrupo L3 y sus subhaplogrupos exhiben una alta diversidad en poblaciones chilenas y también en poblaciones africanas, observándose varios haplotipos compartidos. En el caso de L3b, se encontró el linaje L3b2 en una muestra de Iquique que comparte el haplotipo con un individuo de Camerún. El haplogrupo mayor L3b tendría un origen en África central, sin embargo, el linaje L3b2 es un suhaplogrupo raro con menores frecuencias que su linaje hermano (L3b1) (Soares y col., 2012).

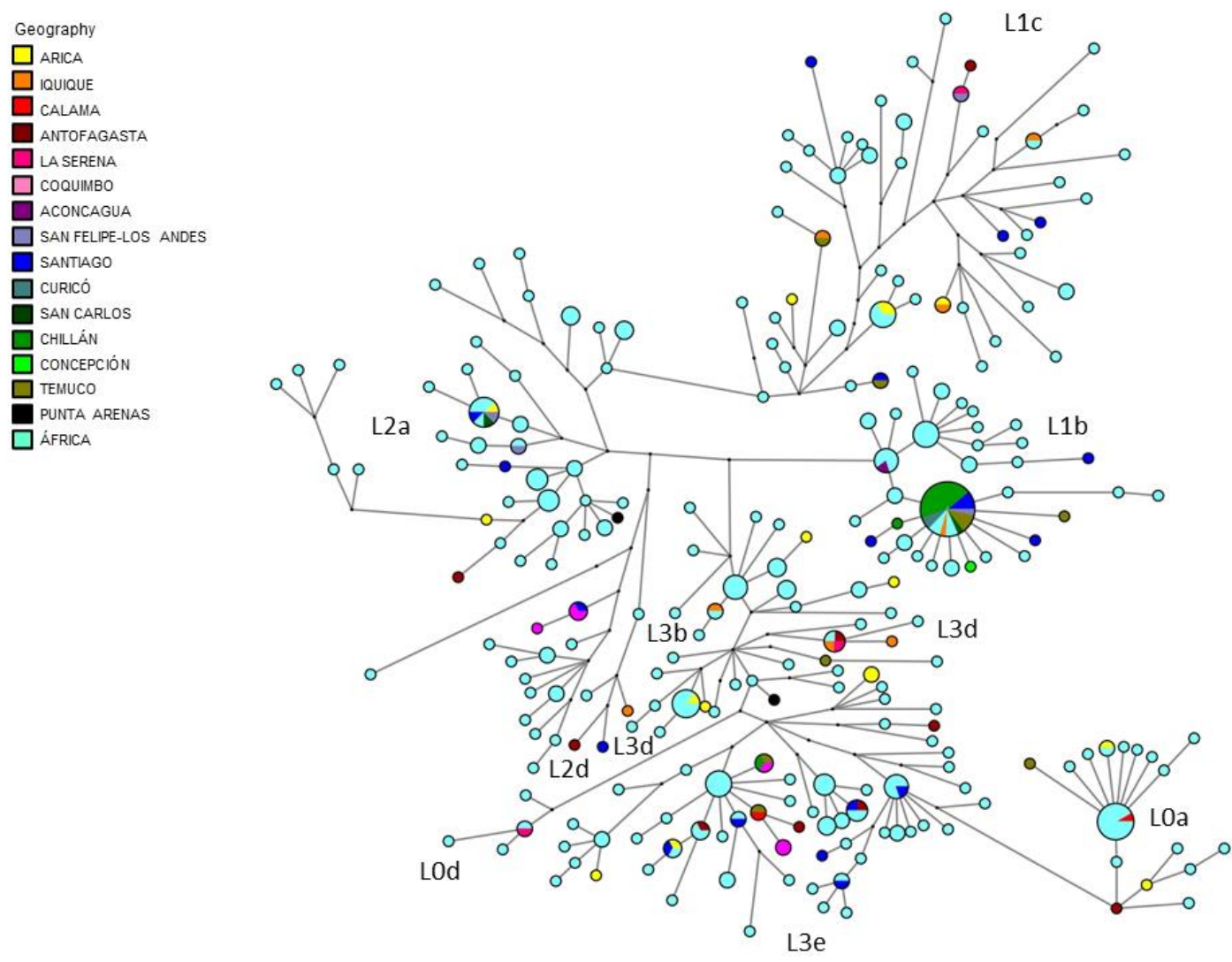


Figura 31: Red de haplotipos del macrohaplogrupo L en poblaciones chilenas y africanas. Ver Anexo Tabla S6 para detalle de muestras africanas consideradas.

En el caso del haplogrupo L3d, éste también tendría un origen en el centro del continente y se ha encontrado en la parte occidental y oriental de África (Soares y col. 2012). En la red se encuentra un haplotipo compartido (L3d2b) por poblaciones del norte de Chile, Iquique, Antofagasta, y La Serena, y en Guinea-Bisáu de la región Senegambia. De este haplogrupo mayor también se encontró en Arica una secuencia D-loop perteneciente al linaje L3d1a1a que coincide con individuos de Angola, Mozambique y Madagascar. En cuanto al haplogrupo L3e, se observa haplotipos compartidos que derivan de otros representados solo por individuos africanos. En el caso de los haplotipos compartidos, las poblaciones chilenas presentes son principalmente Santiago, Arica y Antofagasta. Por ejemplo, el haplogrupo L3e1d que presenta un origen posible en el sur de África y también se distribuye por el centro del continente (Soares y col. 2012), está presente en un individuo de Antofagasta y el haplotipo es compartido con Angola y Gabón. Con un paso mutacional de diferencia (16256T) se encuentra otro haplotipo perteneciente al mismo haplogrupo, que es compartido por individuos de Arica, Santiago y Angola. Por su parte, el linaje L3e3b2 es compartido por Antofagasta, Santiago y Gabón. El haplogrupo L3e2b+152, asociado a una distribución en el oeste de África se encontró en Santiago y también en Ghana, Congo, Angola y Nigeria.

Por último, con respecto al haplogrupo L0, vemos al linaje L0a que se separa del resto de la red por 13 pasos mutacionales y que posee distintos linajes asociados. En primer lugar, se encuentra el haplotipo único que corresponde al linaje L0a1'4 encontrado en Antofagasta. De éste deriva un haplotipo que se diferencia de otro haplotipo más frecuente en la red correspondiente al haplogrupo L0a1b1 que se encontró en Calama y también en Mozambique, Madagascar, Camerún y Angola, las dos primeras poblaciones

con más de un individuo. Este haplogrupo es uno de los más frecuentes en el sureste de África (Cerezo y col., 2016). En cuanto al haplogrupo L0d característico del sur de África, un individuo de La Serena posee un haplotipo que pertenece a L0d1c que coincide con una muestra de Mozambique.

3.2.3. Filogenia de linajes mitocondriales no amerindios en Chile a partir de mitogenomas

3.2.3.1. Caracterización de mitogenomas no amerindios encontrados en Chile

La secuenciación NGS del genoma mitocondrial completo permitió caracterizar en mayor profundidad a 45 muestras que, de acuerdo con la región control, eran portadoras de linajes africanos, euroasiáticos y asiáticos. En la Tabla 8 se detalla el haplogrupo de cada muestra descrito a partir del D-loop y del mitogenoma. Se observa que en la mayoría de las muestras los haplogrupos coinciden, debido a que los polimorfismos que definen a esos linajes se encontraban mayormente en la región control, mientras que en otros la resolución más alta del mitogenoma dio cuenta de un linaje derivado o se mejoró la asignación del haplogrupo. Esto último ocurrió con el linaje D4p mencionado previamente y también con el haplogrupo H, donde la región control era similar a la secuencia de referencia rCRS y por tanto fueron caracterizados como haplogrupo H2a2a, pero luego de la identificación de mutaciones en la región codificante se clasificaron con otro haplogrupo dentro del macrohaplogrupo H.

Tabla 8: Muestras secuenciadas por NGS con su respectivo haplogrupo determinado a partir de D-loop y del genoma mitocondrial completo.

ID muestra	Población	Hg D-loop	Hg mitogenoma
ARI088	Arica	L0a1b1	L0a1b1a1
ARI095	Arica	L3e1d	L3e1d1
ARI212	Arica	HV0	V
ARI220	Arica	U8b1a2b	U8b1a2b
CDSJ049	Santiago	U4c1	U4c1
CDSJ172	Santiago	HV0	V1a1b
CDSJ180	Santiago	H5a3a1	H5a3a1
CDSJ219	Santiago	H1c	H1c
CDSJ322	Santiago	L2d+16129	L2d1a
CHI488	Chillán	L1b1a+189	L1b1a3
CLSM007	Santiago	L1b1a+189	L1b1a3a
CLSM022	Santiago	T2b3+151	T2b3+151
COQ127	Coquimbo	L3e1f*	L3e1f*
COQ164	Coquimbo	L2a'b'c'd	L2c
CUR005	Curicó	U4c1	U4c1
CUR021	Curicó	L1b1a+189	L1b1a3a
LA015	Aconcagua	HV0	V
MFMCH051	Santiago	L2a1+143	L2a1c4a1
MFMCH067	Santiago	L3e1a	L3e1a3a
MFMCH160	Santiago	W3a1+199	W5
MPUC084	Santiago	R0a+60.1T	R0a2a
MPUC085	Santiago	K1b1a2	K1b1a2
PT045	Aconcagua	U8b1a2b	U8b1a2b
PT067	Puerto Montt	T2b7a3	T2b7a3
PT102	Puerto Montt	K1a4a1a+195	K1a3a
PUQ041	Punta Arenas	K1a4a1a+195	K1a16
PUQ077	Punta Arenas	H2a2a	H4
PUQ081	Punta Arenas	D4p	D1g
SC071	San Carlos	H1e1a1	H
SER039	La Serena	L0d1c	L0d1c
SER117	La Serena	L1c1'2'4'6	L1c2'4
SER130	La Serena	L3d2b	L3d2a
SER190	La Serena	U5a1b1	U5a1b1
SP030	Concepción	H2a2a	H1
SP125	Concepción	T2b3+151	T2b3+151
SP142	Concepción	U5b2a2	U5b2a2a1
SP171	Concepción	X2d	X2d2
SP205	Concepción	N1b1	N1b1a2

TEM203	Temuco	H1cf	H3
UL034	Osorno	W1+119	W1+119
XL027	San Felipe-Los Andes	H3p	H1
XL068	San Felipe-Los Andes	L2a1+143	L2a1c4a1
XL078	San Felipe-Los Andes	L1c1'2'4'6	L1c2'4
XL118	San Felipe-Los Andes	J1c2e2	J1c2e2
XL188	San Felipe-Los Andes	L1b1a+189	L1b1a3

En la figura S4 de Anexos se encuentra el árbol filogenético generado por Haplogrep 2 que está enraizado en la secuencia de referencia rCRS. Ahí podemos ver a los mitogenomas encontrados en Chile en contexto con otras muestras del mundo que están incluidas en la base de datos de Phylotree. En negro se indican las mutaciones diagnósticas para cada haplogrupo, en azul (*local private mutation*) aquellas que no están asociadas al haplogrupo, pero que se han reportado al menos una vez en Phylotree y en rojo (*global private mutation*) las mutaciones que no están asociadas al haplogrupo y no se encuentran en la base de datos de Phylotree (Kloss-Brandstätter y col. 2011). Los polimorfismos que se encuentran en las puntas del árbol son candidatos a nuevos haplogrupos, por lo que se deben revisar con mayor detalle, como ocurre con las muestras XL078 y SER117 que fueron asignadas al haplogrupo L1c2'4 y presentan una serie de mutaciones privadas locales. En general, la mayoría de las muestras secuenciadas poseen una o más mutaciones privadas y en algunos casos se evidenció el mismo haplotipo en dos muestras que no son emparentadas y se ubican lejanas geográficamente. Se identificaron pares de secuencias idénticas entre ARI212-LA015, ARI220-PT045, CDSJ049-CUR005, CHI488-XL188, CLSM007-CUR021, MFMCH051-XL068, SER117-XL078 (ver tabla 8 y Figura S4 en Anexos). Considerando que la base de datos

de Phylotree no se actualiza desde el año 2016, existe la posibilidad de que a la fecha se hayan descrito nuevos linajes que pueden coincidir con los encontrados aquí.

3.2.3.2. Reconstrucción filogenética de ADN mitocondrial

Los tres métodos empleados para la reconstrucción filogenética de los linajes de ADN mitocondrial recuperaron una topología de árbol similar, por lo que se decidió mostrar solo el árbol de consenso obtenido a partir de inferencia bayesiana (Figura 32). Sin embargo, cada análisis entregó índices de calidad o convergencia que se reportan a continuación. En las figuras S5 y S6 de Anexos se encuentran los árboles filogenéticos obtenidos por los otros métodos.

En primer lugar, por medio del programa PAUP se realizó un test de Homogeneidad de Partición para conocer cómo se comportan las dos regiones del ADN mitocondrial definidas (región control y región codificante). En un máximo de 100 árboles retenidos, se obtuvieron datos generales de la matriz de alineamiento donde se señala que 16076 caracteres son constantes y 267 son informativos desde un punto de vista parsimonioso. En cuanto a los resultados propios del test, se obtuvo un p-value significativo de 0,05, indicando que el largo del árbol y sus caracteres informativos no se distribuyen homogéneamente entre las dos particiones. Esto muestra que una de las particiones predomina en los análisis, ya sea porque varía más rápido o presenta más sitios informativos. Los antecedentes han mostrado que la región control o d-loop representa la zona más variable del ADN mitocondrial y se debe considerar esta información para análisis posteriores.

Máxima Parsimonia:

En el análisis según Parsimonia, la búsqueda heurística en PAUP nos entrega que 267 caracteres fueron incluidos por ser informativos, los que permitieron generar y retener un total de 119 árboles, donde la mejor solución tuvo una longitud de 374. A continuación, se muestran las medidas de resumen para la solución más parsimoniosa encontrada:

- Índice de consistencia (CI) = 0.7219
- Índice de homoplasia (HI) = 0.2781
- Índice de retención (RI) = 0.8866
- Índice de consistencia reescalado (RC) = 0.6400

Máxima Verosimilitud:

A partir del análisis realizado usando *Máximum Likelihood* en RaxML, el mejor árbol obtenido tuvo un valor de *likelihood* de -26883.799754. Este árbol es aquel que maximiza la probabilidad de ser observado dados los datos y parámetros definidos.

Inferencia bayesiana:

Antes de proceder con el método de inferencia bayesiana para la reconstrucción filogenética, fue necesario encontrar un modelo de evolución que mejor se ajuste a cada región del mitogenoma o partición. Para ello se ingresaron los datos en el programa ModelFinder del servidor IQ-Tree, en la tabla 9 se muestran los resultados obtenidos.

Tabla 9: Modelos de sustitución para cada partición junto a sus parámetros obtenidos por IQ-Tree y que fueron ingresados en MrBayes.

Partición	INF	INV	Modelo	Parámetros
DLOOP	62	1008	TIM3+F+I +G4	TIM3{1.89817, 50.9897, 82.6566}+ F{0.299246, 0.330999, 0.137537, 0.232219}+I{0.751898}+G4{0.642683}
REGIÓN CODIFICANTE	205	15058	TN+F+I	TN{83.3978,35.5283}+ F{0.309473, 0.31136, 0.130939, 0.248228}+I{0.858087}

INF: Sitios informativos; INV: Sitios invariantes

Una vez realizado el análisis en MrBayes, se obtuvieron resultados con respecto a los índices diagnósticos de convergencia (Tabla 10) que mostraron ser los adecuados en cada caso. Los valores de ESS (*Estimated Sample Size*) fueron superiores a 200, con un ESS promedio de 89025.91. En el caso del índice PSRF (*Potential Scale Reduction Factor*) todos los valores fueron igual a 1. Estos resultados están indicando que se alcanzó la convergencia y cada parámetro fue bien muestreado. Otro indicador de convergencia es el espacio de *likelihood* de las 4 corridas. Considerando 10 millones de generaciones, el espacio está reducido a un intervalo aproximado de 3 unidades de *likelihood* [(-27204.59) – (-27207.80)], valor indicativo de que se alcanzó la convergencia y que en esta zona acotada llegan las búsquedas con valores óptimos.

Tabla 10: Índices diagnósticos del análisis bayesiano, resumen de todos los parámetros.

