



Universidad de Chile
Facultad de Ciencias Sociales
Departamento de Antropología

Selección natural en variantes genéticas de la proteína surfactante D en población altiplánica

Memoria para optar al título de
Antropóloga física

Fernanda Andrea Herrera Mesías
Profesor guía: Dr. Mauricio Moraga

2017

“逃げたりあきらめるコトは 誰も
一瞬あればできるから 歩く続けよう

君にしかできないコトがある 青い星に
光がなくせぬように

つかめ！ 描いた夢を
まもれ！ 大事な友を
たくましい自分になれるさ
知らないパワーが宿る ハートに火がついたら
どんな願いも 嘘じゃない
きっとかなうから。。。 Show me your brave heart”
-デジモン.,(1999)

Agradecimientos

Este trabajo fue realizado bajo el marco del proyecto Fondecyt 1140544 “Señales recientes de adaptación genética a climas extremos en poblaciones de cazadores recolectores marítimos de Patagonia y Tierra del Fuego”. Se agradece la colaboración del Dr. A. Moreno-Estrada (Stanford University), quien facilitó el acceso a las muestras de Puno. También se agradece el trabajo de todos los investigadores que se dedican al desarrollo de herramientas informáticas de libre acceso destinadas al análisis genético. Sin su generoso aporte, este y muchos otros trabajos no serían posibles.

El camino para llegar aquí estuvo lleno de desafíos, pero tengo el honor de decir que nunca lo recorrí sola. Agradezco al Profesor M. Moraga por darme esta oportunidad, por acogerme como una de sus Padawan, por sus incontables consejos de todo tipo, por respetar mis “particularidades” y por siempre creer en que todo saldría bien al final. Un agradecimiento especial a M. Apata, quien se dio el tiempo de orientarme de principio a fin y jugar el rol de tutor extraoficial. Lo quieras o no, siempre seré tu primera tesista.

A N. Montalva, S. Krapivka, M. de Saint Pierre y D. Véliz, por todos los consejos y comentarios que hicieron para que este trabajo fuese todo lo bueno que podía llegar a ser. Sus enseñanzas me motivan a querer llegar tan lejos como pueda.

Agradezco a todos los miembros del Laboratorio de Genética de Poblaciones y Evolución Humana del ICBM, por su constante apoyo y por tenerme paciencia durante todo este tiempo. Han sido como una familia para mí. A M. Orellana, S. Flores y P. Contreras por su importantísima ayuda en todo lo que fue la parte metodológica, así como por su buena onda y por dar siempre lo mejor de sí mismos, algún día me gustaría ser tan cool como ustedes! A mi inseparable compañera A. Miranda Cortés, por su constante cariño y preocupación incluso en los momentos más duros. Sabes que te adoro!

A mis ahijadas de la Universidad, A. Eyquem y M. Faure, por su amistad sincera demostrada a lo largo de los años, y porque si no hubiese tomado el ramo de Análisis Genético sólo para pasar más tiempo juntas mi futuro habría sido muy distinto.

A Y. Valenzuela por toda su amabilidad y por cubrirme para que pudiera terminar este trabajo de la mejor manera posible.

Agradecimientos infinitos a mis padres por alentarme siempre a alcanzar nuevas alturas, por protegerme desde que nací y por darme su apoyo incondicional en todo lo que he querido hacer (incluyendo estudiar Antropología). Por darme con su sudor y sangre lo más útil que puede tener un adulto joven chileno de esta generación: una vida sin deudas y la posibilidad de moldear mi destino a mi antojo. No dejaré que su sacrificio sea en vano.

Finalmente, agradezco a mi camarada de la juventud, amigo, confidente y actualmente esposo M. Gárate, quien hace años me dedicó su tesis y que hoy en día sigue trabajando duramente por nuestro futuro. Mi querido Matcha, Gemeinsam wir sind stark!

Tabla de contenido

Resumen.....	8
1. Antecedentes.....	9
1.a Adaptación humana a ambientes extremos.....	9
1.a.1 Adaptación y selección natural en poblaciones de altura.....	9
1. a.2 El altiplano: características biogeográficas.....	11
1. a. 3 Poblaciones altiplánicas.....	12
1.a.4 Microevolución en población altiplánica y adaptación a la altura.....	14
1.b Surfactante pulmonar.....	17
1.b.1 Importancia adaptativa del surfactante pulmonar en el altiplano.....	17
1. b. 2 Señales de selección y polimorfismos en el gen SFTPD.....	18
2. Hipótesis:.....	21
3. Objetivos.....	21
3.1 Objetivo general:.....	21
3.2 Objetivos específicos:.....	21
4. Materiales y métodos.....	22
4.a Poblaciones y Muestras.....	22
4.b Genotipificación.....	23
4.b.1 Base de datos y archivos digitales.....	23
4.b.2 Amplificación del ADN (PCR).....	25
4.b.3 Secuenciación.....	26
4.c Análisis genéticos y estadísticos.....	26
4.c.1 Análisis y Programas.....	26
4.c.2 Desequilibrio de ligamiento y haplotipos.....	26
4.c.3 Fst y Neighbour-Joining.....	27
4.c.4 Análisis de Componentes Principales (PCA).....	28
5. Resultados.....	29
5.a Equilibrio de Hardy-Weinberg.....	29
5. b Frecuencias alélicas.....	31
5.b.1 Frecuencias SNPs exónicos.....	31
5.b.2 Diferencias significativas: Prueba exacta de Fisher.....	32
5. c Frecuencias genotípicas.....	34

5.d Desequilibrio de ligamiento	36
5.d.1 Haplotipos SNPs exónicos.....	36
5. d. 2 Bloques haplotípicos	37
5.e Valores de Fst y Neighbour-Joining.....	45
5.f Análisis de Componentes Principales	51
5.f.1 PCA : Frecuencias alélicas	51
5. f.2 PCA: Frecuencias haplotípicas	54
6. Discusión	59
6. 1 <i>SFTPD</i> y el patrón adaptativo de las poblaciones altiplánicas	59
6.2 Análisis genómicos y métodos para detectar selección natural	61
6.3 Diversidad haplotípica y diferenciación genética entre poblaciones.....	64
7. Conclusiones.....	70
8. Referencias	71

Índice de figuras

Figura 1: Disminución de la presión parcial de oxígeno inspirado	12
Figura 2: Posición del gen <i>SFTPD</i> en el cromosoma 10	24
Figura 3: Resumen de las metodologías utilizadas	29
Figura 4: Frecuencias alélicas de los SNPs exónicos.....	31
Figuras 5(A, B, C, D, E, F, G): Gráficos de las frecuencias alélicas de SNPs con diferencias significativas entre Puno y SUR.....	34
Figuras 6.a – 6.k: LD-plots	45
Figura 7.a: Valores de F_{st} entre poblaciones (SNPs <i>SFTPD</i>).....	47
Figura 7.b: Valores de F_{st} entre poblaciones (región neutral).....	47
Figura 8: Dendrograma de Neighbor-Joining	48
Figura 9: Gráfico de F_{st} vs Heterocigosidad (general)	50
Figura 10.a-10.b: Gráfico de F_{st} vs Heterocigosidad (pares de poblaciones).....	51
Figura 11: Análisis de Componentes Principales, Biplot PC1/PC2 (SNPs)	52
Figura 12: Análisis de Componentes Principales, Biplot PC1/PC2 (Haplotípos)	56
Figura 13: Análisis de Componentes Principales, Biplot PC1/PC3 (Haplotípos)	58
Figura 14: Desequilibrio de ligamiento en diferentes poblaciones	63

Índice de tablas:

Tabla 1: Frecuencias alélicas en población amerindia habitante de altura y de tierras bajas	19
Tabla 2: Concentración SP-D en sangre según genotipo (ng/mL)	20
Tabla 3: Condiciones de amplificación para el SNP rs2243639.....	25
Tabla 4: Protocolo de PCR	25
Tabla 5: SNPs zona neutral cromosoma10.....	28
Tabla 6: Chi-cuadrado asociado a la prueba de equilibrio de Hardy-Weinberg de los SNPs de <i>SFTPD</i> en todas las poblaciones.....	30
Tabla 7: Diferencias significativas entre pares de poblaciones para las frecuencias alélicas de los 23 polimorfismos de <i>SFTPD</i>	32
Tabla 8: Comparación frecuencias genotípicas entre las poblaciones Amerindias.....	35
Tabla 9: Haplotipos conformados por los tres SNPs exónicos y sus frecuencias	37
Tabla 10: Valores de Fst entre poblaciones y p-valor correspondiente (SNPs <i>SFTPD</i>)	46
Tabla 11: Valores de Fst entre poblaciones y p-valor correspondiente (región neutral)	46
Tabla 12: MAF de los SNPs que presentaron correlaciones importantes con los componentes	53
Tabla 13: Haplotipos <i>SFTPD</i>	54
Tabla 14: Frecuencias haplotípicas por población.....	55

Resumen

Los ambientes de altura representan un desafío para la capacidad de los seres humanos de sobrevivir y reproducirse debido a factores ambientales extremos, tales como las condiciones de hipoxia. En tres regiones del mundo las poblaciones humanas presentan características físicas que les permiten habitar permanentemente la altura: la meseta Tibetana-Qinghai (Asia Oriental), la meseta Semien (Etiopía) y el Altiplano Andino (Sudamérica). Los patrones genéticos que subyacen a estas características físicas son diferentes en cada población siendo un caso de convergencia evolutiva.