Parameter	Mean	Variance	Lower	Upper	Median	min ESS*	avg ESS	PSRF+
TL	0.0375	0.000003	0.0342	0.0409	0.0375	84370.55	89025.91	1.000

En la Figura 32 se observa el árbol filogenético para los linajes de ADN mitocondrial encontrados en Chile. Se escogió la filogenia obtenida por el método bayesiano ya que representa mejor las relaciones filogenéticas entre los linajes y además fue el análisis con mayor número de replicas, lo que le otorga más soporte frente a los otros métodos. Se indican los valores de soporte de nodo para los tres métodos y se distinguió con color los linajes asignados a cierto haplogrupo. En el caso de las muestras que presentaron secuencias idénticas se optó por colapsar la rama (detalle de las muestras en la tabla 8), aunque no fue posible para el par CHI488-XL188, puesto que se encuentran en relación con un linaje derivado.

Considerando como un buen soporte de nodo aquel >95 de valor bootstrap y >0.95 de probabilidad posterior, la filogenia de inferencia bayesiana es la que presentó valores altos en la mayoría de los nodos. A nivel de haplogrupo mayor, se observa que las relaciones filogenéticas al interior de L poseen un alto soporte en general. En la base de la filogenia se encuentra el outgroup RSRS como linaje ancestral a la bifurcación de las dos ramas principales, los haplogrupo L0 y L1'3, éste último dando origen al resto de linajes presentes en la filogenia. En el haplogrupo L3 se observaron discordancias entre los métodos, ya que en el caso de la inferencia bayesiana y máxima verosimilitud el haplogrupo L3d diverge antes del resto de linajes L3 con un buen soporte de nodo (1/99), mientras que en el análisis de parsimonia se produce una politomía entre dichos linajes junto al grupo que deriva de L3.

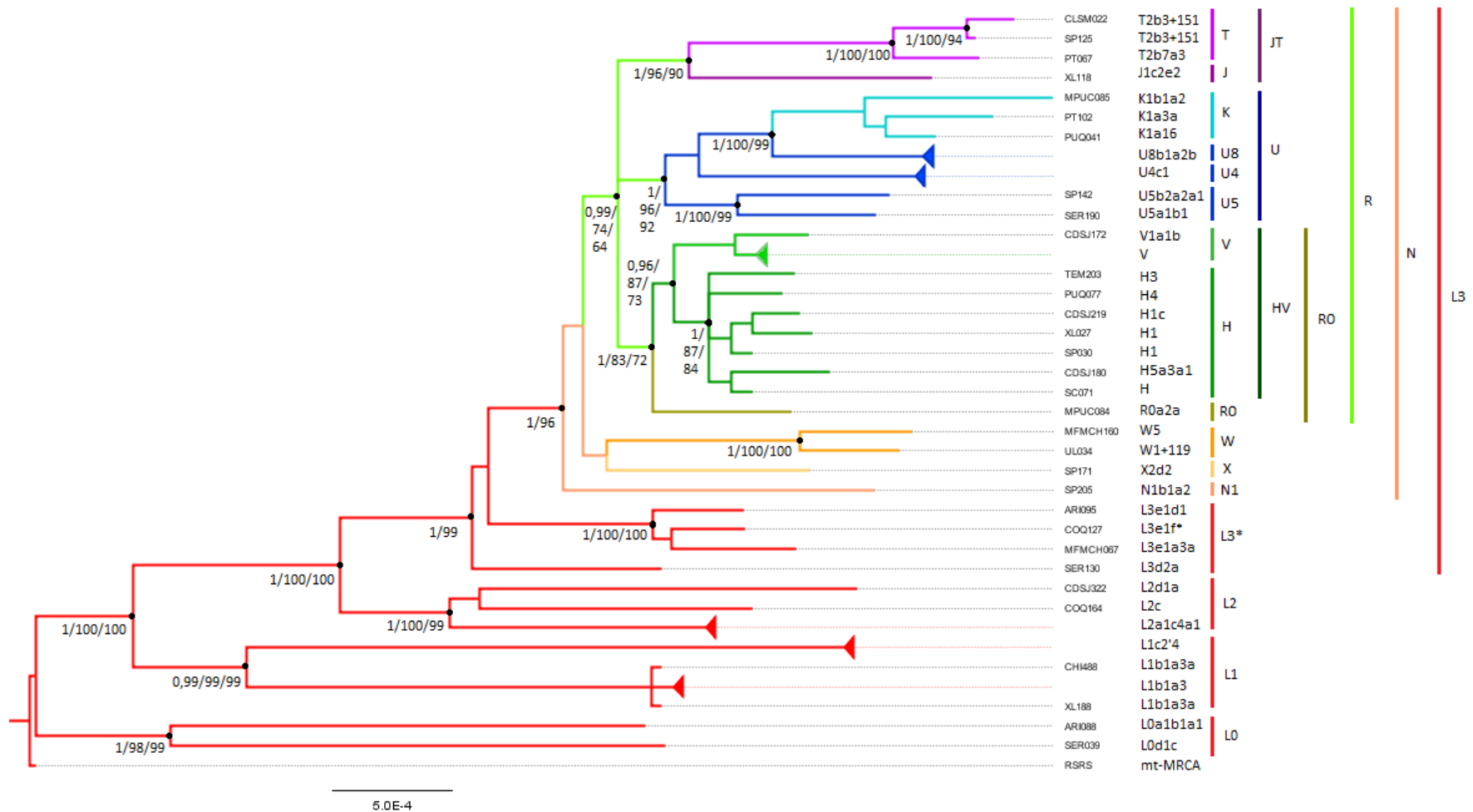


Figura 32: Árbol filogenético de consenso de mayoría obtenido por medio de inferencia bayesiana para linajes mitocondriales no amerindios en Chile, a partir de mitogenomas. Los valores de los nodos representan en orden: la probabilidad posterior del análisis bayesiano y el soporte bootstrap de los análisis de máximo likelihood y parsimonia. El color de las ramas simboliza los haplogrupos asignados y a la derecha se indican los haplogrupos mayores.

Como se ha mencionado anteriormente, del haplogrupo africano L3 se originaron los dos grupos hermanos M y N que se dispersaron fuera de África. En el caso de esta filogenia no se cuenta con linajes derivados de M, pero sí de N. De este haplogrupo, surge a su vez, el haplogrupo R y otros linajes como X y W. Las relaciones observadas entre estos linajes mostraron incongruencias entre los tres métodos utilizados. En el caso del haplogrupo X, tanto en la figura 32 y S6 aparece como linaje hermano del grupo W y con un bajo soporte de nodo (0,6/39). Lo mismo ocurre en la bifurcación dada entre este grupo y el haplogrupo R. Por su parte, en el análisis de parsimonia los linajes X, W y R se disponen dentro de una politomía y el soporte de nodo también es bajo según los estándares definidos (84).

Al interior del haplogrupo mayor R se distinguen tres clústeres principales, el grupo R0 conformado por linajes R0a, H y V, el haplogrupo U que incluye a K y el macrohaplogrupo JT. En el caso del haplogrupo K, éste no presenta soporte de nodo debido a que la topología es distinta entre los métodos. El análisis de parsimonia no resuelve las relaciones de los tres linajes y los presenta con una politomía, mientras que los otros dos métodos separan al linaje K1b1a del grupo K1a con el soporte de nodo más alto (1/100).

4. DISCUSIÓN

4.1. Diversidad mitocondrial no amerindia en Chile

Los resultados obtenidos en este estudio permitieron caracterizar la variabilidad mitocondrial no amerindia encontrada en distintas poblaciones a lo largo de Chile. Se confirmó la baja frecuencia de linajes no amerindios en el pool genético de las poblaciones mixtas de Chile, descrita en estudios previos con este marcador (Gómez-Carballea y col. 2016; de Saint Pierre y col. 2020; Flores-Alvarado y col. 2021; Pezo y col. 2022). A nivel país se obtuvo un 10,2% de haplogrupos no amerindios, proporción que es esperada considerando los procesos de mixtura y flujo migratorio que caracterizaron el periodo de conquista y colonia española (Ruiz-Linares, 2015). Este porcentaje es similar a otros países sudamericanos como Argentina con un 9,59% (Motti y col. 2013) o Paraguay con un 12,8% (Simão y col. 2021), mientras que en otros países esta cifra puede ser menor como es el caso de población mixta en Bolivia con un 1,6%, o mayor como se ha reportado en Río de Janeiro, Brasil con un 74% (Simão y col. 2018). Estas diferencias a nivel genético dan cuenta de las distintas historias locales que han experimentado las poblaciones. Al interior de Chile, estas proporciones también varían por población, de forma que los resultados obtenidos serán interpretados a la luz de la historiografía expuesta en los antecedentes.

La diversidad de linajes observada en el país se puede resumir con fines prácticos en la presencia de los siguientes haplogrupos según frecuencia: H, L, K, T, U, HV (xH), W, J, D, X y otros. Sin embargo, la distribución y frecuencia de éstos varía entre las distintas poblaciones, lo que ha llevado a una estructuración poblacional distinguible en

los diversos análisis realizados con poblaciones chilenas. En primer lugar, el índice de FST realizado entre pares de poblaciones y visualizado a través de un dendrograma, diferenció a las poblaciones en dos grupos, uno formado por las ciudades del norte y centro de Chile, mientras que el otro contiene a las poblaciones del centro-sur y sur del país. Este análisis se llevó a cabo a partir de dos aproximaciones y en ambos casos se observa esta separación principal, aunque algunas relaciones internas cambian. Por ejemplo, en la figura 21a la agrupación de San Carlos, Curicó y Chillán se separa del resto de las poblaciones chilenas, lo que podría explicarse por el haplogrupo L1b1a+189 que presenta frecuencias altas en estas poblaciones. Sus haplotipos asociados generan mayores diferencias entre las secuencias, mientras que en el dendrograma creado a partir de las frecuencias de haplogrupos (Figura 21b) la diversidad se reduce y tiene menor impacto en la diferenciación de las poblaciones.

Los dos PCAs realizados (Figuras 22 y 23) también muestran esta diferenciación geográfica de las poblaciones y fue posible conocer las variables o haplogrupos involucrados. En el primer caso que consideró a todos los haplogrupos, se observó que los linajes africanos (a excepción de L1b) tuvieron mayor peso en las poblaciones del norte y centro norte del país. Santiago, en este análisis se ubica más cerca de las poblaciones del sur donde priman los haplogrupos europeos. Luego, al descartar el haplogrupo L del análisis se mantiene esta estructura, por lo que también existen linajes euroasiáticos que son responsables de esta diferenciación, HV0 en el norte y K para las poblaciones del sur. Estas diferencias observadas en las poblaciones chilenas responden a procesos migratorios y demográficos particulares que caracterizaron la historia local y/o regional de las poblaciones. Si bien el posible origen y llegada a Chile de los distintos linajes no se puede

trazar de manera certera, algunas inferencias pueden realizarse. A continuación, se discuten los resultados para los tres momentos históricos de interés definidos en esta tesis.

4.2. Mujeres migrantes en el periodo de conquista y colonia española

Una de las consecuencias de la conquista europea en Sudamérica como en Chile fue el mestizaje asimétrico entre conquistadores españoles y mujeres indígenas, que hoy se ve reflejado en la proporción de linajes maternos amerindios y no amerindios. Los antecedentes históricos dan cuenta de la llegada de mujeres migrantes desde España al territorio chileno, quienes también fueron parte del proceso de mezcla. En muchos casos, esto se dio en el contexto de captura y cautiverio blanco, práctica que se llevó a cabo de manera recurrente en la zona fronteriza al sur del Biobío, donde mujeres españolas fueron tomadas prisioneras e integradas a la sociedad mapuche (González, 2002).

Si revisamos el lugar de origen de los españoles que viajaron a América y Chile en esta época, las crónicas indican que migraron de diversas regiones de la península Ibérica como Andalucía, Extremadura, región Vasca y Castilla y León, pero sería la región al sur de España, Andalucía, de donde provino la mayoría de los conquistadores y mujeres (Monserrat, 2014; Cussen, 2016). En la Península Ibérica como en Europa, un 40-50% del pool genético mitocondrial está representado por el haplogrupo H (Pereira y col. 2005). En específico, los sublinajes H1 - H3 presentan frecuencias altas en las poblaciones ibéricas, especialmente en la región cantábrica. Por su parte, Andalucía ha mostrado frecuencias similares a otras poblaciones ibéricas, con H1 y H3 sumando un 53% del haplogrupo H (Hernández y col. 2017). En Chile, H1 es el más frecuente, seguido de H2 y H3, sin embargo, es importante mencionar que las frecuencias de los subhaplogrupos de

H pueden estar subestimadas considerando las cifras para el linaje H2a2a. La secuencia de referencia rCRS pertenece a dicho haplogrupo y al momento de trabajar linajes cercanos a esa secuencia, la región control es poco informativa para la asignación de haplogrupos, por lo que se vuelve necesario la caracterización de polimorfismos en la región codificante. En este estudio no se aplicó un protocolo de PCR-RFLP para ello, pero la secuenciación por NGS permitió caracterizar en la resolución más alta a dos muestras que fueron previamente asignadas a H2a2a a partir del D-loop. El procesamiento de los mitogenomas nos entregó resultados que asignaron a uno de los linajes como H1 y el otro como H4.

Para facilitar la lectura de esta sección, nos apoyamos en mapas de isofrecuencias que se presentan a continuación. Es importante mencionar que las frecuencias en cada continente son muy diferentes, por lo que cada mapa dentro de una figura presenta su propia leyenda. De acuerdo con la figura 33, la distribución de linajes H difiere a lo largo del país, concentrándose en algunas poblaciones del centro y sur de Chile, mientras que en el norte se encuentra en baja frecuencia o ausente. Santiago es una de las poblaciones con la frecuencia más alta para el haplogrupo H y no sorprende considerando que durante el periodo colonial se estableció como centro y capital del Reino de Chile, por lo que recibió un gran número de migrantes. Junto a Santiago, las ciudades de Concepción y Chillán también concentran un alto porcentaje de haplotipos H, siendo dos ciudades que se fundaron tempranamente y donde se asentaron españoles y criollos desde los primeros años coloniales. Más al sur, Punta Arenas también presentó una alta frecuencia de haplogrupo H1. Sin embargo, esta ciudad se caracteriza por una fundación tardía que se

remonta al siglo XIX, por lo que se vuelve difícil distinguir entre procesos migratorios más tempranos o tardíos.

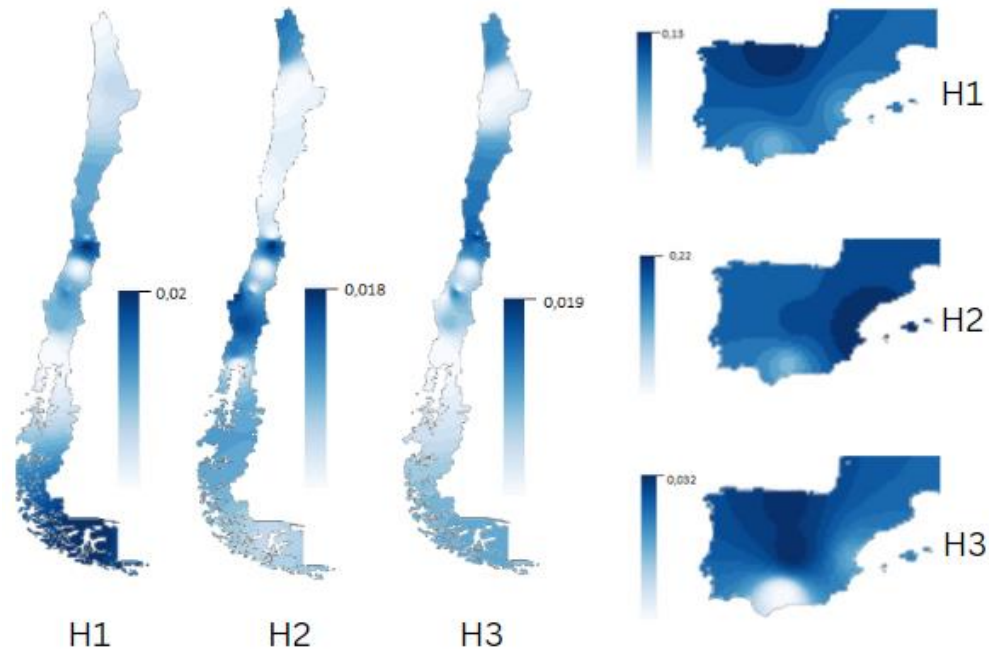


Figura 33: Mapas de isofrecuencia por método de *Kriging* para 3 linajes europeos H en base a la frecuencia de haplogrupos en poblaciones de la Península Ibérica y Chile.

En la Península Ibérica, el país Vasco ha sido foco de estudio por el rol que cumplió como refugio franco-cantábrico en la recolonización postglacial de Europa. Se han identificado linajes autóctonos de esta región y vasco hablantes que corresponden a H1j1, H1t1, H2a5a1, H1av1, H3c2a, and H1e1a1 (Behar y col. 2012b). En Chile se identificaron dos de estos haplogrupos, H2a5a1 y H1e1a1. En la red de haplotipos de la figura 29 se observa que las secuencias pertenecientes a esos haplogrupos coincidieron con muestras de España. La distribución geográfica del haplogrupo HV0 en la Península Ibérica se restringe a la región del país Vasco (Figura 34), lo que apoyaría la idea de que

esta zona fue parte del refugio que permitió la reexpansión de estos linajes por el resto de Europa durante el Último Máximo Glacial (Barral-Arca y col. 2016). El linaje HV0 presenta sus frecuencias más altas en las poblaciones chilenas del norte y centro norte, apoyando la diferenciación de éstas con respecto al resto. Este linaje contiene a V que es difícil de asignar a partir de la región control (van Oven y Kayser, 2009) y así quedó demostrado al momento de secuenciar dos mitogenomas que inicialmente fueron caracterizados a partir de D-loop como HV0 y finalmente fueron asignados como haplogrupo V. De acuerdo con la bibliografía revisada, la región Vasca es considerada como lugar de origen de los conquistadores que llegaron a América, por lo que es posible que este linaje haya llegado en tiempos históricos tempranos cuando el norte era parte del Virreinato del Perú.

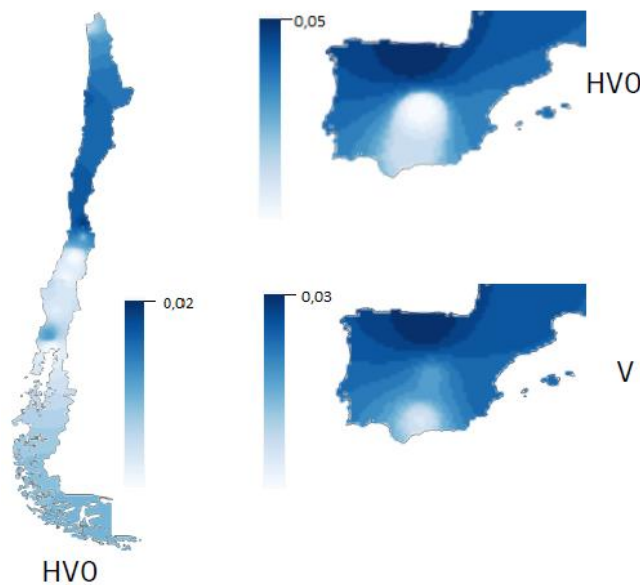


Figura 34: Mapas de isofrecuencia por método de *Kriging* para el linaje europeo HV0 en poblaciones de la Península Ibérica y de Chile, y para el haplogrupo V en la Península Ibérica.

4.3. Mujeres migrantes en el periodo de tráfico de esclavos

El tráfico de esclavos es uno de los hitos más importantes en la migración humana ocurrido en la historia reciente. Durante los siglos XVI y XIX, el desplazamiento forzoso de millones de personas desde el continente africano hacia las Américas y el Caribe tuvo consecuencias sociodemográficas que hacen eco hasta el día de hoy. Aproximadamente, un tercio de las personas esclavizadas que embarcaron desde África eran mujeres (Eltis y Richardson, 2010), quienes también fueron parte del proceso de mestizaje con población europea o indígena. Desde la literatura se menciona que en un comienzo la relación entre españoles y negras, mulatas y zambas era aceptada, reportándose el matrimonio y/o nacimiento de hijos entre ellos (Mellafe, 1984; Soto, 1992). Con respecto al lugar de origen de los esclavos africanos es difícil de conocer, ya que no se tiene registro de ello y la información que existe al respecto es sobre el lugar de embarque de los esclavos. En la figura 1 se presentan las denominaciones que recibieron las distintas regiones de la costa africana.

En Chile, el haplogrupo mayor L es el segundo más frecuente y sus sublinajes se distribuyen de distinta manera en el país, observando frecuencias más altas en el norte de Chile. Allí habrían llegado esclavas con diversos linajes africanos, zona que pertenecía al Virreinato del Perú, entidad político-territorial donde la demanda de mano de obra esclava era mayor solicitando esclavos a Chile y Argentina para trabajos agrícolas y el servicio doméstico (Mellafe, 1984, Vial, 1957). Entre los distintos linajes, destaca el haplogrupo L0a que presenta una distribución asociada al este del continente africano y a pesar de su origen en esa zona, igual se encuentra presente en el centro y sur de África (Rito y col.,

2013). En el caso chileno, este linaje se encuentra en baja frecuencia, pero con una distribución principal en el norte de Chile, a excepción de un individuo de Temuco (Figura 35). De acuerdo con la historiografía, el tráfico de esclavos se habría extendido hasta la costa sureste de África a comienzos de 1800, cuando la demanda en las Américas significó expandir los lugares de captura de esclavos, convirtiéndose en la segunda zona de embarque más importante después de la costa centro oriental durante esa fecha (Eltis y Richardson, 2013). Es posible que las mujeres esclavas portadoras de estos linajes tuvieran un origen en esa zona de África y hayan llegado a Chile en este periodo, pero tampoco se puede descartar su presencia en un momento previo a lo descrito aquí o que haya llegado desde otra zona del continente africano.

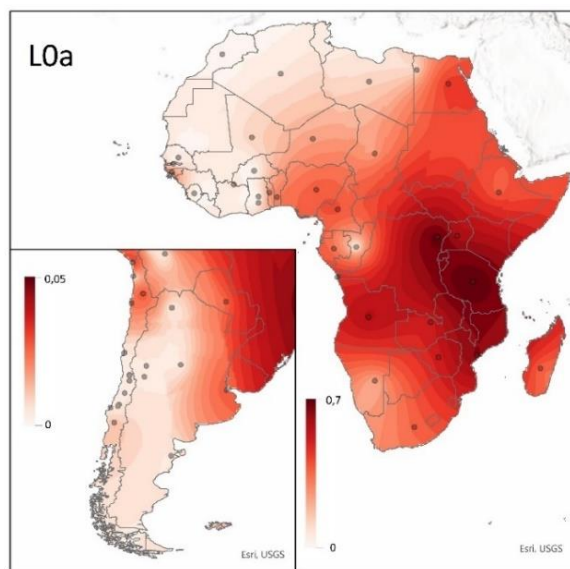


Figura 35: Mapas de isofrecuencia por método de *Kriging* para linaje africano L0a en base a su frecuencia en poblaciones de África y Chile.

Otro linaje que posee una distribución particular es el haplogrupo L1b (Figura 36). Este linaje es común en la costa africana, pero alcanza sus frecuencias más altas en el

sector de Senegambia, Sierra Leona y Camerún (Fortes-Lima, 2015). También se encuentra en las poblaciones del sur de la península ibérica, Andalucía, en frecuencias del 2,53%. Por su parte, otro linaje africano U6 también está presente, alcanzado frecuencias más altas de 8,86% (Hernández y col. 2017). La presencia de estos linajes da cuenta del flujo genético ocurrido a través del mediterráneo y en el caso del haplogrupo L se ha asociado al tráfico de esclavos y el rol que tuvieron Portugal y España en ese proceso. También puede relacionarse con la conquista árabe de la península ibérica o a eventos demográficos más ancestrales (Cerezo y col., 2012; Barral-Arca y col. 2016).

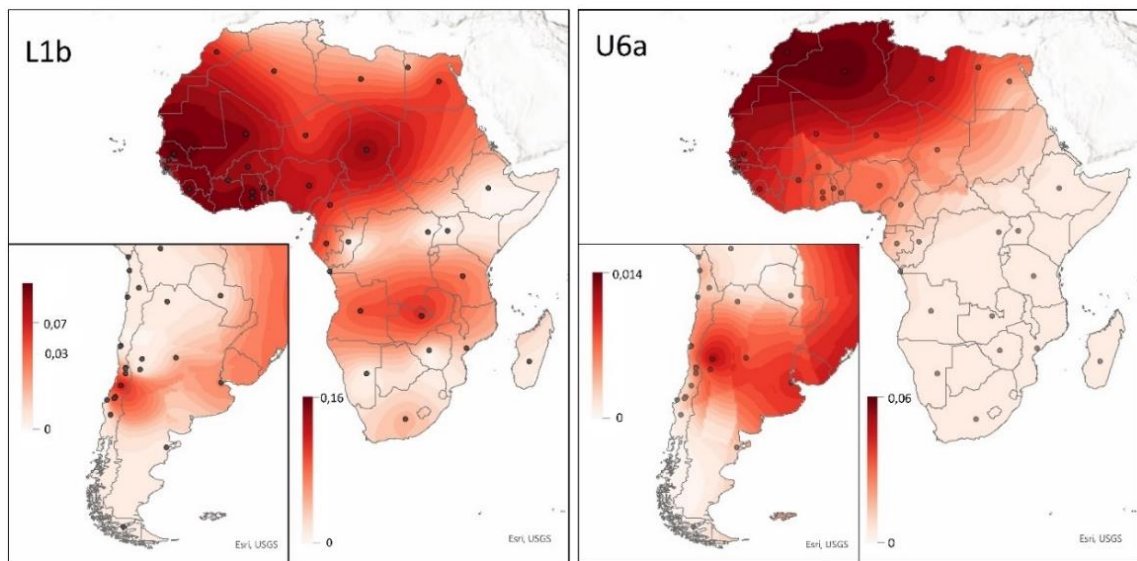


Figura 36: Mapas de isofrecuencia por método de *Kriging* para linajes africanos L1b y U6a en base a sus frecuencias en poblaciones de África y Chile.

En el caso chileno, el haplogrupo L1b está representado mayoritariamente por el linaje L1b1a+189 que posee una distribución restringida al sur de Chile, en Chillán y ciudades aledañas. Si bien es posible que el linaje L1b llegara con mujeres esclavas provenientes del sector de Senegambia y la costa occidental africana, aquí se propone que

este linaje llegara en el genoma de las españolas durante el periodo de conquista y colonia española. Si se consideran los resultados obtenidos en el análisis de datos mixtos, se observó a individuos con linajes L1b que presentaron una mayor ancestría europea que otros. De esta manera, se plantea la posibilidad de que se hayan mezclado tempranamente en el continente europeo, desde donde habrían migrado mujeres españolas a territorio chileno. Además, los antecedentes revisados para Chillán no nos muestran información en cuanto a la presencia africana en la ciudad, por lo que la alta representatividad de este linaje en la zona de Ñuble se relaciona con un efecto fundador que ocurrió con la llegada y establecimiento de españoles en el periodo colonial. Este posible origen europeo de los haplogrupos africanos también se considera para el linaje U6, que se encuentra en bajas frecuencias y presentó una distribución más amplia en Chile (Figura 36).

En ciudades aledañas a Chillán como lo es Concepción, estudios históricos como bioantropológicos han visibilizado la presencia africana en tiempos históricos, dando cuenta del aumento de esclavos durante el siglo XVIII en dicha zona, especialmente mujeres (Andrade, Argo y Soto-Cerda, 2022). Sin embargo, llama la atención que solo se encontrara en un individuo portador de linaje africano en la ciudad, pero es posible que el muestreo haya influido en la visibilidad de estos linajes.

Los linajes representados en los mapas de la figura 37 y figura S7 poseen distribuciones más amplias y se encuentran principalmente en el norte y centro de Chile. Existen otros linajes como L3e que se extiende a otras ciudades al sur del país y también suele ser el más frecuente en poblaciones de Sudamérica (Schaan y col., 2017).

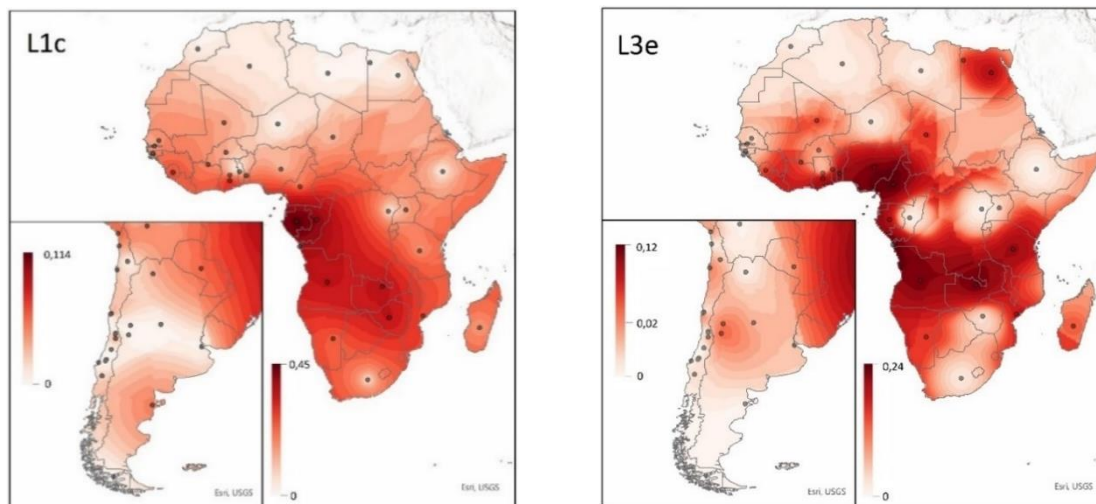


Figura 37: Mapas de isofrecuencia por método de *Kriging* para los linajes africanos L1c y L3e en base a sus frecuencias en poblaciones de África y Chile.

4.4. Mujeres migrantes en los siglos XIX y XX

Volviendo a los haplogrupos euroasiáticos, la diversidad de linajes no amerindios presentes en las poblaciones estaría explicado por procesos migratorios posteriores, como los vividos a partir de la segunda mitad del siglo XIX y durante la primera del XX que contribuyeron a un aporte foráneo mayor. Este periodo se caracterizó por un auge de la inmigración europea a nivel país y sudamericano, la que tuvo distintos objetivos a lo largo del país. Por ejemplo, en el norte de Chile estas oleadas migratorias estuvieron marcadas por personas que decidieron migrar en busca de oportunidades laborales, estabilidad económica y/o política, quienes llegaron en el contexto minero del salitre y cobre, y se asentaron en la región (González y col., 2016). En el caso del sur de Chile, luego de la Guerra del Pacífico, el Estado Chileno pudo volver a impulsar el poblamiento del sur de Chile con colonos europeos. En este caso se propicia un contexto familiar, mientras que

en el norte el patrón de migración está masculinizado, pero eso no impidió que también viajaran mujeres como pudimos ver en las figuras 3 y 4 en Antecedentes.

Tanto la literatura como los resultados obtenidos en el análisis de componentes principales con población europea (Figura 28a) nos indican que en Europa existe un patrón de frecuencias para los haplogrupos mitocondriales mayores que es similar entre sí. Por otro lado, este análisis también evidenció que a Chile llegó una combinación particular de linajes mitocondriales no amerindios que vincula a todas las poblaciones chilenas y que las distingue de la diversidad presente en las poblaciones europeas. En este caso, Santiago fue la excepción, al presentar una diversidad similar a la de Europa, siendo una ciudad cosmopolita que ha recibido una gran variedad de migrantes, lo que se ve reflejado en una alta diversidad haplotípica alcanzada. En la figura 28b, Santiago se agrupa cerca de países del Medio Oriente, principalmente por la presencia de linajes menos frecuentes en Europa occidental como R0a+60.1T o linajes M. En el caso de R0a+60.1T, este posee una distribución en la región arábiga (Gandini y col. 2016), lugar desde donde pudieron provenir migrantes a Chile en el contexto de las migraciones árabes.

En cuanto a la distribución de los linajes europeos en Chile (Figura 19) se observa que en el sur del país existe una mayor diversidad de estos. El linaje K mostró una distribución diferencial (Figura 38), concentrando al linaje K1a4a1a+195 en poblaciones del sur de Chile. Gómez-Carballa y col. (2016) reportó este dato como señal de un efecto fundador en poblaciones de Concepción y Temuco y aquí también se encontró en las ciudades de Chillán, San Carlos, Puerto Montt y Santiago. El haplogrupos K1a4a1a tendría un origen europeo (Costa y col. 2013) y posee una distribución en dicho continente y Cercano Oriente. En el caso de Europa, su distribución se da en el norte y centro del

continente, con frecuencias en Gran Bretaña, Irlanda, Alemania, Hungría, entre otros. Es posible que estos linajes llegaran al país de esos lugares de origen, como británicas, suecas, finlandesas y alemanas en el contexto de la colonización del sur.