Un estudio genómico encontró señales de selección positiva en la población altiplánica en la región del gen *SFTPD*, que codifica para la proteína surfactante D (SP-D) y diferencias significativas en las frecuencias alélicas de tres mutaciones no sinónimas (rs3088308,rs2243639,rs721917) entre esta población y grupos Guaraní (Valverde et al., 2015). Cambios en la conformación de esta proteína podría ser ventajosos en los ambientes de altura, debido a la importancia de la SP-D en la función respiratoria.

En esta memoria, se caracterizaron las frecuencias alélicas, genotípicas y haplotípicas de las mutaciones mencionadas por Valverde et al., 2015, y de un conjunto de mutaciones intrónicas en la población altiplánica y en poblaciones que no habitan la altura. Se comparó una muestra de población Aymara de Puno con una muestra de población amerindia del sur de Chile que no habita la altura, y con muestras de siete poblaciones del 1000Genomes Project, utilizando análisis genéticos descriptivos: comparación de frecuencias alélicas, genotípicas y haplotípicas mediante la prueba exacta de Fisher, estimación de bloques haplotípicos y valores de *Fst*, construcción de dendrograma de Neighbour-Joining y Análisis de Componentes Principales.

Los resultados de esta memoria indican que la población de Puno, las poblaciones amerindias del sur de Chile y la población mestiza de Lima (PEL) son muy similares en cuanto a frecuencias alélicas y haplotípicas. Las variantes encontradas en Puno son también frecuentes en varias poblaciones a nivel global, por lo que una relación entre este gen y el patrón adaptativo de las poblaciones que habitan el altiplano andino no parece probable. Los resultados de este estudio no apoyan la señal de selección natural encontrada por Valverde et al., 2015.

Se encontró variabilidad interpoblacional en las frecuencias de los haplotipos de *SFTPD*, siendo algunos de estos característicos de ciertas regiones geográficas. En particular, se encontraron altos niveles de diferenciación genética al comparar las poblaciones occidentales con las población Han del sur de China (CHS) y con la población Japonesa de Tokyo (JPT), debido a la alta frecuencia de un haplotipo particular en estas poblaciones (>40%), el cual es muy raro en el resto del mundo (< 6%). Este haplotipo posee alelos raros de las tres mutaciones no sinónimas mencionadas, que han sido relacionados por estudios médicos a menores niveles de proteína surfactante D en sangre. Sin embargo, es más probable que esté panorama esté dado por eventos relacionados a expansiones territoriales en las poblaciones asiáticas, que por efecto de la selección natural.

Palabras clave: Adaptación a la altura, selección, surfactante pulmonar, Altiplano.

Financiamiento: Fondecyt 1140544

1. Antecedentes

1.a Adaptación humana a ambientes extremos

Los seres humanos a lo largo de su historia evolutiva se han distribuido por todos los territorios del planeta, colonizando diferentes ambientes a los cuales se han adaptado fisiológicamente, genéticamente y culturalmente. Debido a esto, las poblaciones humanas modernas presentan una notable diversidad fenotípica, influenciada por diferentes variables medioambientales, siendo esta situación más notoria en entornos de condiciones climáticas extremas.

Conceptualmente, la selección natural se basa en que existe diversidad de rasgos fenotípicos heredables en una población, los cuales son resultado (al menos en parte) de diferencias genotípicas, y en que individuos con diferentes genotipos tendrán diferentes probabilidades de supervivencia y reproducción (Nielsen & Slatkin, 2013) . La selección natural actúa sobre organismos que difieren en uno o más de estos rasgos, aumentando la frecuencia de aquellas características hereditarias que incrementan el éxito reproductivo (“*fitness*”) en un ambiente determinado y/o eliminando de la población aquellos cambios perjudiciales para el *fitness*. Una característica hereditaria es considerada una adaptación cuando mejora la supervivencia y/o la reproducción de los organismos que la portan y de su descendencia, respecto a otros que poseen estados alternativos del carácter (Futuyma, 2009). Esto permite incrementar la adecuación biológica a un determinado medio, aspecto importante en la colonización de ambientes exigentes como son los ambientes extremos.

Los ambientes extremos pueden ser definidos como aquellos caracterizados por valores en parámetros físicos, geoquímicos o biológicos que desafían la capacidad de supervivencia de los organismos (Amils, Ellis-Evans, & Hinghofer-Szalkay, 2007). En el caso de los seres humanos, se considera que la supervivencia en estos ambientes implica adaptación, siendo los ejemplos más comunes las regiones polares y las alturas (HPEE, 2016), donde sus habitantes presentan características biológicas específicas que favorecen la supervivencia y la reproducción.

1.a.1 Adaptación y selección natural en poblaciones de altura

Sobrevivir en la altura es un desafío que implica adaptación a bajas temperaturas, alta radiación UV, alta demanda energética, estrés oxidativo e hipoxia de altura. De estos aspectos, la disminución en los niveles de oxígeno causada por la baja presión barométrica (hipoxia de altura) desafía particularmente la capacidad de los seres humanos de sobrevivir y reproducirse en estos ambientes (Bigham et al., 2013).

El oxígeno es fundamental para la actividad mitocondrial y para más de 1000 reacciones enzimáticas oxígeno-dependiente en varias células y tejidos (Beall, 2007). A medida que

se asciende la tasa de difusión del oxígeno es menor, a diferentes niveles en el organismo. Esto genera un considerable estrés sobre el cuerpo, pues la hipoxia de altura es causante a su vez de cantidades de oxígeno menores a las óptimas al interior del organismo (hipoxia fisiológica). Esta condición da lugar a una serie de trastornos que incluso pueden provocar la muerte (Beall, 2007).

La migración hacia zonas de altura está asociada a problemas en las condiciones de salud y reproducción de los migrantes. Por ejemplo, durante el periodo colonial los españoles notaron una alta frecuencia de fracasos reproductivos entre sus pobladores mientras intentaban establecer presencia permanente en la ciudad de Potosí (4000 msnm), pues la mayoría de los niños morían al nacer o poco tiempo después. Por esta razón las mujeres españolas descendían hacia alturas menores durante el embarazo, el parto y durante el primer año de vida del niño (Wiley, 2004), evitando de esta manera abortos y episodios de dificultad respiratoria potencialmente letales en los recién nacidos e infantes.

Aunque estrategias culturales como estas permiten amortiguar los efectos de la hipoxia, la ocupación continua de la altura por poblaciones humanas requiere de adaptación biológica. Subsistir en estos ambientes de manera permanente y prolongada implica que no se los puede abandonar en cualquier momento, por lo que ni la tecnología tradicional ni las prácticas culturales ofrecen protección contra el estrés ambiental causado por la hipoxia de altura (Jobling, et al., 2014). Por estos motivos, la hipoxia de altura puede considerarse un potencial factor de selección natural. Sin embargo, actualmente más de 140 millones de personas viven en ambientes de altura, definidos como aquellos con una elevación superior a 2500 msnm, siendo esta la elevación a la cual la mayoría de las personas muestran una caída en la saturación de oxígeno en la hemoglobina (Moore, 2001).

Se han señalado únicamente 3 regiones geográficas en donde las poblaciones humanas poseen características fisiológicas que les permiten habitar permanentemente las regiones de gran altitud, a pesar de las severas condiciones de hipoxia y otros estresores ambientales (Valverde, et al., 2015); la meseta Tibetana-Qinghai (Asia Oriental), la meseta Semien (Etiopía) y el Altiplano Andino (Bigham et al., 2013). Las adaptaciones de estos tres grupos humanos han seguido rutas diferentes, pues cada una presenta adaptaciones respiratorias, circulatorias y hematológicas propias (Beall, 2014).

La genética humana ha proporcionado un importante enfoque para estudiar la variabilidad de las poblaciones de altura y sus adaptaciones a estos ambientes, pues la información contenida en el ADN es modelada no sólo por fuerzas evolutivas neutrales, sino también por procesos históricos de selección natural (Futuyma, 2009). Las adaptaciones de las poblaciones de altura son resultado de evolución convergente, pues las bases genéticas que subyacen a sus diferentes genotipos no son las mismas, siendo difícil identificar los genes que controlan el/los diferente/s fenotipo/s de adaptación a la hipoxia. Esto explica porque genes candidatos como el *EGLN1* y el *EPAS1*, que han sido relevantes para el estudio del patrón adaptativo de la población Tibetana, no han

proporcionado información satisfactoria sobre la población altiplánica (Bigham et al., 2013; Beall, 2014)

Los grupos altiplánicos exhiben características hereditarias que los diferencian de las otras poblaciones de altura, tales como: bajos niveles de ventilación en reposo, altos niveles de presión arterial pulmonar, altos niveles de hipertensión pulmonar, niveles de oxígeno arterial altos respecto a las poblaciones que no habitan la altura, altas concentraciones y saturaciones de hemoglobina, altas concentraciones de eritropoyetina, un hematocrito alto y un incremento en la frecuencia de “Mal de Montaña Crónico” si lo comparamos con el caso de los tibetanos (Beall et al.2007; Bigham & Lee 2014 ; Valverde et al. 2015).

Estos antecedentes plantean la necesidad de proponer nuevos enfoques de investigación, tomando en cuenta las características de las adaptaciones de la población altiplánica y las particularidades de su ambiente.