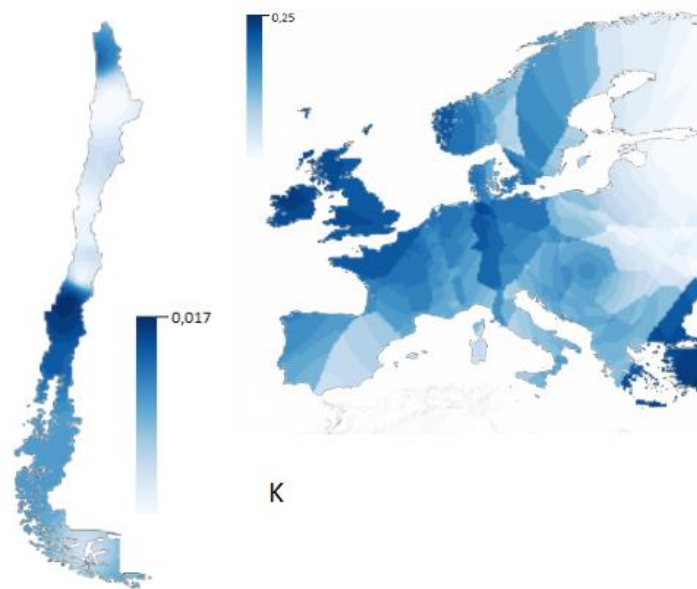


Figura 38: Mapas de isofrecuencia por método de *Kriging* para linaje europeo K en base a su frecuencia en poblaciones de Europa y Chile.

También destaca el haplogrupo T2 (Figura 39) que se encuentra en el sur del país, con la frecuencia más alta en la ciudad de Punta Arenas y representado por los linajes T2b3+151 y T2b7a3. El haplogrupo T2b es el más común al interior de T2 y posee una distribución por el área mediterránea y Europa central (Barral-Arca y col. 2016; Pala y col. 2012), por lo que se pueden inferir patrones migratorios desde esa zona en el contexto del poblamiento reciente de la ciudad.

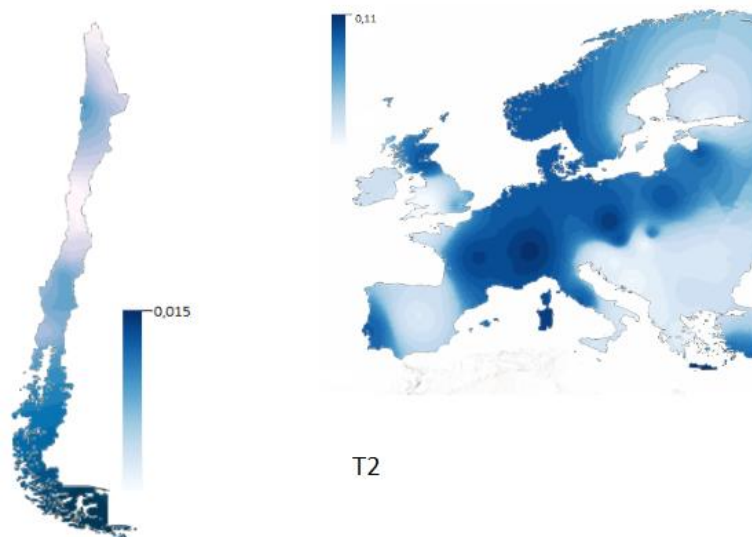


Figura 39: Mapas de isofrecuencia por método de *Kriging* para linaje europeo T2 en base a su frecuencia en poblaciones de Europa y Chile.

Con respecto a las poblaciones del norte y centro-norte del país, desde Arica a San Felipe-Los Andes, presentan frecuencias para el haplogrupo U8b1a2b, mientras que no se reportó en el sur de Chile (Figura S8). Este linaje se ha asociado a poblaciones del Cáucaso, Medio Oriente y los Balcanes (Cabrera, 2022). Considerando los antecedentes para esta zona del país, es posible que este subhaplogrupo esté relacionado con las migraciones árabes ocurridas a fines del siglo XIX hasta la caída del imperio Otomano en 1914. Si bien se asentaron dispersamente por el país y a nivel de provincia, destaca su presencia en el Norte Chico (Cussen, 2016).

Considerando los antecedentes que registran las oleadas migratorias de croatas hacia las ciudades de Antofagasta y Punta Arenas, llama la atención que no se encontraron haplotipos compartidos entre dichas ciudades, pero sí entre Punta Arenas y Calama, de la misma región de Antofagasta, correspondiente al haplogrupo H1e+16129. Para ahondar

en este linaje, se ingresó una de las secuencias fasta a BLAST y se encontró más de una coincidencia del 100% con muestras antiguas de la región vasca, donde es descrito como H1j1, linaje autóctono de la región Franco-Cantábrica. Este linaje tiene altas frecuencias en el norte de Iberia y también está presente en poblaciones con ancestría vasca (Palencia-Madrid y col. 2017). En el estudio de Flores-Alvarado et al. (2021) con la población de Punta Arenas se describieron haplogrupos presentes en población croata, pero no se puede descartar su llegada desde otras migraciones europeas.

Con respecto a las migraciones asiáticas, la baja frecuencia de linajes de dicho origen en Chile concuerda con lo descrito desde la historiografía, ya que estas migraciones fueron menores en número y se suelen comportar como colonias, contrayendo matrimonio entre personas de la misma comunidad.

4.5. Consideraciones finales y proyecciones

La alta variabilidad mitocondrial junto a la alta movilidad que ha caracterizado a las poblaciones europeas se convirtió en una limitación para este estudio, ya que no permitió distinguir entre los procesos descritos desde la historiografía en Chile. Por una parte, no fue posible diferenciar entre procesos más tempranos y recientes, y, por otro lado, los linajes mitocondriales fueron poco informativos para una distribución geográfica específica. Al complementar la información entregada por el ADN mitocondrial con datos sobre la ancestría genómica fue posible identificar individuos portadores de linajes no amerindios que presentaron un perfil de ancestría indicativo de mezcla más reciente.

Los resultados obtenidos al comparar las poblaciones chilenas con países de Europa y África dan cuenta de que en Chile llegaron linajes mitocondriales particulares que corresponden a una submuestra de la diversidad que hoy se observa en esos continentes. Considerando lo obtenido para el país, se vuelve interesante comparar la diversidad mitocondrial no amerindia a nivel regional, incluyendo a otros países sudamericanos. De esta forma, para el caso de los linajes africanos se podría indagar en la presencia de patrones migratorios entre los lugares que fueron parte de las rutas esclavistas en América.

El procesamiento de los mitogenomas mostró cómo la máxima resolución que entrega la secuenciación completa del ADN mitocondrial permite una detallada clasificación filogenética de los linajes, que a partir del D-loop se ve limitada. Además, este tipo de datos entrega información sobre la diversidad genética que permite analizar en mayor profundidad la historia demográfica de las poblaciones. En cuanto a la obtención de mitogenomas, el trabajo de laboratorio contempló algunos aspectos metodológicos que son importantes de considerar. A partir de la amplificación del ADN mitocondrial en varios fragmentos y utilizando la enzima Taq, DNA polimerasa fue posible generar librerías genómicas para ser secuenciadas por NGS, a pesar de que los ADNs genómicos se encontraban parcialmente degradados. Durante la aplicación del protocolo, la pérdida de ADN en la purificación es un aspecto que puede evitarse, mediante la fragmentación del ADN por más tiempo, de modo que no sea necesario una purificación doble. A pesar de estas consideraciones técnicas, se obtuvo una buena cobertura de la secuencia del mitogenoma, en longitud como profundidad, entregándole soporte a las variantes identificadas. El uso de mitogenomas abre una ventana a nuevos análisis que permitan

establecer relaciones más claras entre los haplogrupos no amerindios en Chile y en poblaciones africanas y euroasiáticas. Como se mencionó previamente, la región control se vuelve poco informativa para la asignación de ciertos haplogrupos, por lo que resulta necesario complementar el análisis con la caracterización de polimorfismos en la región codificante, si es que la obtención de mitogenomas no es posible.

5. CONCLUSIONES

El análisis de D-loop y mitogenomas en distintas poblaciones mixtas a lo largo de Chile permitió caracterizar la diversidad mitocondrial no amerindia en el país, de manera que se puso a prueba la hipótesis que indicaba la contribución de las mujeres migrantes en la conformación de las poblaciones chilenas, evidenciando una distribución geográfica diferencial de linajes de ADN mitocondrial. A continuación, se resumen las principales conclusiones obtenidas:

- Las poblaciones chilenas mostraron una estructura poblacional que separa geográficamente a las poblaciones que habitan el norte y centro de Chile, de aquellas ubicadas al sur del país. Este resultado fue evidenciado en todos los análisis realizados.
- Esta diferenciación de las poblaciones responde a patrones migratorios que fueron particulares a la región y/o población, definiendo una distribución geográfica diferencial de los linajes mitocondriales no amerindios.
- Con respecto a la primera predicción, se encontró una mayor frecuencia y diversidad de linajes provenientes de la Península Ibérica en poblaciones del centro de Chile, especialmente Santiago, Chillán y Concepción.
- Con respecto a la segunda predicción, en el norte de Chile se evidenció una mayor cantidad y diversidad de linajes africanos a consecuencia del tráfico de esclavos en dicha zona, pero se encontró otro foco con una alta frecuencia del haplogrupo

africano L1b en el sur de Chile, presente principalmente en Chillán y ciudades aledañas.

- De acuerdo con la tercera predicción se identificaron linajes mitocondriales que poseen una distribución geográfica restringida a ciertas poblaciones del país, asociado a los procesos migratorios post-mestizaje ocurridos durante los siglos XIX y XX.
- Este estudio permitió evidenciar que las mujeres que migraron voluntaria o forzosamente a territorio chileno durante tiempos históricos tempranos como recientes se establecieron e integraron en la sociedad y formaron parte del proceso de mestizaje, generando una impronta a nivel genético en las poblaciones chilenas.

6. BIBLIOGRAFÍA

- Aboukhalid, R., Sturk-Andreaggi, K., Bouabdellah, M., Squalli, D., Irwin, J. A., & Amzazi, S. (2013). Mitochondrial DNA control region variation from samples of the Moroccan population. *International journal of legal medicine*, 127(4), 757-759.
- Achilli, A., Rengo, C., Magri, C., Battaglia, V., Olivieri, A., Scozzari, R., ... & Torroni, A. (2004). The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *The American Journal of Human Genetics*, 75(5), 910-918.
- Achilli, A., Perego, U., Bravi, C., Coble, M., Kong, Q., Woodward, S., ... y Bandelt, H. (2008). The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies. *PLoS One*, 3(3), e1764.
- Agar, L. (2015). Migraciones externas en Chile: bases históricas de un fenómeno complejo. *Oasis*, 22, 49–91.
- Adhikari, K., Mendoza-Revilla, J., Chacón-Duque, J. C., Fuentes-Guajardo, M., y Ruiz-Linares, A. (2016). Admixture in Latin America. *Current opinion in genetics y development*, 41, 106-114.
- Anderson, S., Bankier, A., Barrell, B., de Bruijn, M., Coulson, A., Drouin, J., Eperon, I., Nierlich, D., Roe, B., Sanger, F., Schreier, P., Smith, A., Staden, R., Young, I. (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290, 457– 465.
- Andrade, P., Argo, M. y Soto-Cerda, A. (2022). The presence of Africans and their descendants in colonial Concepción, Chile: a first approach. *Revista de Historia de América*, (162), 13-56.
- Andrews R., Kubacka I., Chinnery P., Lightowlers R., Turnbull D., Howell N. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Gene*, 23, 147.
- Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online].
- Apata, M. (2015). *Señales de adaptación humana al hidroarsenicismo en los valles andinos de la Región de Arica y Parinacota: Evidencias de variantes protectoras del Gen Arsénico [+3] Metiltransferasa* (Tesis de Pregrado). Universidad de Chile, Santiago, Chile.
- Badro, D., Douaihy, B., Haber, M., Youhanna, S., Salloum, A., Ghassibe-Sabbagh, M., ... y Wells, R. (2013). Y-chromosome and mtDNA genetics reveal significant contrasts in affinities of modern Middle Eastern populations with European and African populations. *PloS One*, 8(1), e54616.

- Ballester, B., San Francisco, A. y Gallardo, F. (2010). Modo de vida y economía doméstica de las comunidades cazadoras recolectoras costeras del desierto de Atacama en tiempos coloniales y republicanos. *Taltalia*, (3), 21-32.
- Bandelt, H., Forster, P. y Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*, 16, 37-48.
- Bailliet, G., Rothhammer, F., Carnese, F. R., Bravi, C. M., & Bianchi, N. O. (1994). Founder mitochondrial haplotypes in Amerindian populations. *American Journal of Human Genetics*, 55(1), 27.
- Barbieri, C., Whitten, M., Beyer, K., Schreiber, H., Li, M. & Pakendorf, B. (2012). Contrasting maternal and paternal histories in the linguistic context of Burkina Faso. *Mol Biol Evol*, 29(4); 1213-1223.
- Barbieri, C., Guldemann, T., Naumann, C., Gerlach, L., Berthold, F., Nakagawa, H., Mpoloka, S.W., Stoneking, M. & Pakendorf, B. (2014). Unraveling the complex maternal history of Southern African Khoisan populations. *Am J Phys Anthropol*, 153(3); 435-448.
- Barnett, D. W., Garrison, E. K., Quinlan, A. R., Strömberg, M. P., y Marth, G. T. (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics*, 27(12), 1691-1692.
- Barral-Arca, R., Pischedda, S., Gómez-Carballa, A., Pastoriza, A., Mosquera-Miguel, A., López-Soto, M., ... y Salas, A. (2016). Meta-analysis of mitochondrial DNA variation in the Iberian Peninsula. *PLoS One*, 11(7), e0159735.
- Batini, C., Lopes, J., Behar, D. M., Calafell, F., Jorde, L. B., Van der Veen, L., ... & Comas, D. (2011). Insights into the demographic history of African Pygmies from complete mitochondrial genomes. *Molecular biology and evolution*, 28(2), 1099-1110.
- Behar, D., Villems, R., Soodyall, H., Blue-Smith, J., Pereira, L., Metspalu, E., ... & Rosset, S. (2008). The dawn of human matrilineal diversity. *The American Journal of Human Genetics*, 82(5), 1130-1140.
- Behar, D., van Oven, M., Rosset, S., Metspalu, M., Loogväli, E-L., Silva, N., Kivisild, T., Torroni, A., Villems, R. (2012a). A “Copernican” Reassessment of the Human Mitochondrial DNA Tree from its Root. *Am J Hum Genet*, 90, 675–684.
- Behar, D., Harmant, C., Manry, J., Van Oven, M., Haak, W., Martinez-Cruz, B., ... & Genographic Consortium. (2012b). The Basque paradigm: genetic evidence of a maternal continuity in the Franco-Cantabrian region since pre-Neolithic times. *The American Journal of Human Genetics*, 90(3), 486-493.
- Bengoa, J. (2000). *Historia del pueblo Mapuche*. Lom Ediciones.
- Bisselle, L. (2019). *El papel de mujeres españolas en la conquista y colonización de las Américas: las mujeres más importantes en el Pacífico sudamericano durante el siglo XVI* (Tesis de honor). UVM College of Arts and Sciences.

- Bittmann, B. (1977). *Cobija y alrededores en la época colonial (1600-1750)*. *Actas del VII Congreso de Arqueología de Chile* (págs. 327-352). Altos de Vilches: Kultrum.
- Bodner, M., Perego, U. A., Huber, G., Fendt, L., Röck, A. W., Zimmermann, B., ... y Parson, W. (2012). Rapid coastal spread of First Americans: Novel insights from South America's Southern Cone mitochondrial genomes. *Genome research*, 22(5), 811-820.
- Brandstätter, A., Peterson, C. T., Irwin, J. A., Mpoke, S., Koech, D. K., Parson, W., & Parsons, T. J. (2004). Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database. *International Journal of Legal Medicine*, 118, 294-306.
- Brandstätter A, Klein R, Duftner N, Wiegand P, Parson W. (2006). Application of a quasi-median network analysis for the visualization of character conflicts to a population sample of mitochondrial DNA control region sequences from southern Germany (Ulm). *Int. J. Legal Med.*, 120, 310–314.
- Brisighelli, F., Alvarez-Iglesias, V., Fondevila, M., Blanco-Verea, A., Carracedo, A., Pascali, V. L., ... & Salas, A. (2012). Uniparental markers of contemporary Italian population reveals details on its pre-Roman heritage. *PLoS One*, 7(12), e50794.
- Brown, W., Matthew, G. y Wilson, A. (1979). Rapid Evolution of animal mitochondrial DNA. *Proceedings of the National Academy of science of the United States of America*, 79, 1967-1971.
- Brucato, N., Cassar, O., Tonasso, L., Tortevoeye, P., Migot-Nabias, F., Plancoulaine, S., ... & Dugoujon, J. M. (2010). The imprint of the Slave Trade in an African American population: mitochondrial DNA, Y chromosome and HTLV-1 analysis in the Noir Marron of French Guiana. *BMC evolutionary biology*, 10(1), 1-19.
- Cabrera, V. (2022). Updating the Phylogeography and Temporal Evolution of Mitochondrial DNA Haplogroup U8 with Special Mention to the Basques. *DNA*, 2(2), 104-115.
- Catelli, M., Álvarez-Iglesias, V., Gómez-Carballa, A., Mosquera-Miguel, A., Romanini, C., Borosky, A., ... y Salas, A. (2011). The impact of modern migrations on present-day multi-ethnic Argentina as recorded on the mitochondrial DNA genome. *BMC genetics*, 12(1), 77.
- Castillo, D. (2021). *Contribución de la mujer indígena al poblamiento de la ciudad de San Carlos, Región de Ñuble. Análisis a partir de marcadores de ADN mitocondrial* (Tesis de pregrado). Universidad de Chile.
- Castro, J., Picornell, A., Ramon, M. (1998) Mitochondrial DNA: a tool for populational genetics studies. *Int Microbiol*, 1, 327–332.
- Cerda, P. (1988) Las mujeres en la sociedad fronteriza del Chile colonial. (Ponencia en congreso) *46 Congreso de Americanistas*, Amsterdam.

- Cerezo, M., Achilli, A., Olivieri, A., Perego, U.A., Gomez-Carballa, A., Brisighelli, F., Lancioni, H., Woodward, S.R., Lopez-Soto, M., Carracedo, A., Capelli, C., Torroni, A. & Salas, A. (2012). Reconstructing ancient mitochondrial DNA links between Africa and Europe. *Genome Res*, 22(5); 821-826.
- Cerezo, M., Gusmao, L., Černý, V., Uddin, N., Syndercombe-Court, D., Gomez-Carballa, A., ... y Salas, A. (2016). Comprehensive analysis of pan-African mitochondrial DNA variation provides new insights into continental variation and demography. *Journal of Genetics and Genomics*, 43(3), 133-143.
- Chaitanya, L., van Oven, M., Brauer, S., Zimmermann, B., Huber, G., Xavier, C., ... & Kayser, M. (2016). High-quality mtDNA control region sequences from 680 individuals sampled across the Netherlands to establish a national forensic mtDNA reference database. *Forensic Science International: Genetics*, 21, 158-167.
- Chernomor, O., Von Haeseler, A., y Minh, B. Q. (2016). Terrace aware data structure for phylogenomic inference from supermatrices. *Systematic biology*, 65(6), 997-1008.
- Chinnery, P. (2006) Mitochondrial DNA in Homo Sapiens. En: Bandelt H-J, Macaulay V, Richards M, editores. *Human Mitochondrial DNA and the Evolution of Homo sapiens*. Springer-Verlag.
- Chinnery, P. y Gomez-Duran, A. (2018). Oldies but goldies mtDNA population variants and neurodegenerative diseases. *Frontiers in neuroscience*, 12, 682.
- Cocoş, R., Schipor, S., Hervella, M., Cianga, P., Popescu, R., Bănescu, C., ... y Raicu, F. (2017). Genetic affinities among the historical provinces of Romania and Central Europe as revealed by an mtDNA analysis. *BMC genetics*, 18(1), 1-11.
- Coelho, M., Sequeira, F., Luiselli, D., Beleza, S., y Rocha, J. (2009). On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola. *BMC evolutionary biology*, 9(1), 1-18.
- Córdoba, J. (2018). El viaje femenino a América (1493-1600). *X Congreso virtual sobre Historia de las Mujeres (Octubre de 2018)*. España: Archivo Histórico Diocesano de Jaén.
- Costa, M. D., Pereira, J. B., Pala, M., Fernandes, V., Olivieri, A., Achilli, A., ... & Richards, M. B. (2013). A substantial prehistoric European ancestry amongst Ashkenazi maternal lineages. *Nature communications*, 4(1), 1-10.
- Coudray, C., Olivieri, A., Achilli, A., Pala, M., Melhaoui, M., Cherkaoui, M., ... & Dugoujon, J. M. (2009). The complex and diversified mitochondrial gene pool of Berber populations. *Annals of Human Genetics*, 73(2), 196-214.
- Cussen, C. (2016). Relato histórico de la formación del actual pueblo chileno. En S. Berríos (ed.), *El ADN de los Chilenos y sus orígenes genéticos* (pp. 17-36). Santiago, Chile: Editorial Universitaria.
- de Ramón, A. (2000). *Santiago de Chile (1541-1991): Historia de una sociedad urbana*. Santiago: Catalonia.

- de Saint Pierre, M., Faure, M., y Morales, H. (2020). Análisis histórico genético de la población de la región de Antofagasta revela discontinuidad en la costa y valles interiores. *Estudios atacameños*, (64), 183-198.
- Edler, D., Klein, J., Antonelli, A. y Silvestro, D. (2020). raxmlGUI 2.0: a graphical interface and toolkit for phylogenetic analyses using RAxML. *Methods Ecol. Evol.*, 1–5.
- Edgar, R. C. (2004), MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792-97.
- Eltis, D. y Richardson, D. (2010). *Atlas of the transatlantic slave trade*. New Haven, Conn: Yale University Press.
- Eltis, D. y Richardson, D. (2013). *Routes to Slavery: Direction, Ethnicity and Mortality in the Transatlantic Slave Trade*. New York, USA: Routledge.
- Encina, F. (1940). *Historia de Chile: desde la prehistoria hasta 1891*. Santiago, Chile: Nascimento.
- Esri Inc. (2021). *ArcGIS Pro* (versión 2.9). <https://www.esri.com/es-es/arcgis/products/arcgis-pro/overview>
- Ewels, P., Magnusson, M., Lundin, S., y Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047-3048.
- Excoffier, L. y Lischer, L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10, 564-567.
- Eyheramendy, S., Martínez, F. I., Manevy, F., Vial, C., y Repetto, G. M. (2015). Genetic structure characterization of Chileans reflects historical immigration patterns. *Nature communications*, 6(1), 1-10.
- Fadhlaoui-Zid, K., Martínez-Cruz, B., Khodjet-el-khil, H., Mendizabal, I., Benammar-Elgaaied, A., & Comas, D. (2011). Genetic structure of Tunisian ethnic groups revealed by paternal lineages. *American journal of physical anthropology*, 146(2), 271-280.
- Faure, M. (2018). *Estructura genética mitocondrial en la Región de Antofagasta, Chile* (Tesis de pregrado). Universidad de Chile. Santiago, Chile.
- Feliú, G. (1973). *La abolición de la esclavitud en Chile: estudio histórico y social*. Santiago: Universitaria.
- Fernández, E. (2006) La emigración francesa en Chile, 1875–1914: Entre integración social y mantenimiento de la especificidad. *Amérique Latine Histoire et Mémoire. Les Cahiers ALHIM online*, (12).
- Fendt, L., Zimmermann, B., Daniaux, M., & Parson, W. (2009). Sequencing strategy for the whole mitochondrial genome resulting in high quality sequences. *BMC genomics*, 10, 1-11.

- Fendt, L., Röck, A., Zimmermann, B., Bodner, M., Thye, T., Tschentscher, F., ... & Parson, W. (2012). MtDNA diversity of Ghana: a forensic and phylogeographic view. *Forensic Science International: Genetics*, 6(2), 244-249.
- Flores-Alvarado, S., Orellana-Soto, M., & Moraga, M. (2021). Ancestry and admixture of a southernmost Chilean population: The reflection of a migratory history. *American Journal of Human Biology*, 34, e23598.
- Flores, S. (2016). *Ancestría y Mestizaje de poblaciones Chilota y Croata en Punta Arenas: Un estudio a través de marcadores uniparentales* (Tesis de pregrado). Universidad de Chile, Santiago, Chile.
- Fortes-Lima, C. (2015) *Tracing the genetic origin of African descendants from South America*. (Tesis de doctorado), Université Toulouse III Paul Sabatier.
- Fuentes, M., Pulgar, I., Gallo, C., Bortolini, M. C., Canizales-Quinteros, S., Bedoya, G., ... y Rothhammer, F. (2014). Geografía génica de Chile: Distribución regional de los aportes genéticos americanos, europeos y africanos. *Revista médica de Chile*, 142(3), 281-289.
- Galanter, J., Fernandez-Lopez, J., Gignoux, C., Barnholtz-Sloan, J., Fernandez-Rozadilla, C., y col. (2012). Development of a Panel of Genome-Wide Ancestry Informative Markers to Study Admixture Throughout the Americas. *PLoS Genet*, 8(3): e1002554.
- Gandini, F., Achilli, A., Pala, M., Bodner, M., Brandini, S., Huber, G., ... & Olivieri, A. (2016). Mapping human dispersals into the Horn of Africa from Arabian Ice Age refugia using mitogenomes. *Scientific Reports*, 6(1), 1-13.
- Garrison, E., y Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*.
- Garrison, E. (2015). *vcflib*. In GitHub repository. GitHub. <https://github.com/ekg/vcflib>
- Guerrero, G. (2019). *Linajes maternos: Una aproximación de la distribución de los haplogrupos mitocondriales en la población actual de Osorno* (Tesis de pregrado). Universidad de Concepción.
- Göbel, T. M., Bodner, M., Robino, C., Augustin, C., Huber, G. E., Marra, M., ... & Parson, W. (2020). Mitochondrial DNA variation in Sub-Saharan Africa: Forensic data from a mixed West African sample, Côte d'Ivoire (Ivory Coast), and Rwanda. *Forensic Science International: Genetics*, 44, 102202.
- Gómez, L. (2014). Travesía de las primeras viajeras a la Nueva España. *Temas de Ciencia y Tecnología*, 18(52), 21-29.
- Gómez-Carballa, A., Moreno, F., Álvarez-Iglesias, V., Martín-Torres, F., García-Magariños, M., Pantoja-Astudillo, J. A., ... & Salas, A. (2016). Revealing latitudinal patterns of mitochondrial DNA diversity in Chileans. *Forensic Science International: Genetics*, 20, 81-88.

- Gonder, M., Mortensen, H., Reed, F., de Sousa, A., y Tishkoff, S. (2007). Whole-mtDNA genome sequence analysis of ancient African lineages. *Molecular biology and evolution*, 24(3), 757-768.
- González, J. (2011). Notas sobre la inmigración argentina en la precordillera antofagastina durante el ciclo salitrero. *Estudios atacameños*, (42), 189-204.
- González, J., Lufin, M., y Galeno, C. (2016). Mujeres europeas en el ciclo salitrero de la región de Antofagasta: 1880-1930. La inmigración de italianas, yugoslavas y alemanas. *Historia 396*, (2), 257-292.
- González, Y. (2002). Cautiverio femenino y mestizaje al revés en la Araucanía del siglo XVII. La conquista de la mujer del otro. *Educación y Humanidades*, (10-11), 72-95.
- Guardia, S. (Ed.). (2013). *Historia de las mujeres en América Latina*. México: CEMHAL.
- Handt, O., Krings, M., Ward, R. H., & Pääbo, S. (1996). The retrieval of ancient human DNA sequences. *American journal of human genetics*, 59(2), 368.
- Hernández, C. L., Reales, G., Dugoujon, J. M., Novelletto, A., Rodríguez, J. N., Cuesta, P., & Calderón, R. (2014). Human maternal heritage in Andalusia (Spain): its composition reveals high internal complexity and distinctive influences of mtDNA haplogroups U6 and L in the western and eastern side of region. *BMC genetics*, 15(1), 1-16.
- Hernández, C. L., Soares, P., Dugoujon, J. M., Novelletto, A., Rodríguez, J. N., Rito, T., ... & Calderón, R. (2015). Early holocenic and historic mtDNA African signatures in the Iberian Peninsula: the Andalusian Region as a paradigm. *PLoS One*, 10(10), e0139784.
- Hernández, C. L., Dugoujon, J. M., Novelletto, A., Rodríguez, J. N., Cuesta, P., & Calderón, R. (2017). The distribution of mitochondrial DNA haplogroup H in southern Iberia indicates ancient human genetic exchanges along the western edge of the Mediterranean. *BMC genetics*, 18(1), 1-14.
- Hidalgo, J. (2012). Capítulo 4: Atacama colonial. De la Conquista a la Colonia. En C. Aldunate (Ed.), *Atacama* (pp. 111-155). Museo Chileno de Arte Precolombino; Banco Santander.
- Illumina Sequencing Platform. Human mtDNA Genome. Disponible en: https://emea.support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_legacy/human-mtdna-genome-guide-15037958-01.pdf
- INE Instituto Nacional de Estadísticas (1908). Censo de la República de Chile levantado el 28 de noviembre de 1907. Recuperado de <http://ine.cl/estadisticas/censos/censos-de-poblacion-y-vivienda>
- INE Instituto Nacional de Estadísticas (1931). Resultados del X Censo de la Población efectuado el 27 de noviembre de 1930 y Estadísticas comparativas con censos

- anteriores. Recuperado de <http://ine.cl/estadisticas/censos/censos-de-poblacion-y-vivienda>
- Jobling, M., Hollox, E., Hurles, M., Kivisild, T., Tyler-Smith, C. (2013) *Human Evolutionary Genetics*. New York: Garland Science.
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., y Jermin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods*, 14(6), 587-589.
- Karachanak, S., Carossa, V., Nesheva, D., Olivieri, A., Pala, M., Hooshar Kashani, B., ... & Torroni, A. (2012). Bulgarians vs the other European populations: a mitochondrial DNA perspective. *International journal of legal medicine*, 126(4), 497-503.
- Kassambara, A. (2017). *Practical guide to principal component methods in R: PCA, M (CA), FAMD, MFA, HCPC, factoextra* (Vol. 2). Sthda.
- Katoh, K., Rozewicki, J., y Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in bioinformatics*, 20(4), 1160-1166.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., ... y Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647-1649.
- Kivisild, T. (2015) Maternal ancestry and population history from whole mitochondrial genomes. *Investigative Genetics*, 6:3.
- Kloss-Brandstätter, A., Pacher, D., Schönherr, S., Weissensteiner, H., Binna, R., Specht, G., y Kronenberg, F. (2011). HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Human mutation*, 32(1), 25-32.
- Laricchia, K. M., Lake, N. J., Watts, N. A., Shand, M., Haessly, A., Gauthier, L., ... y Genome Aggregation Database Consortium. (2022). Mitochondrial DNA variation across 56,434 individuals in gnomAD. *Genome research*, 32(3), 569-582.
- Leiva, X. (2010). *Efectos de las migraciones recientes en la composición genética de la población de Santiago de Chile*. (Tesis de Pregrado). Universidad de Chile.
- Legati, A., Zanetti, N., Nasca, A., Peron, C., Lamperti, C., Lamantea, E., y Ghezzi, D. (2021). Current and new next-generation sequencing approaches to study mitochondrial DNA. *The Journal of Molecular Diagnostics*, 23(6), 732-741.
- León, M. (2007). Pasajeros del segundo viaje de Cristóbal Colón. *Revista de Estudios Colombinos*, (3), 19-32.
- León, M. (2014). Visión didáctica de la mujer pobladora en América en el siglo XVI. *Revista de Estudios Colombinos*, (10), 65-74.

- Li, H. y Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25, 1754-60.
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... y Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079.
- Lippold, S., Xu, H., Ko, A., Li, M., Renaud, G., Butthof, A., ... & Stoneking, M. (2014). Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. *Investigative genetics*, 5(1), 1-17.
- López, B. (2013). El viaje a la Nueva España entre 1540 y 1625: El trayecto femenino. En S. Guardia (Ed.), *Historia de las mujeres en América Latina* (pp. 47-56). México: CEMHAL.
- Maldonado, D. (2019). *El cautiverio blanco: nuevas dimensiones de la realidad fronteriza hispano-indígena en el sur de Chile (1540-1810)*. (Informe para optar al grado de Licenciado). Facultad de Filosofía y Humanidades, Universidad de Chile.
- Manríquez, G., Puschel, T., Flores, S., González, T., Moraga, M. y Rothhammer, F. (2016). El origen y evolución de la población chilena desde un enfoque bioantropológico. En F. Falabella, M. Uribe, L. Sanhueza, C. Aldunate, J. Hidalgo (Eds.), *Prehistoria en Chile: Desde sus primeros habitantes hasta los Incas* (pp. 573-586). Santiago, Chile: Editorial Universitaria.
- McEwan, B. (1991). The Archaeology of Women in the Spanish New World. *Historical Archaeology*, 25(4), 33-41.
- Malyarchuk B, Litvinov A, Derenko M, Skonieczna K, Grzybowski T, Grosheva A, Shneider Y, Rychkov S, Zhukova O. (2017). Mitogenomic diversity in Russians and Poles. *Forensic Sci Int Genet.*, 30, 51-56.
- Malyarchuk, B., Derenko, M., Denisova, G., Litvinov, A., Rogalla, U., Skonieczna, K., ... & Molnár, M. J. (2018). Whole mitochondrial genome diversity in two Hungarian populations. *Molecular Genetics and Genomics*, 293(5), 1255-1263.
- Marínez, M. (2020). Migración y colonización. La Agencia de Colonización como actor indispensable en el proyecto de colonización chileno. 1882-1901. *Tzintzun. Revista de Estudios Históricos*, 71, 67-91.
- Martinić, M. (1977). *Historia del Estrecho de Magallanes* (1ra edición). Santiago, Chile: Editorial Andrés Bello.
- Martinić, M. (2002). *Breve historia de Magallanes*. Punta Arenas, Chile: Ediciones de la Universidad de Magallanes.
- Martínez, J. L. (1983). *Pasajeros a Indias. Viajes trasatlánticos en el siglo XVI*. Madrid, España: Alianza Editorial.