1. a.2 El altiplano: características biogeográficas

El altiplano es una meseta elevada (3000-4500 msnm) que se ubica en la región central de la cordillera de los Andes y que se extiende hacia Perú, Bolivia, el Norte de Chile y Argentina. En él se localizan numerosos emplazamientos humanos, desde pequeñas villas en el Centro de Perú y pueblos alrededor del lago Titicaca hasta ciudades como Cuzco y La Paz (Rupert & Hochachka, 2001). Algunas de estas ciudades están entre las más altas del mundo, como es el caso de Puno (3810 a 4050 msnm).

Como otras regiones montañosas, posee una distribución compleja de comunidades bióticas, determinada por la zonificación vertical, la irregularidad de la distribución biótica y características geográficas como la pendiente y la desigualdad del terreno, los cuales han producido micro hábitats con poblaciones biológicas aisladas, únicas en el mundo, que se distribuyen de acuerdo a características zonales (Moran, 2008). Posee un clima semiárido, con precipitaciones irregulares y una elevada radicación solar, debido a su latitud, su topografía plana y escaso espesor atmosférico (Sarricolea Espinoza & Romero Aravena, 2015).

Las temperaturas medias en esta región son relativamente bajas, limitando el crecimiento de la vegetación (Sarricolea Espinoza & Romero Aravena, 2015). La temperatura media anual es de 8°C, existiendo una diferencia de 20°C entre la temperatura mínima y la máxima, y es frecuente encontrar escarcha producto de las bajas temperaturas nocturnas o por el movimiento de masas frías de aire, generando estrés por el frío en las poblaciones que residen allí (Moran, 2008).

La cubierta vegetal cambia significativamente sobre los 2500 msnm. A esta altura los árboles son prácticamente inexistentes debido al hielo y en su lugar se encuentran matorrales y estepas (Moran, 2008).La productividad agrícola en el altiplano es baja debido a la aridez de su superficie, las frías temperaturas y a los bajos niveles de materia orgánica en el suelo, razón por la cual los grupos humanos han utilizado estrategias como el comercio, las redes de intercambio y las migraciones estacionales a

zonas más bajas y productivas para suplir los déficits nutricionales asociados a estas condiciones (Moran, 2008).

Finalmente, en el altiplano la presión parcial de oxígeno disminuye drásticamente porque el peso del aire sobre el organismo (el cual es responsable de la presión atmosférica) es menor. Aunque el porcentaje de oxígeno en el aire inspirado es constante a diferentes alturas, la disminución de la presión atmosférica en la altura disminuye la presión parcial de oxígeno inspirado y, por lo tanto, la presión para el intercambio de gases en los pulmones (Peacock, 1998). La presión atmosférica y la presión de oxígeno inspirado (21% de la presión barométrica) caen de manera aproximadamente lineal con la altitud, siendo a 5500 msnm sólo un 50% del valor esperado a nivel del mar (Beall, 2014; Peacock, 1998), como se puede observar en la **Figura 1**. Esta disminución de oxígeno provoca respuestas fisiológicas compensatorias a diferentes niveles en el organismo. Todos estos elementos en conjunto han presentado desafíos para el asentamiento de los seres humanos, que han habitado la región desde finales del Pleistoceno.

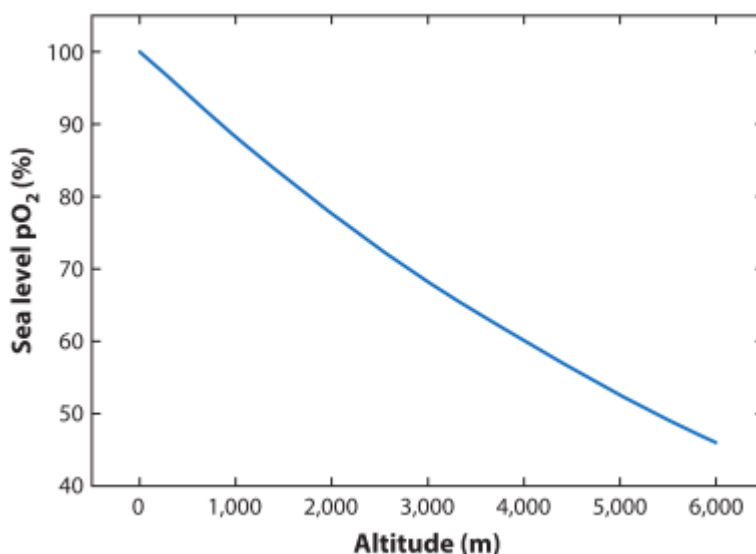


Figura 1: Disminución de la presión parcial de oxígeno inspirado en función de la altitud, expresado como porcentaje respecto a su valor a nivel del mar (extraído de Beall, 2014).

1. a. 3 Poblaciones altiplánicas

Los principales grupos étnico-lingüísticos que actualmente habitan el altiplano son los Aymaras y los Quechuas, con 2.5 y 10 millones de hablantes respectivamente (Gayà-Vidal et al., 2011) . Estas poblaciones comparten ambientes y estilos de vida similares, estando muy relacionadas genéticamente (Rupert & Hochachka, 2001). Tradicionalmente sus formas de producción se centran en pastoreo de camélidos y ovejas, y en la agricultura estacional en zonas más bajas (Moran, 2008).

La homogeneidad del cromosoma-Y y la similitud en cuanto a marcadores Alu autosómicos en estas poblaciones altiplánicas, sugiere que ambas comparten un origen

común y que ha existido una alta movilidad masculina en el área andina (Gayà-Vidal et al., 2010; Gayà-Vidal et al., 2011).

Respecto al mestizaje entre estos grupos y la población europea post-contacto, un estudio realizado en poblaciones altiplánicas de Bolivia sugiere que los Aymara presentan sólo un 3% de cromosoma-Y no amerindio y un 81% de haplogrupo mitocondrial B2 (el cual presenta frecuencias particularmente altas en el área andina), mientras que se estima que los Quechua poseerían alrededor de un 22% de cromosomas-Y de origen no amerindio y un 62% de haplogrupo mitocondrial B2 (Gayà-Vidal et al., 2011). Esto sugiere un mestizaje muy limitado con grupos foráneos en el caso de los Aymara, siendo mayor en el caso de los grupos Quechua.

La relación entre estas poblaciones altiplánicas actuales y las poblaciones prehistóricas del altiplano, así como el tiempo que llevan habitando la región es materia de discusión. La hipótesis propuesta originalmente por Bennett & Bird (1964), plantea que las actuales poblaciones altiplánicas son descendientes de grupos Paleoindios presentes en los Andes centrales hace ya 12.000 años. Este modelo migratorio sugiere que grupos nómades cazadores-recolectores cruzaron el istmo de Panamá e ingresaron por el norte a la región andina, vía los ríos Cauca y Magdalena en Colombia. Posteriormente un subgrupo de este conjunto original se habría movilizó hacia el este (hacia Venezuela y Guayana), mientras que un segundo grupo se habría movilizó hacia el Sur, hacia la región central y sur de los Andes (revisado en Rothhammer, Llop, Carvallo, & Moraga, 2001). Esta hipótesis se sustenta en el hallazgo de evidencia arqueológica que sugiere una temprana ocupación de los ambientes de altura. Sitios como Pucuncho (4355msn) y Cuncaicha (4480 msnm) en Perú, y Cueva Bautista (3800msnm) en Bolivia, poseen fechas entre los 12400 y 13000 años A.P (Capriles & Albarracín-Jordan, 2013; Rademaker et al., 2014). Todos estos sitios se encuentran a alturas donde las condiciones de hipoxia son considerables.

Otra alternativa, propuesta por Lathrap (1970), plantea que el origen de las poblaciones andinas está relacionado con las culturas de los bosques tropicales, que se desarrollaron en el área Amazónica. Hacia el 7000-9000 AP estos grupos se habrían desplazado desde las tierras bajas húmedas hacia el este de los Andes, debido al aumento poblacional relacionado al cultivo de la mandioca y al desarrollo de la agricultura (revisado en Rothhammer et al., 2001). Estudios genéticos han indicado que hay una relación más estrecha entre los grupos Andinos actuales (Aymara-Quechua) y los grupos de lengua Arawak, que con respecto a otras poblaciones amerindias del sur de América, respaldando la hipótesis de que el origen de los grupos andinos son las poblaciones amazónicas (Rothhammer et al., 2001).

En el caso de la población Aymara que comprende este estudio, es sabido que producto de la fragmentación del Imperio Tiwanaku (100AP-1200DC) aparecen los señoríos Aymara en los Andes sud-centrales, presentando posibles vinculaciones lingüísticas con Tiwanaku. También es sabido que estos señoríos posteriormente fueron conquistados por el Imperio Inca e incorporados al "Collasuyu" (Gayà-Vidal et al., 2011). Sin embargo, no se tiene certeza de si los señoríos Aymara presentan continuidad genética con sus

predecesores en la región o si simplemente sus miembros se asentaron en el territorio post caída de Tiwanaku. Algunos estudios que han comparado información genética de poblaciones andinas modernas con ADN de restos óseos precolombinos, sugieren que ha habido continuidad genética entre las poblaciones del altiplano por al menos los últimos 1000 años (Carnese et al., 2010; Fehren-Schmitz et al., 2011).

Probablemente, el origen de los grupos Aymara y de otras poblaciones altiplánicas, así como el tiempo que estas han permanecido en la región, siga debatiéndose hasta que se cuente con antecedentes que permitan determinar la continuidad o discontinuidad genética y cultural de las poblaciones que han habitado el altiplano en los diferentes periodos temporales a partir del Pleistoceno.