- Mellafe, R. (1984). *La introducción de la esclavitud negra en Chile: tráfico y rutas*. Santiago, Chile: Editorial universitaria.
- Moraga, M., Rocco, P., Miquel, J., Nervi, F., Llop, E., Chakraborty, R., ... & Carvalho, P. (2000). Mitochondrial DNA polymorphisms in Chilean aboriginal populations: implications for the peopling of the southern cone of the continent. *American Journal of Physical Anthropology*, 113(1), 19-29.
- Moraga, M., Pezo, P. y de Saint Pierre, M. (2016). El genoma de herencia uniparental en el estudio de las poblaciones fundadoras. In S. Berríos (Ed.), *El ADN de los Chilenos y sus orígenes genéticos* (pp. 62-88). Editorial Universitaria.
- Motti, J., Muzzio, M., Ramallo, V., Rodenak Kladniew, B., Alfaro, E. L., Dipierri, J. E., ... & Bravi, C. M. (2013). Origen y distribución espacial de linajes maternos nativos en el noroeste y centro oeste argentinos. *Revista argentina de antropología biológica*, 15(1), 3-14.
- Nei, M. (1987). *Molecular evolutionary genetics*. Columbia University Press.
- Orellana, M. (2014). *Patrones de diferenciación entre las poblaciones rurales de del Norte y Sur del Curso Superior del río Aconcagua: Aproximación desde la Antropología Molecular* (Tesis de pregrado). Universidad de Concepción, Chile.
- Pakendorf, B. & Stoneking, M. (2005) Mitochondrial DNA and human evolution. *Annu Rev Genomics Hum Genet*, 6, 165–183.
- Pala, M., Olivieri, A., Achilli, A., Accetturo, M., Metspalu, E., Reidla, M., ... & Richards, M. B. (2012). Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. *The American journal of human genetics*, 90(5), 915-924.
- Palencia-Madrid, L., Cardoso, S., Keyser, C., López-Quintana, J. C., Guenaga-Lizasu, A., & de Pancorbo, M. M. (2017). Ancient mitochondrial lineages support the prehistoric maternal root of Basques in Northern Iberian Peninsula. *European Journal of Human Genetics*, 25(5), 631-636.
- Pereira, L., Macaulay, V., Torroni, A., Scozzari, R., Prata, M. J., y Amorim, A. (2001). Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. *Annals of human genetics*, 65(5), 439-458.
- Pereira L., Richards M., Goios A., Alonso A., Albarran C., Garcia O., y col. (2005). High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. *Genome Res.*, 15, 19–24.
- Pezo, P. (2010). *Determinación del aporte indígena materno sobre la población de la comuna de Concepción mediante el análisis de polimorfismos de ADN mitochondrial* (Tesis de pregrado). Universidad de Concepción, Chile.
- Pezo, P., Orellana-Soto, M., de la Fuente, C., Leiva, X., Herrera, L., Flores-Alvarado, S., ... & Moraga, M. (2022). Native American mitochondrial lineages in admixed populations from Chile: Detecting recent migrations during post-Columbian times

- using geographically restricted lineages. *American Journal of Biological Anthropology*, 178(3), 504-512.
- Pierron, D., Heiske, M., Razafindrazaka, H., Rakoto, I., Rabetokotany, N., Ravololomanga, B., ... & Letellier, T. (2017). Genomic landscape of human diversity across Madagascar. *Proceedings of the National Academy of Sciences*, 114(32), E6498-E6506.
- Piotrowska-Nowak, A., Kosior-Jarecka, E., Schab, A., Wrobel-Dudzinska, D., Bartnik, E., Zarnowski, T., & Tonska, K. (2019). Investigation of whole mitochondrial genome variation in normal tension glaucoma. *Experimental Eye Research*, 178, 186-197.
- Podgorná, E., Soares, P., Pereira, L., & Černý, V. (2013). The genetic impact of the lake chad basin population in North Africa as documented by mitochondrial diversity and internal variation of the L3e5 haplogroup. *Annals of Human Genetics*, 77(6), 513-523.
- Powers, K. (2005). *Women in the Crucible of Conquest: the gendered genesis of Spanish American society, 1500-1600*. University of New Mexico Press.
- Promega Corporation (2017-2018). *TECHNICAL MANUAL ProNex® Size-Selective Purification System*. Madison, USA. Disponible en: <https://worldwide.promega.com/products/sequencing/ngs-library-prep/pronex-size-selective-purification-system?catNum=NG2001>
- Quintana-Murci, L., Quach, H., Harmant, C., Luca, F., Massonnet, B., Patin, E., ... y Behar, D. M. (2008). Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. *Proceedings of the National Academy of Sciences*, 105(5), 1596-1601.
- Rambaut, A. (2018). *FigTree version 1.4.4*. Disponible online en: <http://tree.bio.ed.ac.uk/software/figtree/>
- R Core Team. (2018). *R: A language and environment for statistical computing (computer software)*. R Foundation for Statistical Computing.
- Rito, T., Richards, M.B., Fernandes, V., Alshamali, F., Cerny, V., Pereira, L. & Soares, P. (2013). The first modern human dispersals across Africa. *PLoS One*, 8(11); e80031.
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D., Darling, A., Höhna, S., Larget, B., ... y Huelsenbeck, J.P. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst. Biol.*, 61, 539–542.
- Roostalu, U., Kutuev, I., Loogväli, E., Metspalu, E., Tambets, K., Reidla, M., ... y Villems, R. (2007). Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: the Near Eastern and Caucasian perspective. *Molecular biology and evolution*, 24(2), 436-448.

- Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular biology and evolution*, *34*(12), 3299-3302.
- Ruiz-Linares, A. (2015). How genes have illuminated the history of early Americans and Latino Americans. *Cold Spring Harbor Perspectives in Biology*, *7*(6), a008557.
- Schaan, A., Costa, L., Santos, D., Modesto, A., Amador, M., Lopes, C., ... y Yoshioka, F. (2017). mtDNA structure: the women who formed the Brazilian Northeast. *BMC evolutionary biology*, *17*(1), 185.
- Silva, M., Alshamali, F., Silva, P., Carrilho, C., Mandlate, F., Jesus Trovoada, M., Cerny, V., Pereira, L. & Soares, P. (2015). 60,000 years of interactions between Central and Eastern Africa documented by major African mitochondrial haplogroup L2. *Sci Rep*, *5*; 12526.
- Silva, P. (2021). *Aporte de linajes mitocondriales indígenas a la actual población de Curicó* (Tesis de pregrado). Universidad de Chile.
- Simão, F., Ferreira, A. P., de Carvalho, E. F., Parson, W., & Gusmão, L. (2018). Defining mtDNA origins and population stratification in Rio de Janeiro. *Forensic Science International: Genetics*, *34*, 97-104.
- Simão, F., Ribeiro, J., Vullo, C., Catelli, L., Gomes, V., Xavier, C., ... & Gusmão, L. (2021). The ancestry of Eastern Paraguay: a typical South American profile with a unique pattern of admixture. *Genes*, *12*(11), 1788.
- Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., ... & Richards, M. B. (2009). Correcting for purifying selection: an improved human mitochondrial molecular clock. *The American Journal of Human Genetics*, *84*(6), 740-759.
- Soares, P., Alshamali, F., Pereira, J.B., Fernandes, V., Silva, N.M., Afonso, C., Costa, M.D., Musilova, E., Macaulay, V., Richards, M.B., Cerny, V. & Pereira, L. (2012). The Expansion of mtDNA Haplogroup L3 within and out of Africa. *Mol Biol Evol*, *29*(3), 915-927.
- Soto, R. (1992). Negras esclavas. Las otras mujeres de la Colonia. *Proposiciones*, *21*, 21-31.
- Stefflova, K., Dulik, M. C., Barnholtz-Sloan, J. S., Pai, A. A., Walker, A. H., & Rebbeck, T. R. (2011). Dissecting the within-Africa ancestry of populations of African descent in the Americas. *PloS one*, *6*(1), e14495.
- Sullivan-Beare, N. (1956). *Las mujeres de los conquistadores: La mujer española en los comienzos de la colonización americana*. Compañía Bibliográfica Española.
- Swofford, D. (2003). *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Version 4. Sinauer Associates, Sunderland, Massachusetts.
- Tajima, F. (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* *105*, 437-460.

- Tamm, E., Kivisild, T., Reidla, M., Metspalu, M., Smith, D., Mulligan, C., ... y Fedorova, S. (2007). Beringian standstill and spread of Native American founders. *PloS One*, 2(9), e829.
- Tamura, K., Stecher, G., y Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Molecular biology and evolution*, 38(7), 3022-3027.
- Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on mathematics in the life sciences*, 17(2), 57-86.
- Téllez, E. (1984). La guerra atacameña del siglo XVI: Implicancias y trascendencia de un siglo de insurrecciones indígenas en el despoblado de Atacama. *Estudios Atacameños*, (7), 295-310.
- The Galaxy Community (2022). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Research*, 50(W1), W345-W351.
- Tillería, J. y Vela, F. (2017). Las viviendas de la colonización alemana en el sur de Chile: su influencia en la arquitectura tradicional de Valdivia a Puerto Montt. *Cuaderno de Notas*, (18), 54-72.
- Torrioni, A., Schurr, T., Yang, C-C., Szathmary, E., Williams, R., Schanfield, M., Troup, G., Knowler, W., Lawrence, D., Weiss, K., Wallace, D. (1992). Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics* 130, 153–162.
- Trifinopoulos, J., Nguyen, L. T., von Haeseler, A., y Minh, B. Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic acids research*, 44(W1), W232-W235.
- van Oven, M., Kayser, M. (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat*, 30, 386–394.
- Verdugo, R., Di Genova, A., Herrera, L., Moraga, M., Acuña, M., Berríos, S., ... y Symon, A. (2020). Development of a small panel of SNPs to infer ancestry in Chileans that distinguishes Aymara and Mapuche components. *Biological research*, 53, 1-11.
- Vergara, S. (1981). Edad y vida en el grupo conquistador. un estudio de la existencia humana en el siglo XVI. *Cuadernos de Historia* 1, 65-86.
- Vial, G. (1957). *El africano en el Reino de Chile: Ensayo histórico-jurídico*. Instituto de Investigaciones Históricas, Universidad Católica de Chile.
- Vieira-Machado, C., Tostes, M., Alves, G., Nazer, J., Martinez, L., Wettig, E., ... y Dutra, M. (2016). Uniparental ancestry markers in Chilean populations. *Genetics and molecular biology*, 39(4), 573-579.
- Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H. J., ... y Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic acids research*, 44(W1), W58-W63.

- Yang, I. S., Lee, H. Y., Yang, W. I., y Shin, K. J. (2013). mtDNAprofiler: a Web application for the nomenclature and comparison of human mitochondrial DNA sequences. *Journal of forensic sciences*, 58(4), 972–980.
- Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution*, 39(3), 306-314.
- Zhang W, Cui H, Wong L. (2012). Comprehensive one-step molecular analyses of mitochondrial genome by massively parallel sequencing. *Clin Chem*, 58, 1322-1331
- Zlatař, V. (2002). Inmigración croata en Antofagasta. Oñate Impresores.

7. ANEXOS

Tabla S1. Protocolos de amplificación de fragmentos largos para la secuenciación de mitogenomas.

Protocolo	Partidor	Secuencia (5' - 3')	Posición	Tamaño fragmento	Referencia
1	H2698	ttgacctgccccgtgaagagg	2679-2698	5721 pb	Laboratorio Uchile
	L8351	ttggggcatttactgtaaagagg	8351-8374		
	H7801	ctatcctgcccccatcatc	7782-7801	6530 pb	
	L14267	gaggggtcagggttgattcg	14267-14286		
	H13305	tggcatcaaccaaccacac	13286-13305	6735 pb	
	L3407	tacaacgttggggcctttgc	3407-3426		
2	FampA	aaatcttaccgccctgttt	2480-2499	8378 pb	Gonder y col. 2007
	RampA	aattaggctgtgggtggttg	10839-10858		
	FampB	gccatactagtctttgccgc	10653-10672	8604 pb	
	RampB	ggcaggtaatttactggt	2669-2688		
3	MTL-F1	aaagcacataccaaggccac	9397-9416	9065 pb	Illumina Sequencing Platform
	MTL-F2	tatccgcatcccatatt	15195-15214		
	MTL-R1	ttggctctccttgcaaagtt	1873-1892	11170 pb	
	MTL-R2	aatgttgagccgtagatgcc	9777-9796		

Tabla S2: Matriz de FST entre pares de poblaciones chilenas obtenida a partir de las frecuencias de haplotipos estimados por DNAsp.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Punta Arenas		0.05405	0.06306	0.00000	0.00000	0.00000	0.34234	0.00000	0.00000	0.00901	0.03604	0.00000	0.04505	0.00000	0.02703	0.00000	0.00000
2. Puerto Montt	0.04535		0.61261	0.11712	0.15315	0.08108	0.68468	0.01802	0.00901	0.00901	0.02703	0.00000	0.00901	0.00000	0.02703	0.03604	0.00901
3. Osorno	0.05582	-0.02548		0.31532	0.39640	0.14414	0.37838	0.06306	0.13514	0.17117	0.15315	0.10811	0.19820	0.16216	0.34234	0.20721	0.13514
4. Temuco	0.03053	0.01496	0.01016		0.38739	0.00901	0.77477	0.01802	0.10811	0.18018	0.11712	0.00000	0.12613	0.01802	0.13514	0.23423	0.00000
5. Concepción	0.04513	0.01665	0.00676	0.00033		0.02703	0.48649	0.00901	0.01802	0.21622	0.15315	0.00000	0.10811	0.00901	0.21622	0.04505	0.00000
6. Chillan	0.07225	0.04186	0.03093	0.01619	0.02688		0.72973	0.34234	0.00000	0.17117	0.36937	0.01802	0.14414	0.04505	0.14414	0.05405	0.00000
7. San Carlos	0.00439	-0.02773	-0.00540	-0.02063	-0.00661	-0.01897		0.38739	0.25225	0.49550	0.14414	0.00901	0.20721	0.17117	0.09910	0.36036	0.12613
8. Curicó	0.17577	0.11559	0.13911	0.09649	0.14330	0.01014	0.00502		0.01802	0.04505	0.25225	0.00901	0.03604	0.04505	0.09009	0.05405	0.06306
9. Santiago	0.03037	0.04428	0.02150	0.00385	0.01455	0.02775	0.00741	0.11027		0.57658	0.32432	0.02703	0.45946	0.02703	0.17117	0.18919	0.00000
10. SF-LA	0.05736	0.04799	0.03452	0.00884	0.01090	0.01232	-0.00381	0.10001	-0.00393		0.63964	0.03604	0.79279	0.24324	0.32432	0.72072	0.05405
11. Aconcagua	0.08299	0.08175	0.05882	0.01931	0.03000	0.00724	0.03553	0.11305	0.00241	-0.02409		0.17117	0.93694	0.28829	0.52252	0.58559	0.09910
12. Coquimbo	0.13460	0.12212	0.09465	0.09334	0.10709	0.09795	0.10080	0.19838	0.07334	0.06457	0.08834		0.13514	0.27027	0.29730	0.03604	0.00901
13. La Serena	0.04801	0.06149	0.02143	0.01631	0.02111	0.02889	0.02221	0.10784	-0.00340	-0.02715	-0.04903	0.05202		0.74775	0.89189	0.90090	0.14414
14. Antofagasta	0.06223	0.04995	0.02411	0.03219	0.04572	0.04971	0.01754	0.09674	0.02514	0.01126	0.00438	0.01784	-0.01670		0.72973	0.58559	0.36937
15. Calama	0.09278	0.08139	0.00337	0.03343	0.03310	0.03653	0.04272	0.13924	0.02020	0.00818	-0.00576	0.02869	-0.03911	-0.03494		0.50450	0.28829
16. Iquique	0.04761	0.03644	0.01071	0.00579	0.02094	0.02389	0.00402	0.08493	0.00450	-0.00783	-0.01512	0.05838	-0.02006	-0.00317	-0.00595		0.27928
17. Arica	0.09360	0.06548	0.04010	0.04525	0.07498	0.04862	0.01764	0.06661	0.05122	0.02887	0.03172	0.06584	0.01643	0.00614	0.00872	0.00188	