A pesar de la polémica respecto al origen de las actuales poblaciones altiplánicas, estudios fisiológicos y genéticos realizados tanto en grupos Aymara como Quechua sugieren que estas poblaciones han sufrido procesos de adaptación biológica.

1.a.4 Microevolución en población altiplánica y adaptación a la altura.

Numerosos rasgos morfológicos y fisiológicos han sido descritos como característicos de las poblaciones del altiplano y relacionados a su patrón adaptativo.

Estresores ambientales como el frío y la aridez han sido considerados en los estudios sobre adaptación, debido a su influencia en las estrategias de supervivencia de las poblaciones altiplánicas y a posibles cambios en las respuestas fisiológicas asociadas a las bajas temperaturas y al estrés nutricional (Moran.,2008; Weinstein et al., 2014). Sin embargo, el factor selectivo por excelencia estudiado en la altura es la hipoxia. A diferencia de otras características ambientales, la hipoxia es particularmente difícil de afrontar debido a la escases de estrategias culturales efectivas para aminorar sus efectos y a la mayor relevancia de los cambios fisiológico-adaptativos para la supervivencia.

No está del todo claro si el altiplano fue colonizado por individuos ya genéticamente predispuestos a una menor susceptibilidad a los estresores ambientales de la altura, o si la adaptación genética se dio entre los descendientes de estos primeros pobladores a través del tiempo. Sin embargo, simulaciones computacionales utilizadas para estudiar el cambio en las frecuencias alélicas en *loci* de genes previamente relacionados a la adaptación a la altura, utilizando muestras de ADN antiguo y de población actual, apoyan este último panorama (Fehren-Schmitz & Georges, 2016)

Aunque no todas las particularidades fisiológicas de las poblaciones altiplánicas son necesariamente adaptativas y probablemente muchas de las que sí lo son presentan una compleja interdependencia entre sí (Rupert & Hochachka, 2001), diferentes investigaciones han estudiado la influencia de la hipoxia en el organismo y sus posibles efectos evolutivos en las poblaciones altiplánicas.

Algunos trabajos se han centrado en analizar la heredabilidad de características físicas potencialmente beneficiosas en condiciones de hipoxia, tales como el gran tamaño del tórax que presentan las poblaciones altiplánicas ("*barrel chest*") en comparación con las

poblaciones de baja altura (Rupert & Hochachka, 2001). Un tórax más ancho se relaciona con un mayor volumen pulmonar y una mayor capacidad ventilatoria (Hurtado, 1932). Se ha encontrado evidencia de que algunas medidas torácicas son heredables, por lo que es posible que haya un componente genético relacionado a este fenotipo (Eckhardt & Melton,1992). Al comparar el desarrollo fisiológico de los hijos de sujetos Aymara y Quechua de Puno nacidos y criados en la altura, y de una contraparte nacida y criada en Tacna, se determinó que los niños desarrollaban el esternón alargado y el tórax ancho característico de las poblaciones del altiplano de las que provenían sus padres, independientemente de si eran criados a baja altura (Melton.,1992). Esta y otras investigaciones similares han apoyado la hipótesis de que el pronunciado desarrollo del tórax es una característica intrínseca, y potencialmente adaptativa, de los grupos altiplánicos (Hoff, 1972; Beall et al.,1977).

La heredabilidad de la respuesta ventilatoria en condiciones de hipoxia también ha sido estudiada en las poblaciones de altura, pues la caída en la presión de oxígeno arterial desencadena una respuesta compensatoria que aumenta la frecuencia ventilatoria (Rupert & Hochachka, 2001). En población Aymara, la exposición prolongada a condiciones de hipoxia elevada genera un respuesta ventilatoria “embotada” que tiende a la hipoventilación (Chiodi,1957). Sin embargo, esta característica no es considerada un rasgo adaptativo debido a que implica una serie de desventajas para el organismo, incluyendo una mayor susceptibilidad al sufrir “Mal de Montaña Crónico” (Rupert & Hochachka, 2001).

Otras investigaciones se han centrado en comparar los rasgos físicos de las poblaciones de altura con poblaciones de baja altura, incluyendo poblaciones genéticamente cercanas sometidas a otras condiciones ambientales, pues se espera que rasgos adquiridos bajo condiciones ambientales específicas, no estén presentes o muestren mayor variabilidad entre poblaciones emparentadas que habitan diferentes ambientes (Rupert & Hochachka, 2001).

En la altura, los hijos de sujetos de ancestría altiplánica tienden a tener un tamaño y un peso mayor que los hijos de inmigrantes provenientes de tierras más bajas, rasgo compartido con otras poblaciones de altura. El peso de nacimiento disminuye entre 88 g y 89 g por cada 1000 metros de elevación en neonatos tibetanos y altiplánicos, pero en neonatos de origen europeo y chinos Han, el peso de nacimiento disminuye en 119 g y 153 g respectivamente por cada 1000 metros de elevación (Moore, Charles, & Julian, 2011). Esto podría estar relacionado a un rasgo adaptativo, pues las mujeres de las poblaciones altiplánicas parecen tener un mayor flujo sanguíneo arterial uterino en condiciones de hipoxia durante el embarazo que las mujeres de poblaciones de baja altitud, aumentando la disponibilidad de oxígeno uteroplacentario y permitiendo que los fetos alcancen un mayor tamaño (Beall, 2014; Moore, 2001). Esto es importante si consideramos que un menor peso al nacer está relacionado a un mayor riesgo de mortalidad infantil (Moore, 2001).

Se ha comparado también la morfología esquelética de muestras arqueológicas de poblaciones altiplánicas con muestras de poblaciones andinas de costa, con el objetivo

de encontrar respuestas óseas vinculadas al estrés causado por el frío, las deficiencias nutricionales y la hipoxia (Weinstein et al., 2014). Aunque existe variabilidad intra poblacional en sitios arqueológicos como Machu Picchu, clavículas más largas y costillas con una curvatura más abierta y un área total mayor han sido encontradas en los restos arqueológicos de habitantes del altiplano, e interpretadas como características vinculadas con una función respiratoria adaptada a la hipoxia (Weinstein et al., 2014).

Finalmente, los análisis genéticos han cobrado importancia en el estudio de los patrones adaptativos en el altiplano, utilizando enfoques comparativos y genómicos. Los resultados de estos análisis son más complicados de interpretar que los resultados de los estudios fisiológicos, pues las características fenotípicas involucradas en la adaptación a la altura son rasgos poligénicos. Los fenotipos seleccionados en la altura son influenciados por múltiples alelos interrelacionados, que contribuyen en una pequeña proporción al fenotipo resultante (Hartl, 1998), el cual en muchos casos se comporta más como una gama de fenotipos posibles dentro de una población que como un rasgo discreto (Rupert & Hochachka, 2001). Esto hace difícil aislar los efectos de un gen o de un conjunto de genes específico.

Un enfoque que se ha utilizado en los estudios genéticos, es el de proponer genes candidatos a partir de alelos relacionados a fenotipos potencialmente beneficiosos en la altura que estén sobre representados en la población altiplánica (Rupert & Hochachka, 2001). Por ejemplo, alelos en el gen *β -fibrinogen* asociados a niveles bajos de fibrinógeno fueron encontrados en mayor frecuencia en población Quechua que en población Na-Dene y Caucásica Norteamericana, lo cual se interpretó como selección de fenotipos con una menor viscosidad sanguínea en el altiplano (Rupert, Devine, Monsalve, & Hochachka, 1999). Puesto que una mayor viscosidad sanguínea y un alto hematocrito producto de la hipoxia están asociados a un mayor riesgo de hipertensión y enfermedades cardíacas, tener un fenotipo con bajos niveles de fibrinógeno puede ser beneficioso, en tanto disminuye la probabilidad de sufrir estas complicaciones potencialmente letales en la altura (Rupert et al., 1999).

Los genes candidatos implicados en la adaptación a la altura son generalmente aquellos relacionados al consumo, transporte y utilización del oxígeno, aunque otros genes con una relación menos directa a los efectos de la hipoxia también han sido propuestos (Rupert & Hochachka, 2001). El desarrollo de nuevas técnicas de genotipificación y de la bioinformática han permitido proponer nuevos genes candidatos relacionados a distintas vías metabólicas en las poblaciones de altura a partir de análisis genómicos.

Fuertes señales de selección se han encontrado en genes como *EGLN1* y *PRKAA1*, asociados a las vías metabólicas del Factor Inducible por Hipoxia-1 (*HIF-1* por su sigla en inglés) y en los genes *NOS2* y *NOS3*, asociados a la regulación del óxido nítrico endotelial (revisado en Fehren-Schmitz & Georges, 2016) entre otros.

Vincular estas señales genéticas a características fenotípicas relevantes presenta sus propias complicaciones. En el caso de *EGLN1*, aunque se han encontrado señales de selección tanto en población tibetana como altiplánica (Bigham et al., 2010), trabajos

posteriores no han podido encontrar asociaciones entre genotipos de este gen y la concentración de hemoglobina en población altiplánica (Bigham et al.,2013) ni en ninguna otra característica fisiológica relevante. Otro estudio que ha escaneado el genoma de sujetos altiplánicos, no ha encontrado evidencia de que este gen esté bajo selección natural en esta población (Zhou et al., 2013).

La falta de conocimiento respecto a los patrones genéticos que subyacen a las características fisiológicas de la población altiplánica y su relación con la adaptación a estresores ambientales en la altura plantean la necesidad de evaluar nuevos genes candidatos vinculados a fenotipos potencialmente beneficiosos en la altura.

Puesto que un gran número de las investigaciones presentadas anteriormente han encontrado evidencia a favor de una relación entre la adaptación a la hipoxia y diferentes aspectos morfológicos y fisiológicos del sistema respiratorio, los genes candidatos propuestos por estudios genómicos que estén relacionados a elementos importantes del sistema respiratorio parecen ser prometedores.

1.b Surfactante pulmonar

1.b.1 Importancia adaptativa del surfactante pulmonar en el altiplano

Los alveolos son las estructuras terminales especializadas del árbol bronquial pulmonar donde se realiza el intercambio gaseoso. Para mantener la funcionalidad del alveolo es necesaria la actividad del surfactante pulmonar, un complejo lipo-proteico secretado por los neumocitos tipo II presentes en la superficie alveolar y que se compone principalmente de dos fosfolípidos (dipalmitoilfosfatidilcolina y fosfatidilglicerol) y cuatro proteínas surfactantes: A, B, C y D.

El surfactante pulmonar reduce la tensión superficial sobre el alveolo, impidiendo su colapso, permitiendo la ventilación normal y el intercambio gaseoso de O₂ y CO₂ desde las primeras etapas del desarrollo postnatal, además de proteger los pulmones de lesiones causada por partículas inhaladas y microorganismos (Griese, 1999). Los lactantes que nacen antes de siete meses de gestación aún no producen esta secreción o la producen en cantidades insuficientes, padeciendo en muchos casos de Síndrome de Dificultad Respiratoria neonatal, potencialmente letal (Gartner & Hiatt, 2000).

Según la UNICEF las afecciones respiratorias son una de las seis causas principales que producen más del 70 % de las muertes infantiles en el mundo, fundamentalmente en Asia meridional, Asia Central y África subsahariana, donde se registran las tasas más elevadas, siendo estas afecciones uno de los principales indicadores de morbilidad y mortalidad en niños menores de un año (Castro et al, 2007).

En entornos de altura, la hospitalización de infantes por problemas respiratorios (sobre todo hijos de inmigrantes de zonas bajas) es también frecuente. Wiley (2004) señala que la alta tasa de mortalidad neonatal en asentamientos de altura está probablemente relacionada con los efectos de la hipoxia, tanto durante el embarazo (cuando se restringe el crecimiento fetal) como durante el período neonatal, cuando aumenta la

vulnerabilidad a las enfermedades respiratorias e, indirectamente, el riesgo de enfermedades infecciosas como la septicemia.

Debido a las condiciones de hipoxia del altiplano, las inflamaciones, edemas, enfisemas, infecciones y necrosis pulmonares son situaciones críticas, dado que reducen la capacidad de intercambio gaseoso de la porción respiratoria del sistema ventilatorio. Las temperaturas frías, la baja humedad del aire en varias regiones y la polución doméstica que caracterizan los asentamientos altiplánicos favorecerían la proliferación de patógenos y la transmisión de enfermedades respiratorias en los Andes Peruanos (Budge et al., 2014).

La actividad inmunológica y regulatoria de las proteínas del surfactante pulmonar es fundamental para mantener un correcto funcionamiento de las vías respiratorias en estos ambientes y para la supervivencia durante esta etapa del desarrollo, debido a que los neonatos son susceptibles a sufrir enfermedades respiratorias.

La proteína surfactante D (SP-D) es codificada por el gen *SFTPD* (por su sigla en inglés “*Surfactant Protein D*”). Esta proteína regula la homeostasis del surfactante y posee varias funciones inmuno-moduladoras, incluyendo la aglutinación de virus que causan enfermedad en las vías respiratorias (Lahti, Löfgren, Marttila, Renko, & Klaavuniemi, 2002). Actúa como receptor de reconocimiento de patrones, uniéndose selectivamente a las superficies de las bacterias, virus y hongo en los alveolos, mejorando de este modo la fagocitosis de los macrófagos y la muerte de células infectadas. Se ha sugerido que niveles bajos de SP-D debido a razones genéticas pueden causar susceptibilidad a infecciones pulmonares (Leth-Larsen et al., 2005). Además, tanto la SP-D como la SP-A participan en la unión de fosfolípidos del surfactante pulmonar, los cuales son fundamentales para las funciones biofísicas del complejo. Experimentos en ratones knockout para la SP-D indican que la ausencia de esta proteína altera significativamente el metabolismo de los lípidos del surfactante, generando su acumulación en los espacios alveolares (Orgeig, Morrison, & Daniels, 2015)

Dadas la importancia inmunológica y homeostática de la SP-D, es de esperar que una actividad más eficiente de esta proteína o una mayor disponibilidad de esta fuese ventajosa en los ambientes de altura. En tal caso, variantes con efectos funcionales beneficiosos en esta proteína deberían presentar indicios de selección natural en la población altiplánica.

1. b. 2 Señales de selección y polimorfismos en el gen *SFTPD*

Dentro de la región codificante del gen *SFTPD* (posición genómica 10q22.2-23.1) existen tres polimorfismos de nucleótido simple (SNPs) que producen cambios aminoacídicos asociados a variaciones en la concentración de SP-D en el suero sanguíneo: rs721917, rs3088308 y rs2243639.

El SNP rs721917 es una transición de G a A generando un cambio en el codón 11 de la proteína. El alelo G codifica para Treonina (Thr), mientras que el alelo A codifica para Metionina (Met). El rs3088308 es una transversión de A a T y su consecuencia es el

cambio de una Serina a una Treonina en el codón 270. Esta sustitución aminoácida afectaría la función y/o la estructura de la proteína según evaluaciones hechas con PolyPhen (Valverde et al., 2015) Por último, el SNP rs2243639 corresponde a una transición de C a T y produce el cambio de una Alanina a una Treonina en el codón 160. Todas las mutaciones presentan variación en sus frecuencias alélicas dependiendo de la población, según los antecedentes disponibles en bases de datos (The 1000 Genomes Project Consortium, 2012).

Un estudio genómico reciente llevado a cabo por Valverde et al., 2015 evaluó una región candidata en el cromosoma 10 (81.1-82.6 Mb) en población altiplánica, encontrando mediante el método de “homocigosidad relativa de haplotipo extendido” (*REHH* por su sigla en inglés) una señal de selección en la región del gen *SFTPD*, y diferencias significativas entre población altiplánica y Guaraní en las frecuencias alélicas de las tres mutaciones no sinónimas previamente mencionadas. Para estas tres mutaciones, las variantes con mayor frecuencia en población de altura encontrada en el estudio de Valverde et al., 2015 son aquellas asociada a una mayor producción de proteína surfactante D por estudios médicos (Sorensen et al., 2014), como se puede ver en las **Tablas 1 y 2**.

Tabla 1: Frecuencias alélicas en población amerindia habitante de altura y de tierras bajas
Adaptado de Valverde et al, 2015.

SNP	Frecuencia alélica (Aymara Quechua)		Frecuencia alélica (Guaraní)	
	T	A	T	A
rs3088308	0,118	0,882	0,7	0,3
rs2243639	T	C	T	C
	0,526	0,474	0,2	0,8
rs721917	A	G	A	G
	0,684	0,316	0,28	0,72

Tabla 2: Concentración SP-D en sangre según genotipo (ng/mL).
Adaptado de Sorensen et al, 2014.

SNP	Genotipo	Infante maduro	Infante prematuro
rs721917	AA	964	2403
	AG	833	1913
	GG	650	1373
	TT	n.d	n.d
3088308	AT	691	1382
	AA	867	2037
	TT	893	2213
2243639	TC	921	2089
	CC	736	1649

n.d: no disponible

Se ha demostrado experimentalmente que para el rs721917 los individuos con la variante Metionina/Metionina (Met/Met11) poseen niveles promedio de SP-D en suero sanguíneo significativamente mayores que los individuos con la variante Treonina/Treonina (Thr/Thr11) o Metionina/Treonina (Leth-Larsen et al., 2005). Además, se determinó mediante cromatografía de filtración de gel que el suero de los individuos Met/Met11 presentaba dos formas de SP-D: SP-D alta y SP-D baja. La SP-D alta presente en el suero de individuos con la variante Met/Met11, posee una mayor capacidad de unión a bacterias y virus, por lo que los sujetos con el fenotipo Met/Met11 disponen de una variante inmunológicamente más eficiente (Lahti et al., 2002).

En el caso del rs3088308, la presencia del alelo ancestral que codifica para Serina está asociado a una mayor producción de SP-D, pues los homocigotos para el alelo A presentan niveles de proteína surfactante D en sangre mayores que los heterocigotos (Sorensen et al., 2014).

Finalmente, la presencia del alelo derivado codificante para Treonina del rs2243639 se relaciona con una mayor producción de proteína surfactante D. Los infantes homocigotos para el alelo T y los heterocigotos presentan niveles de SP-D significativamente mayores que aquellos homocigotos para el alelo C (Sorensen et al., 2014).