Tabla S3: Matriz de FST entre pares de poblaciones chilenas obtenido a partir de las frecuencias de 27 haplogrupos.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Arica		0.66696	0.35204	0.31759	0.12692	0.06861	0.01574	0.17048	0.00703	0.02198	0.05039	0.00000	0.00119	0.00238	0.13355	0.00030	0.00000
Iquique	-0.00685		0.18038	0.36353	0.34888	0.02861	0.13187	0.59459	0.23562	0.03366	0.27324	0.01297	0.12078	0.16939	0.56202	0.03485	0.00168
Antofagasta	0.00420	0.01299		0.94555	0.44095	0.81150	0.15890	0.12019	0.00733	0.02792	0.18959	0.00020	0.00059	0.00178	0.20780	0.00188	0.00812
Calama	0.01501	0.00860	-0.06088		0.96297	0.50084	0.58776	0.30066	0.11910	0.03079	0.10296	0.08316	0.12038	0.08702	0.65003	0.03317	0.02980
La Serena	0.02339	0.00469	0.00195	-0.06110		0.19493	0.43877	0.31136	0.03960	0.01733	0.03643	0.00426	0.02851	0.00683	0.43798	0.00505	0.01040
Coquimbo	0.04027	0.04869	-0.03343	-0.02031	0.03122		0.18671	0.13316	0.00297	0.01386	0.03960	0.00059	0.00040	0.00069	0.08069	0.00000	0.00040
Aconcagua	0.07700	0.03604	0.04072	-0.01486	-0.00164	0.05322		0.16177	0.01317	0.10633	0.02396	0.00960	0.00475	0.00396	0.17969	0.00099	0.00218
SF - LA	0.01497	-0.00553	0.02712	0.02059	0.01087	0.03428	0.03727		0.03614	0.02267	0.28334	0.04920	0.02287	0.02010	0.22354	0.01317	0.00099
Santiago	0.02476	0.00466	0.03323	0.03102	0.02899	0.05944	0.06762	0.02154		0.05960	0.31046	0.00020	0.00802	0.05742	0.19404	0.02564	0.00069
Curicó	0.09817	0.07459	0.11104	0.13022	0.12744	0.17868	0.14605	0.09935	0.05258		0.33363	0.18810	0.02218	0.06504	0.02465	0.07306	0.09128
San Carlos	0.03908	0.01205	0.02930	0.06072	0.05730	0.08338	0.10575	0.01835	0.00724	0.03440		0.55589	0.49451	0.82160	0.62677	0.82546	0.54401
Chillán	0.05799	0.02675	0.06715	0.04904	0.06304	0.10419	0.08756	0.02625	0.02580	0.03388	-0.00754		0.02802	0.02723	0.31106	0.04415	0.00000
concepción	0.04596	0.01201	0.06275	0.03970	0.04323	0.10563	0.10355	0.03401	0.01596	0.09440	-0.00414	0.01845		0.71379	0.75101	0.19741	0.00000
Temuco	0.03273	0.00741	0.04651	0.04255	0.04961	0.09318	0.09211	0.02895	0.00607	0.05707	-0.02029	0.01284	-0.00427		0.56400	0.37323	0.00584
Osorno	0.03053	-0.00358	0.02971	-0.01564	0.00326	0.07357	0.06159	0.02648	0.01704	0.11779	-0.01257	0.01066	-0.02401	-0.00721		0.83348	0.11583
Puerto Montt	0.05440	0.02862	0.06290	0.06667	0.06130	0.11279	0.12079	0.04613	0.02550	0.07527	-0.03180	0.02975	0.01267	0.00232	-0.03454		0.17186
Punta Arenas	0.07389	0.04790	0.05169	0.07428	0.06400	0.11036	0.11665	0.07167	0.02710	0.06549	-0.00820	0.06038	0.05569	0.02316	0.04052	0.01466	

Tabla S4: Poblaciones africanas de referencia utilizadas para los análisis comparativos.

Población	Región	N	Referencia bibliográfica
Angola	West central Africa	640	Coelho y col. (2009); Barbieri y col. (2014)
Burkina Faso	West Africa	293	Barbieri y col. (2012)
Camerún	Bight of Biafra	289	Cerezo y col. (2016)
Chad	Central Africa	146	Cerezo y col. (2016)
Congo	West central Africa	29	Hernández y col. (2015)
RDCongo	West central Africa	35	Batini y col. (2011)
Egipto	North Africa	19	Coudray y col. (2009)
Etiopia	East Africa	21	Behar y col. (2008)
Ghana	Gold Coast	192	Fendt y col. (2012)
Costa de Marfil	Windward Coast	100	Göbel y col. (2020)
Marruecos	North Africa	215	Cerezo y col. (2016); Aboukhalid y col. (2013)
Mozambique	Southeast Africa	511	Cerezo y col. (2016)
Niger	West Africa	52	Cerezo y col. (2016)
Nigeria	Bight of Benin	92	Cerezo y col. (2016)
Sierra Leona	Sierra Leone	80	Hernández y col. (2015)
Senegal	Senegambia	194	Stefflova y col. (2011)
Tanzania	East Africa	301	Cerezo y col. (2016)
Togo	Bight of Biafra	54	Göbel y col. (2020); Cerezo y col. (2016)
Uganda	East Africa	211	Cerezo y col. (2016)
Zambia	Southeast Africa	629	Barbieri y col. (2014)
Zimbabue	Southeast Africa	21	Cerezo y col. (2016)
Gabón	West central Africa	203	Quintana-Murci y col. (2008); Batini y col. (2011)
Gambia	Senegambia	112	Hernández y col. (2015)
Guinea-Bisáu	Senegambia	301	Stefflova y col. (2011)
Namibia	Southwest Africa	106	Barbieri y col. (2014)
Libia	North Africa	85	Fadhlaoui-Zid y col. 2011
Madagascar	Southeast Africa	1023	Pierron y col. (2017)
Benin	Bight of Benin	150	Brucato y col. (2010)
Mali	West Africa	201	Stefflova y col. (2011)
Andalucía	Iberia	243	Hernández y col. (2014)

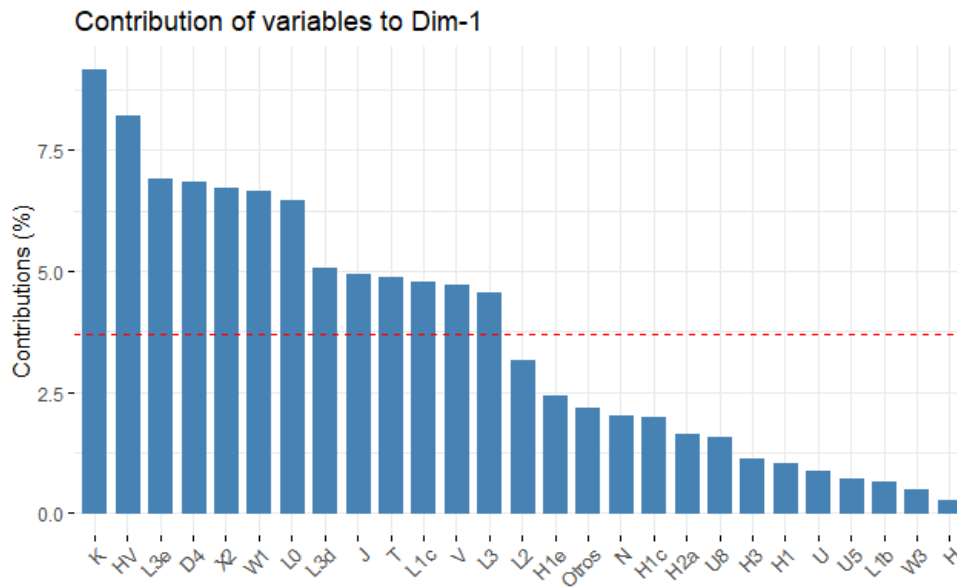
Tabla S5: Poblaciones europeas de referencia utilizadas en la red de haplotipos de HV.

Población	N	Referencia
Alemania	21	Brandstätter y col. (2006)
Andalucía	7	Hernández y col. (2014)
Croacia	1	Family Tree DNA
Hungría	1	Malyarchuk y col. (2018)
Inglaterra	13	Family Tree DNA
Italia	20	Brisighelli y col. (2012)
País Vasco	5	Palencia-Madrid y Pancorbo (2018), sin publicar
España	53	Cerezo y col. (2016)
Polonia	1	Piotrowska-Nowak y col. (2019)
Rusia	1	Malyarchuk y col. (2017)
Suiza	1	Family Tree DNA

Tabla S6: Poblaciones africanas de referencia utilizadas en la red de haplotipos de L.

Población	N	Referencia
Angola	46	Barbieri y col. (2014)
Benin	1	Göbel y col. (2020)
Burkina Faso	1	Barbieri y col. (2012)
Camerún	30	Gonder y col. (2007); Podgorná y col. (2013); Göbel y col. (2020)
Chad	1	Soares y col. (2012)
Congo	16	Batini y col. (2011)
Costa de Marfil	15	Göbel y col. (2020)
Etiopía	1	Rito y col. (2013)
Gabón	26	Batini y col. (2011)
Gambia	3	Göbel y col. (2020)
Ghana	52	Fendt y col. (2012); Göbel y col. (2020)
Guinea-Bisáu	10	Behar y col. (2008)
Madagascar	17	Pierron y col. (2017)
Marruecos	6	Aboukhalid y col. (2013); Hernández y col. (2015)
Mozambique	40	Cerezo y col. (2016)
Nigeria	10	Göbel y col. (2020)
Rep. Central África	1	Zheng y col. (2014), sin publicar
Senegal	11	Lippold y col. (2014)
Togo	9	Göbel y col. (2020)
Zambia	1	Barbieri y col. (2012)

a)



b)

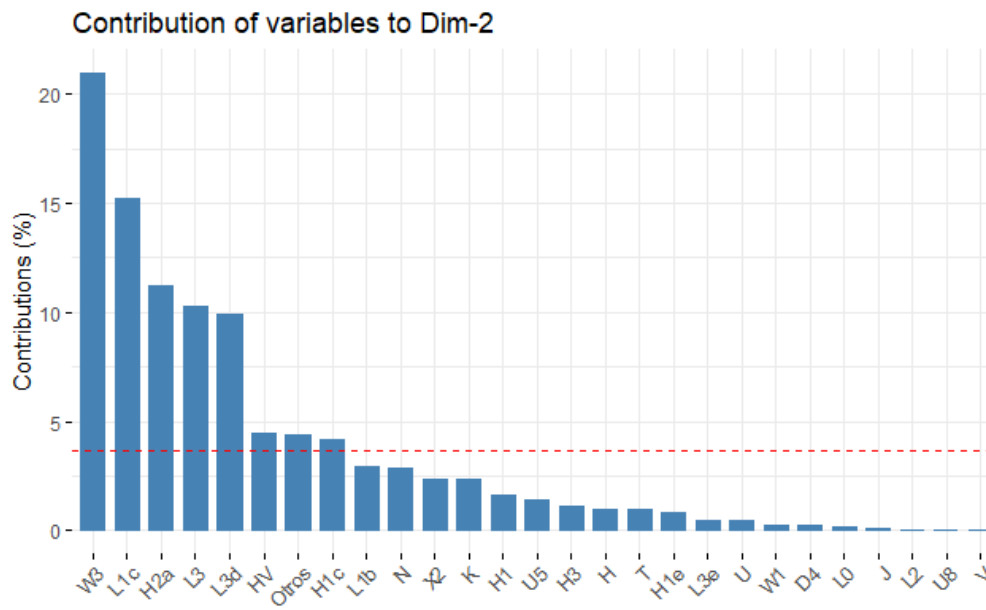
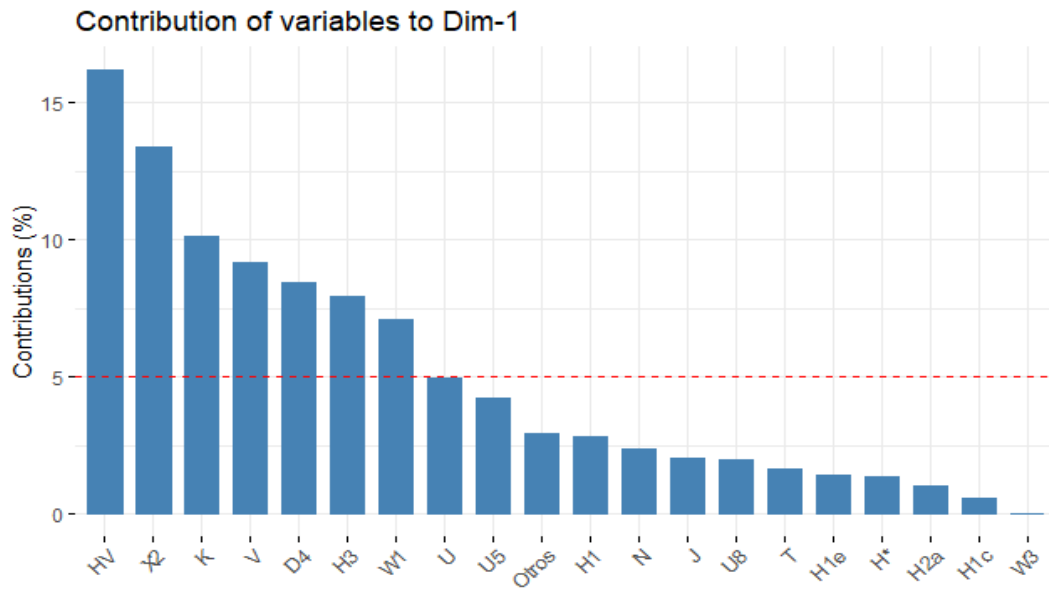


Figura S1: Gráfico de barras con la contribución de cada variable al componente 1 (a) y 2 (b). PCA realizado a partir de las frecuencias de 27 haplogrupos.

a)



b)