Los mismos estudios han sugerido que estos tres polimorfismos están en desequilibrio de ligamiento, conformando un haplotipo (Sorensen et al., 2014). Cuando un alelo ventajoso es llevado hacia la fijación por selección positiva arrastra consigo alelos neutrales ubicados en *loci* que están ligados estrechamente al locus afectado, proceso conocido como “barrido selectivo” (Nielsen & Slatkin, 2013). Esto deja patrones característicos en el genoma, pues la variabilidad se reduce mientras más cerca se está de los polimorfismos seleccionados, lo cual permite identificar regiones bajo barrido selectivo (Nielsen & Slatkin, 2013). Ante un barrido selectivo, una mutación que ha aumentado su frecuencia en la población disminuye la variabilidad en los sitios con los cuales presenta desequilibrio de ligamiento (Pritchard, Pickrell, & Coop, 2010).

Tomando en cuenta las fechas más tempranas para la ocupación del altiplano (alrededor de 10.000 años), si los cambios fenotípicos presentes en las poblaciones altiplánicas son adaptaciones, estas debiesen haber ocurrido rápidamente, siendo el escenario más posible un barrido selectivo fuerte producto de las condiciones de hipoxia. Este barrido debería haber dejado patrones característicos en el genoma de las poblaciones altiplánicas, generando regiones genómicas en alto desequilibrio de ligamiento y con escasa variabilidad, las cuales debiesen estar involucradas en cambios fenotípicos potencialmente beneficiosos en altura.

Datos como la alta frecuencia de alelos de *SFTPD* relacionados a cambios en la concentración de proteína surfactante D en población altiplánica y la señal de selección detectada mediante métodos genómicos en la región en la que se encuentra este gen, sugieren que dicho escenario pudo haber ocurrido, seleccionando variantes específicas del gen *SFTPD*.

Estos antecedentes y la importancia fisiológica de la SP-D, presentan al gen *SFTPD* como un candidato prometedor para estudiar la adaptación a la hipoxia en población altiplánica, el cual debe ser investigado en mayor profundidad. Considerando las condiciones de hipoxia características del altiplano surge la pregunta: *¿Existen diferencias significativas en las frecuencias de variantes del gen SFTPD entre población altiplánica y población amerindia que no habita la altura?*

2. Hipótesis:

Considerando que las condiciones de hipoxia a las cuales está expuesta la población altiplánica ejercen una presión ambiental fuerte sobre los individuos de esta población, se espera que las frecuencias de los alelos y genotipos asociados a mayores niveles de proteína surfactante D de los SNPs rs721917, rs3088308 y rs2243639, sean significativamente mayores en población altiplánica respecto a población amerindia que no habita la altura, conformando un haplotipo más frecuente en esta debido a la presión selectiva.

3. Objetivos

3.1 Objetivo general:

Caracterizar la variabilidad poblacional de las frecuencias génicas, genotípicas y haplotípicas de los SNPs rs721917, rs3088308 y rs2243639, que poseen alelos relacionados a mayores niveles de proteína surfactante D en población altiplánica.

3.2 Objetivos específicos:

- a) Determinar la presencia de diferencias significativas en las frecuencias alélicas, genotípicas y haplotípicas entre población altiplánica y población amerindia que no habita la altura.

- b) Identificar los haplotipos de *SFTPD* presentes en cada población, a partir de las combinaciones de alelos de las tres mutaciones no sinónimas y de 20 SNPs intrónicos del mismo gen.
- c) Determinar evidencia de diferenciación poblacional en los polimorfismos del gen *SFTPD*.
- d) Evaluar indicios de selección natural en los polimorfismos de interés mediante su comparación con marcadores putativamente neutrales.

4. Materiales y métodos

4.a Poblaciones y Muestras

Para representar a la población amerindia altiplánica se utilizó una muestra de sujetos Aymara de Puno (Sureste de Perú), mientras que para representar a la población amerindia de baja altura (en adelante “SUR”) se utilizó una muestra de sujetos que combina población Huilliche (Región de Los ríos y de los Lagos, Chile) y Pehuenche (Región del Bío-Bío, Chile).

Las muestras de Puno (n=82) corresponden a individuos peruanos con al menos un abuelo de ancestría Aymara, y fueron obtenidas en 2013 por el laboratorio del Dr. Carlos Bustamante de la Universidad de Stanford a partir de muestras de 5 ml de sangre. El procesamiento y control de calidad de las muestras se realizó en Stanford siguiendo protocolos estándar. Los sujetos fueron informados y consintieron su participación en el estudio, el cual fue aprobado por el Comité de Ética de la Universidad de Stanford (protocolo 20839).

Las muestras SUR(n=39) fueron recolectadas en el año 1991 y corresponden a sujetos Huilliches de la localidad de Misión, en la comuna de San Juan de la Costa, y a sujetos Pehuenches, provenientes de la comunidad Pehuenche Trapa-Trapa, en la comuna de Alto Bío Bío. El ADN se extrajo a partir de muestras de 5 ml sangre. El análisis de estas muestras en dicha ocasión se realizó con la aprobación del comité de ética de la Facultad de Medicina de la Universidad de Chile.

Todos los sujetos fueron genotipificados previamente mediante el microarreglo “*Axiom Human Genome-Wide LAT 1*” (*World Array 4*) de Affymetrix (Hoffmann et al., 2011). Los datos y las muestras provienen del Laboratorio de Genética de Poblaciones y Evolución Humana (Facultad de Medicina, Universidad de Chile). Los sujetos que componen las muestras de estas dos poblaciones presentan una proporción de ancestría amerindia mayor a 90%, determinada mediante análisis genómicos de ancestría global realizados utilizando ADMIXTURE (Alexander et al., 2011) por Paloma Contreras (P. Contreras, comunicación personal, 30 de Noviembre 2016).

Además de las dos poblaciones amerindias, se incluyó información de 7 poblaciones control (no altiplánicas) disponibles en la base de datos del proyecto 1000Genomes : Peruanos de Lima (PEL, n=85), Mexicanos-Americanos de Los Angeles California (MXL,

n=64), Población Ibérica de España (IBS, n=107), residentes de Utah con ancestría del Norte y Oeste de Europa (CEU, n=99), Luhya de Webuye Kenia (LWK, n=99), Japoneses de Tokyo (JPT, n=104) y Población Han del Sur de China (CHS, n=105), específicamente de las Provincias de Hu Nan y Fu Jian. Las muestras comprenden sólo sujetos no emparentados.

Se escogió la muestra de PEL debido a que es la más próxima a la población altioplánica de interés. La descripción de la muestra de PEL señala que algunos de estos sujetos poseían ancestros provenientes de regiones de Europa y África, por lo que se optó por incluir las muestras de IBS y LWK¹, para evaluar las diferencias respecto a estas. La muestra de MXL se incluyó con el objetivo de contrastar los resultados de PEL con otra población de ancestría en parte amerindia, bajo el supuesto de comparar poblaciones genéticamente cercanas, pero situadas en ambientes diferentes.

Las muestras asiáticas de JPT y CHS se incluyeron para analizar el comportamiento de los alelos en zonas geográficamente alejadas de América. Si bien un porcentaje de la población de Perú tiene ancestros provenientes de estos países, ninguno de los sujetos de la muestra de PEL señaló ser descendiente de inmigrantes asiáticos según la información de la base de datos de 1000Genomes.

Debido a la diferencia de tamaños muestrales entre las poblaciones amerindias (Puno=82 sujetos y SUR=39 sujetos), para los análisis comparativos se optó por comparar SUR (n=39) con un subconjunto de la muestra de Puno (en adelante “Puno99%”) compuesta por un conjunto de sujetos sin datos perdidos, con un 99% de ancestría amerindia (n=39). Estos sujetos fueron escogidos del subconjunto de individuos de la muestra de Puno que poseen un 99% o más de ancestría amerindia, mediante un generador de números aleatorios. Para las comparaciones con las muestras de 1000Genomes se utilizó el total de la muestra de Puno.

4.b Genotipificación

4.b.1 Base de datos y archivos digitales

La información genotípica de 22 SNPs del gen *SFTPD* y los SNPs de ambas muestras amerindias (**Ver Figura 2**) se extrajo de la base de datos disponible en el laboratorio, en forma de archivos binarios (.ped,.bim,.fam). Estos datos incluyeron la información de las mutaciones no sinónimas rs3088308 y rs721917.