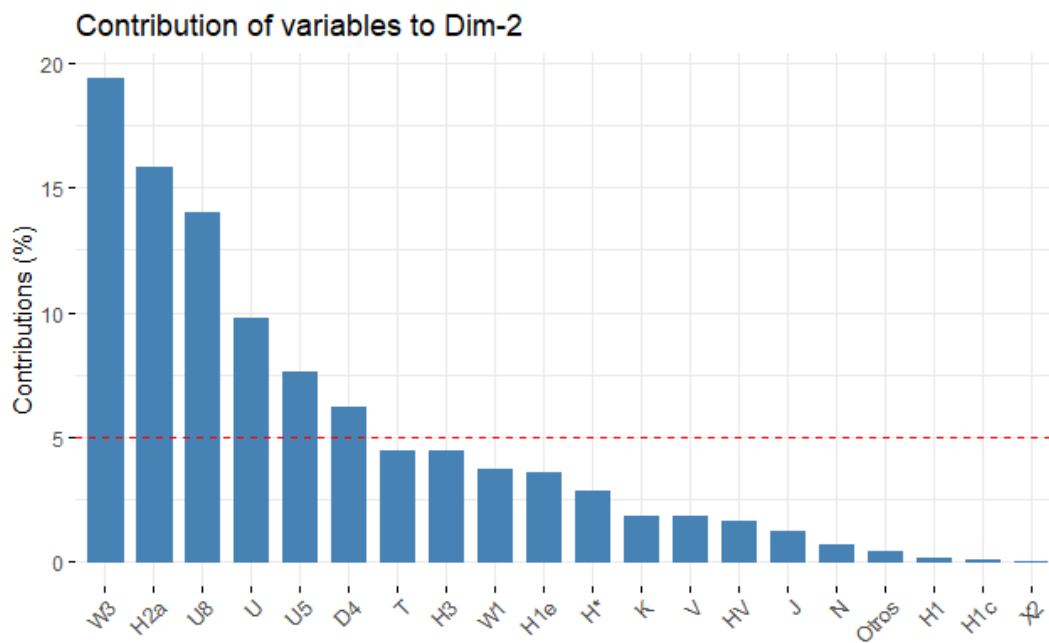
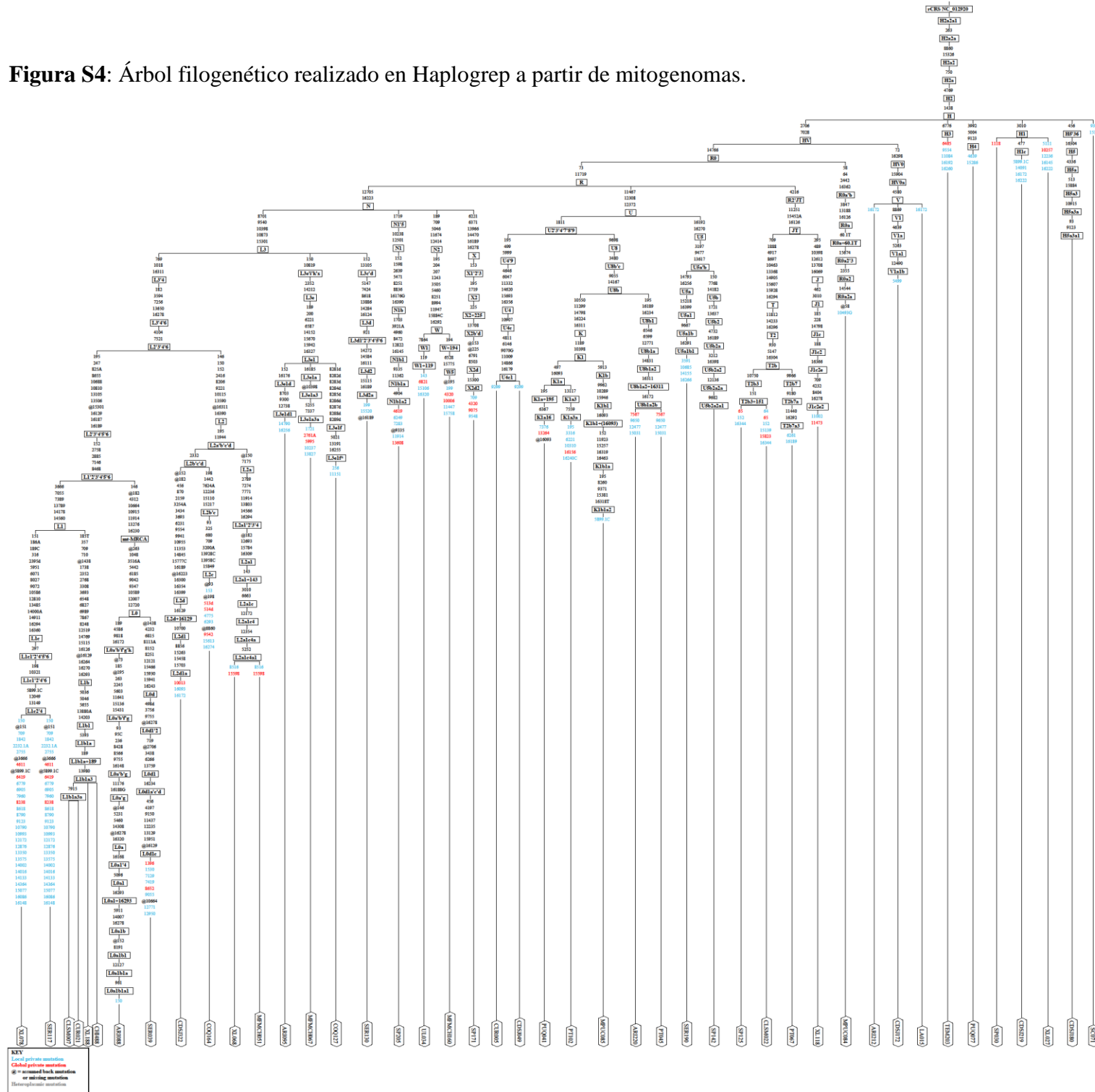


Figura S2: Gráfico de barras con la contribución de cada variable al componente 1 (a) y 2 (b). PCA realizado a partir de las frecuencias de 20 haplogrupos (sin haplogrupo L).



Figura S3: Análisis de Componentes Principales para poblaciones chilenas y españolas. Se utilizó la división regional de Barral-Arca y col. (2016).

Figura S4: Árbol filogenético realizado en Haplogrep a partir de mitogenomas.



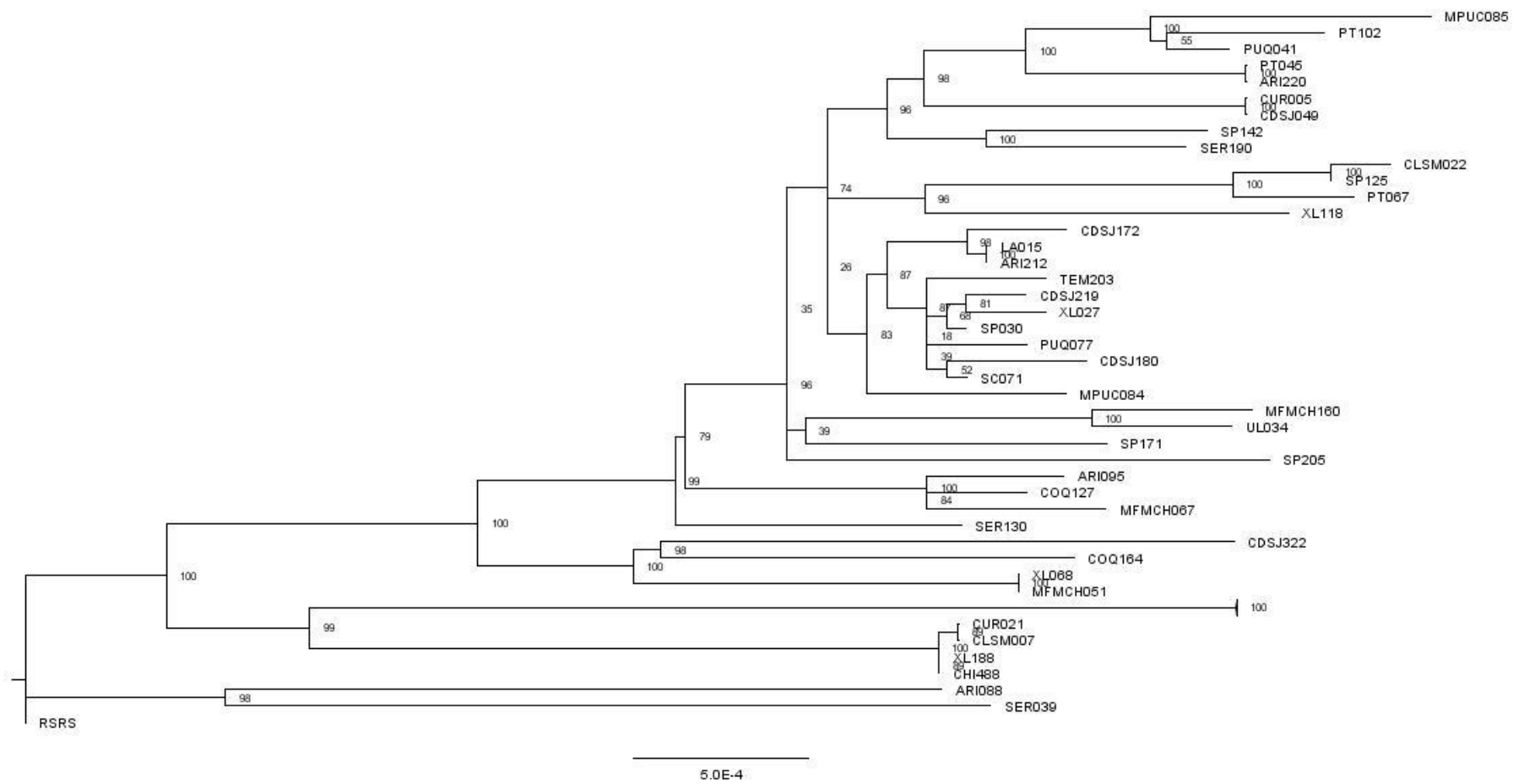


Figura S6: Árbol filogenético de Máximo Likelihood para los linajes mitocondriales no amerindios en Chile, a partir de mitogenomas. Se indican los valores de bootstrap.

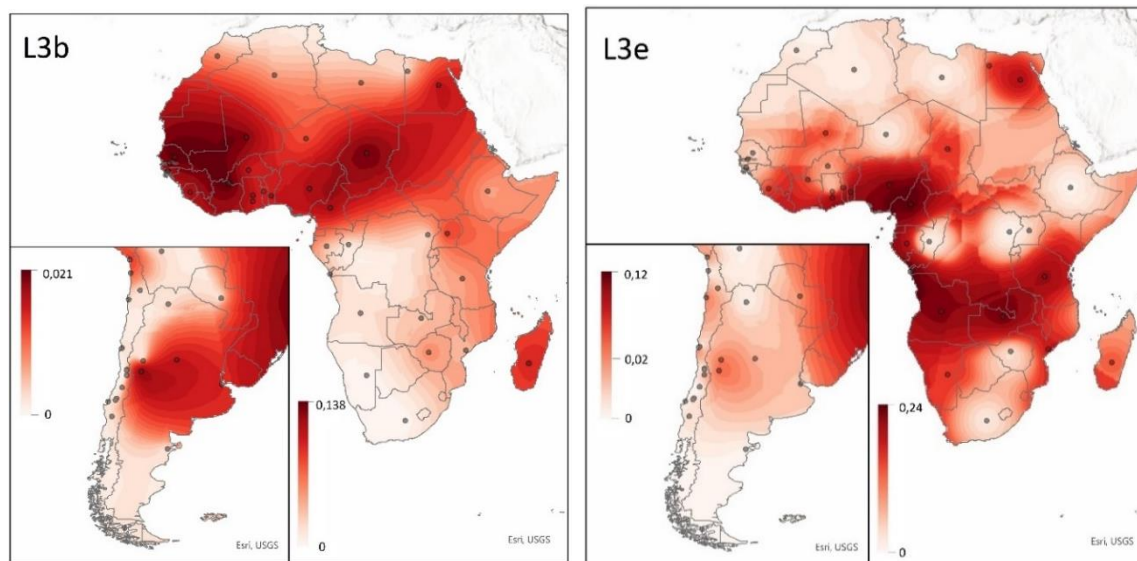
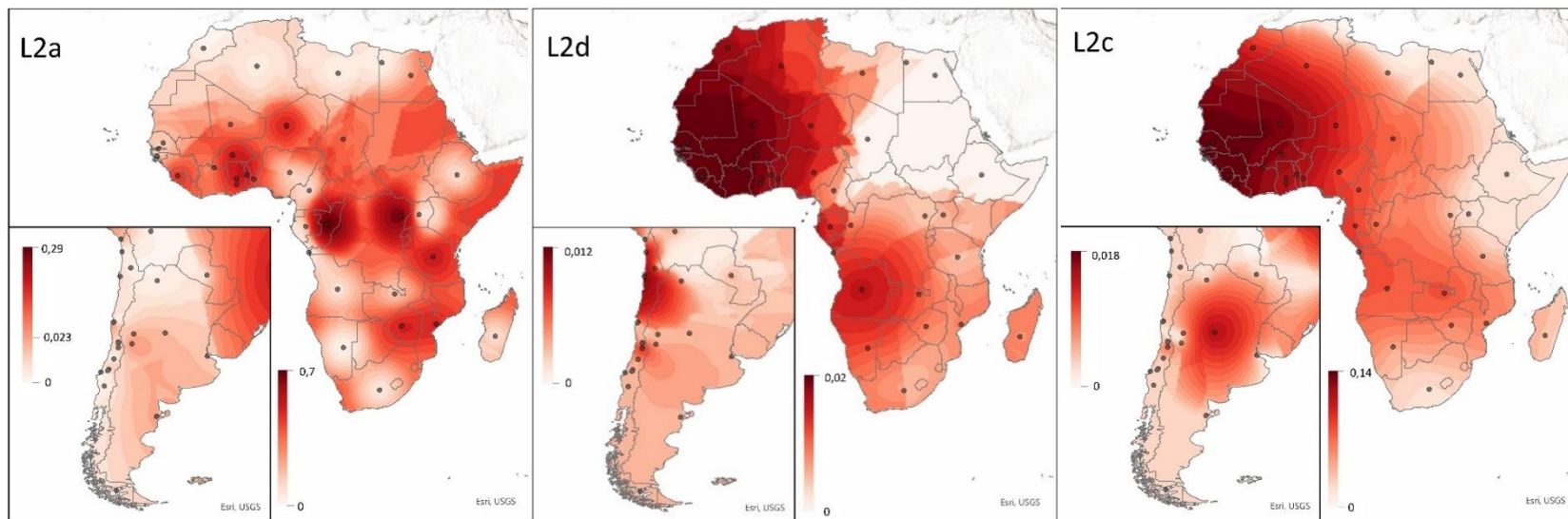


Figura S7: Mapas de isofrecuencias por método kriging para 5 linajes africanos. La simbología del nivel de frecuencia debe considerarse de manera separada.

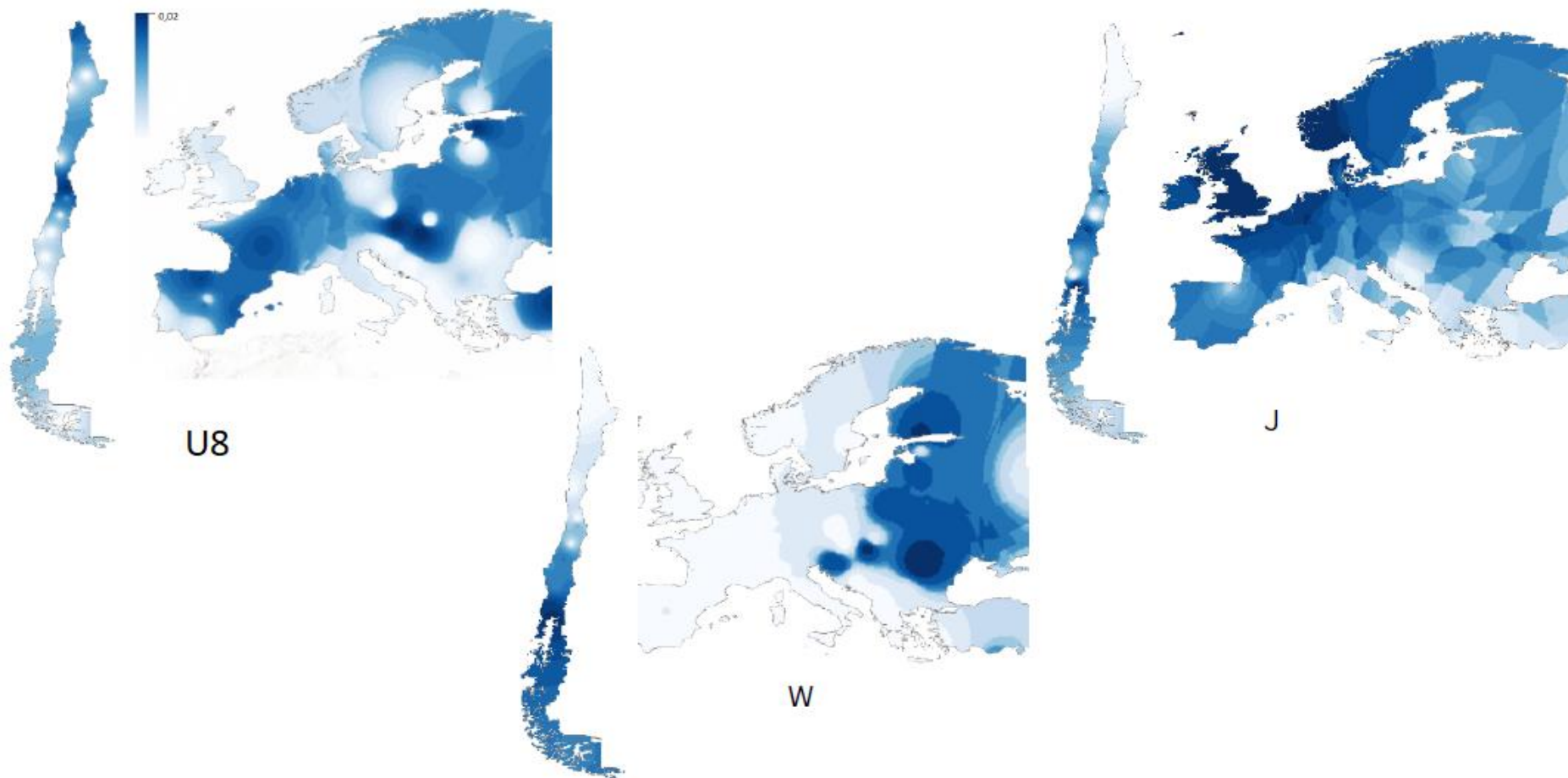


Figura S8: Mapas de isofrecuencias por método *kriging* para 3 linajes europeos. La simbología del nivel de frecuencia debe considerarse de manera separada.