Los genotipos del SNP exónico rs2243639 se incluyeron posteriormente en la muestra SUR, pues al no estar en la base de datos original fue necesario genotipificarlo mediante PCR y Secuenciación. Puesto que sólo se cuenta con los datos digitales de las muestras de Puno, no se pudo genotipificar este SNP para la población altioplánica. No obstante, se consideró de todas maneras dentro del data set, pues se pudo obtener información relevante respecto a sus posibles alelos en diferentes haplotipos utilizando métodos indirectos basados en el desequilibrio de ligamiento con SNPs cercanos. Las muestras

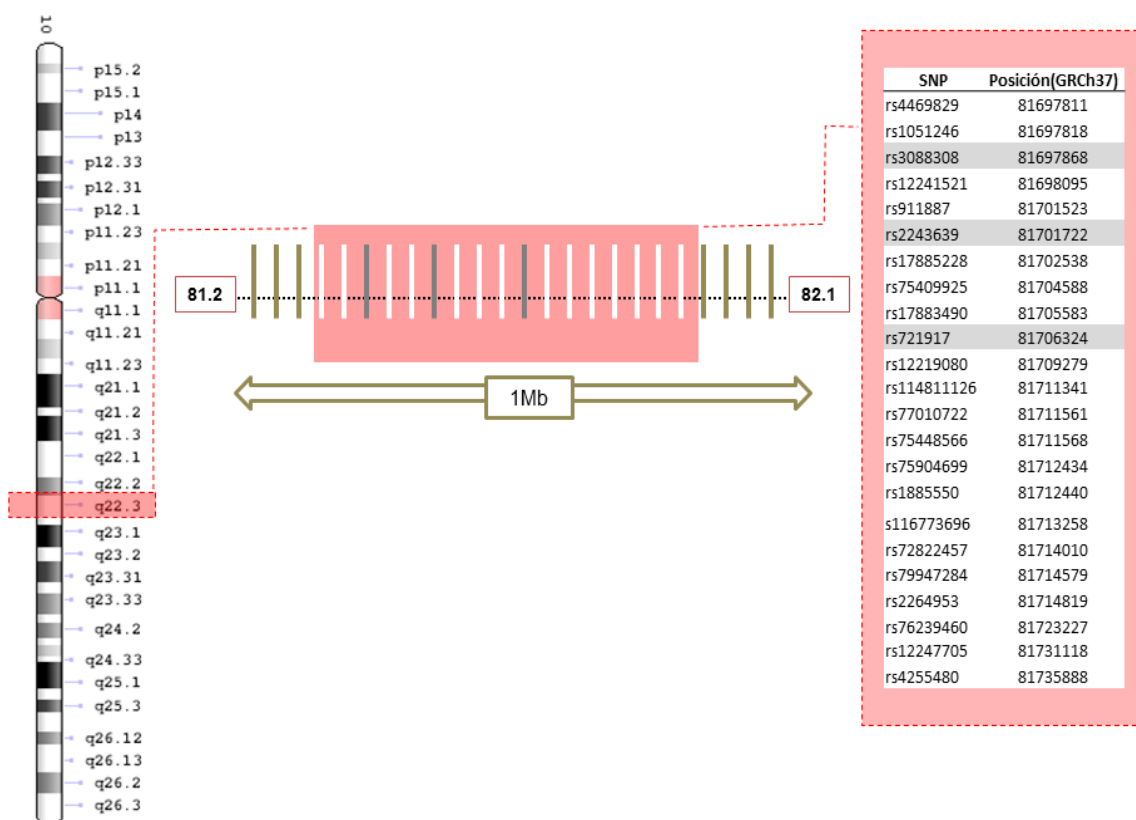
¹ La muestra de LWK fue la única de todas las poblaciones Africanas disponibles que presentó variabilidad en los 3 polimorfismos exónicos de interés. En todas las demás al menos en uno de los SNPs un alelo estaba fijado.

de 1000Genomes contaban con todos los SNPs, incluyendo rs2243639. En total se trabajó con 23 SNPs de *SFTPD*.

En la muestra de Puno se detectaron dos datos perdidos para rs3088308, dos para rs75409925 y uno para rs12219080 y para rs77010722. Al ser tan pocos datos perdidos, no se excluyó ningún sujeto en base a este criterio. En el caso de la muestra SUR, ocho SNPs presentaban datos perdidos para 10 o más sujetos. Sin embargo, como los programas utilizados descartan automáticamente los datos perdidos, y excluir a los sujetos a los que les faltaba información perjudicaba la estimación de frecuencias para los SNPs sin datos perdidos, se optó por incluir todos los sujetos y señalar en los resultados los casos en los que se trabajó con menos datos. Las muestras de 1000Genomes no presentaron datos perdidos.

Para los análisis de desequilibrio de ligamiento se consideró también la información de 142 SNPs del *Axiom Human Genome-Wide LAT 1* situados en áreas adyacentes a la región del gen *SFTPD*. Incluyendo estos SNPs, el data set utilizado consta de una región de 1Mb (posiciones 81212044 a 82189686 en GRCh37) con la región del gen *SFTPD* en medio de esta (Ver **Figura 2**).

Figura 2: Posición del gen *SFTPD* en el cromosoma 10 y data set considerado en este estudio.



Rojo: Región gen *SFTPD*, Gris: Mutaciones no sinónimas mencionadas por Valverde et al., 2015, Blanco: Mutaciones intrónicas del gen *SFTPD* consideradas en este estudio (20SNPs), Café: Área circundante utilizada en los análisis de desequilibrio de ligamiento (142 SNPs).

4.b.2 Amplificación del ADN (PCR)

En primer lugar, se confeccionaron los siguientes partidores utilizando como referencia la literatura disponible y la herramienta Primer-BLAST:

Tabla 3: Condiciones de amplificación para el SNP rs2243639.
Partidores, temperaturas de annealing y tamaño de los amplicones.

SNP	Partidor	Secuencia del primer 5´ =>3	Temperatura de annealing (°C)	Tamaño del amplificado (pb)
rs2243639	F8	CATGTTCCAGCGATACACCT	60	295
	R8	AAACCACACCTGCTGAACGC		

Para amplificar los fragmentos, se utilizaron 2µl de DNA por cada 25 µl de Mix, preparado utilizando las especificaciones descritas en la **Tabla 4**.

Tabla 4: Protocolo de PCR. Reactivos y concentraciones.

	Mix	x1
Buffer 5x		5
dNTPs	(2,5 mM)	2
MgCl₂	(25 mM)	1,5
Part. F: F8	(10 µM)	1,25
Part. R: R8	(10 µM)	1,25
Go Taq®	(5 U/µl)	0,2
Flexi DNA		
H₂O dd		11,8
Vol. Final		25

La amplificación se llevó a cabo en un termociclador, utilizando el programa "SURFACTANTE", que consta de los siguientes pasos:

- 1) Desnaturalización inicial de 5 minutos a 95°C
- 2) 30 ciclos consistentes en: 95°C por 45 segundos, 60°C por 45 segundos (temperatura de annealing) y 72°C por 45 segundos.
- 3) Elongación final a 72°C por 5 minutos.

Los productos de PCR obtenidos fueron analizados mediante electroforesis en gel de agarosa al 3%, utilizando como estándar el marcador de peso molecular

GenRuler100bp. La electroforesis se realizó en tampón TAE 1x (40 mM Trisacetato, 1 mM EDTA) a un voltaje constante de 100 volts durante 40 minutos.

4.b.3 Secuenciación

Los fragmentos amplificados mediante PCR fueron purificados y secuenciados por MacroGen Inc. (Corea), utilizando los partidores F8 y R8 anteriormente descritos.

Las secuencias resultantes se alinearon utilizando el software Genious v8.0.5 (Kearse et al., 2012) y se revisaron mediante la inspección visual de los electroferogramas.

Con los resultados obtenidos de la secuenciación se construyeron archivos .ped y .map, que posteriormente fueron transformados en archivos binarios utilizando el programa PLINK 1.07. La misma herramienta se utilizó para fusionar estos archivos con la base de datos del *Axiom LAT1*, completando de esta manera la información para la muestra SUR.

4.c Análisis genéticos y estadísticos

4.c.1 Análisis y Programas

Se comparó la muestra de población Aymara de Puno con las muestras de poblaciones que no habitan la altura, utilizando análisis genéticos descriptivos y comparativos que incluyeron la estimación y comparación de frecuencias alélicas, genotípicas y haplotípicas mediante la prueba exacta de Fisher, la estimación de bloques haplotípicos y valores de F_{st} , la construcción de un dendrograma de Neighbour-Joining y Análisis de Componentes Principales (PCA²).

Los análisis se realizaron utilizando las herramientas informáticas de libre acceso PLINK1.9 (Purcell et al, 2007)., Haploview 4.2 (Purcell et al, 2007)., R 3.0.0 (R Core Team.,2013), MEGA7 (Kumar, Stecher, & Tamura, 2016), Arlequin 3.5.2.2 (Excoffier&Lischer, 2010) y PGDSpider (Lischer & Excoffier, 2012).

4.c.2 Desequilibrio de ligamiento y haplotipos

Una vez estimadas las frecuencias alélicas y comparadas entre las poblaciones, se evaluó el desequilibrio de ligamiento tanto entre los tres polimorfismos exónicos de interés como entre los SNPs que presentaron diferencias significativas en sus frecuencias alélicas entre las poblaciones amerindias. Para ello se estimó el parámetro r^2 entre pares de SNPs, el cual es un coeficiente de correlación derivado del coeficiente de desequilibrio de ligamiento (D) que, al igual que D', adquiere valores entre 0 y 1 que describen la independencia de los alelos entre dos *loci* (Nielsen & Slatkin, 2013). Los valores cercanos a 1 indican desequilibrio de ligamiento fuerte.

La estimación de estos parámetros se realizó mediante el software Haploview 4.2 y la estimación de los haplotipos en la muestra se realizó utilizando el Algoritmo Esperanza-Maximización (algoritmo EM) incorporado en este mismo software. El algoritmo EM crea estimaciones de frecuencias haplotípicas poblacionales altamente precisas basándose

² Abreviación por su nombre en inglés "*Principal Component Analysis*".

en la máxima verosimilitud del input sin fase (Barrett, Bender, Maller, & Whitworth, 2016).

También se utilizó Haploview 4.2 para estimar el patrón de bloques haplotípicos de cada población mediante el método propuesto por Gabriel et al., 2002. Los bloques haplotípicos son regiones donde sólo se observan unos pocos haplotipos y donde hay muy poca evidencia de recombinación (Gabriel et al., 2002). Estos bloques se determinan a partir de patrones de recombinación entre pares de polimorfismos, en función de intervalos al 95% de confianza del parámetro D' .

Para que se considere que un par de alelos está en desequilibrio de ligamiento fuerte, el valor de D' en el límite superior del intervalo de confianza debe estar sobre 0.98 y su valor en el límite inferior debe estar sobre 0.75. Si el valor de D' en el límite superior del intervalo de confianza está bajo 0.9, se considera que el par presenta evidencia de recombinación (Gabriel et al., 2002). Un bloque de haplotipos se define entonces como una región donde un porcentaje menor al 5% de todas las comparaciones hechas entre pares de polimorfismos presenta evidencia de recombinación (Gabriel et al., 2002).

Para evaluar el desequilibrio de ligamiento entre bloques se utilizó el valor de D' multilocus, también calculado por Haploview.

4.c.3 Fst y Neighbour-Joining

Para determinar la existencia de diferenciación genética entre las poblaciones respecto los SNPs de interés, se estimó el valor de F_{st} utilizando el software *Arlequin* ver 3.5.2.2. Este análisis resume las diferencias de frecuencia de alelos entre pares de poblaciones o entre múltiples poblaciones. Con estos datos se construyó un dendrograma de Neighbour-Joining para visualizar gráficamente las distancias genéticas entre las poblaciones.

Para evaluar diferencias en los valores de F_{st} obtenidos de los SNPs del *SFTPD* respecto a los esperados sólo por efecto de deriva, se compararon estos valores observados con los obtenidos de un conjunto de SNPs presentes en el *Axiom LAT1* que a priori no deberían estar bajo selección.

Estos SNPs putativamente neutrales provienen de una región sin genes del cromosoma 10 (36500000-36700000), alejada de la región donde se localiza el *SFTPD*. Con tal de evitar cualquier SNP que pudiese estar ligado a regiones bajo selección en la muestra de Puno, a la zona escogida se le aplicó un filtro utilizando el programa PLINK 1.9. Con ello se descartaron todos los SNPs que presentaron valores de r^2 mayores a 0.2 con otros polimorfismos de la zona. En total esta "región neutral" consta de 14 SNPs no asociados a gen alguno que se muestran en la **Tabla 5**.

Adicionalmente, se realizaron análisis de F_{st} Outlier utilizando el software de libre acceso LOSITAN (Antao, Lopes, Lopes, Beja-pereira, & Luikart, 2008), basado en el método F_{st} Outlier f_{dist} (Beaumont & Nichols, 1996). Este método evalúa la relación de F_{st} vs heterocigosidad esperada (H_e) bajo un "modelo de migración de isla" mediante el

uso de marcadores neutrales teóricos. Describe la distribución de F_{st} vs H_e bajo este modelo y posteriormente contrasta esta distribución con los valores obtenidos para cada uno de los SNPs analizados. Esto permite identificar *loci* que poseen valores de F_{st} excesivamente altos o bajos respecto a lo esperado según un modelo neutral, así como identificar SNPs candidatos potencialmente afectados por selección natural (Antao et al., 2008) .

Tabla 5: SNPs zona neutral cromosoma 10

SNP	Posición(GRCh37)
rs1219603	36543314
rs2800115	36574697
rs1219615	36590152
rs58700445	36613159
rs16936933	36615953
rs1125556	36620416
rs7905986	36632635
rs1631619	36638425
rs10827630	36646248
rs2800086	36651245
rs17544598	36656814
rs59663372	36668880
rs7911616	36672553
rs10827637	36688528

4.c.4 Análisis de Componentes Principales (PCA)

Finalmente, se realizaron Análisis de Componentes Principales para evaluar el agrupamiento de las poblaciones observado en los análisis previos e identificar componentes clave de la estructuración poblacional observada, utilizando el software *R 3.0.0*.

El primer PCA se realizó a partir de la frecuencia del alelo menor (“MAF” por su sigla en inglés) de los 22 SNPs de *SFTPD* presentes en todas las poblaciones. El segundo PCA se realizó a partir de las frecuencias de los haplotipos de los 23 SNPs totales de *SFTPD*, estimados en las diferentes poblaciones mediante el algoritmo EM.

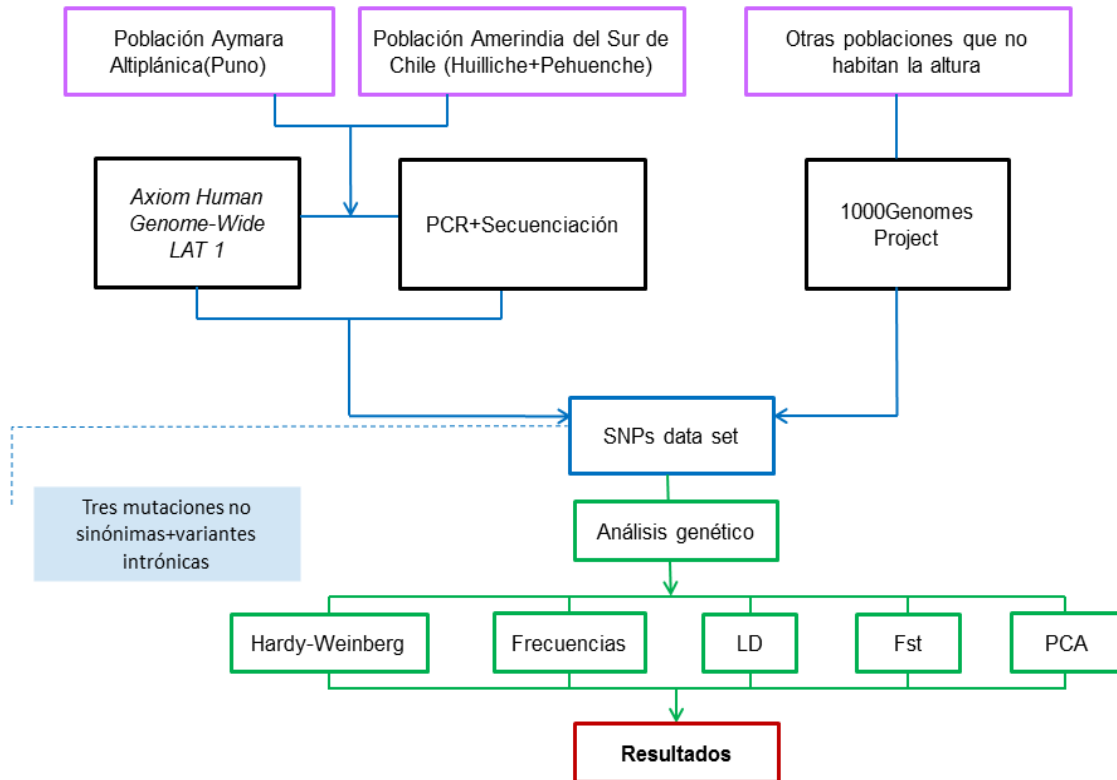


Figura 3: Resumen de las metodologías utilizadas y su relación

5. Resultados

5.a Equilibrio de Hardy-Weinberg

Para cada SNP se obtuvieron los genotipos, las frecuencias alélicas, la heterocigocidad observada y la heterocigocidad esperada en cada población. Se estimó si las poblaciones estaban en equilibrio de Hardy-Weinberg para cada polimorfismo utilizando la herramienta PLINK 1.07.

Sólo los p-valores del test de chi-cuadrado asociado a la prueba de equilibrio de Hardy-Weinberg de la mutación no sinónima rs2243639 en SUR, y de los SNPs intrónicos rs1051246 en PEL y rs72822457 en CEU fueron estadísticamente significativos a un 95% de confianza ($p < 0.05$), lo cual indica desviación respecto al equilibrio de Hardy Weinberg para estos SNPs. El resto de las poblaciones se encuentra en equilibrio genético para los demás polimorfismos estudiados, incluyendo los polimorfismos exónicos de interés. En 16 ocasiones se observó un déficit de heterocigotos respecto a lo esperado, aunque sólo en las tres ocasiones antes mencionadas este déficit se relacionó a desviaciones respecto al equilibrio (Ver **tabla 6**).

Tabla 6: Chi-cuadrado asociado a la prueba de equilibrio de Hardy-Weinberg de los SNPs de *SFTPD* en todas las poblaciones

Población	SNP	Alelo 1	Alelo 2	Genotipos	HET (Obs)	HET(Esp)	p-valor
Puno	rs3088308	T	A	TT=2/TA= 14/AA=64	0.175	0.1997	0.2453
	rs2243639	T	C	X	X	X	X
	rs721917	G	A	GG=5/GA=31/AA=46	0.378	0.375	1
SUR	rs3088308	T	A	TT=3/TA= 6/AA=13	0.2727	0.3967	0.27
	rs2243639*	C	T	CC=12/TC=11/TT=15	0.2895	0.4969	0.01043
	rs721917	G	A	GG=5/GA=15/AA=19	0.3846	0.4356	0.4723
PEL	rs1051246*	G	A	GG=5/GA=17/AA=63	0.2	0.2672	0.03026
	rs3088308	T	A	TT=1/TA= 17/AA=67	0.2	0.1985	1
	rs2243639	T	C	TT=21/TC=37/CC=27	0.4353	0.4975	0.2766
	rs721917	G	A	GG=5/GA=36/AA=44	0.4235	0.3947	0.5933
MXL	rs3088308	T	A	TT=4/TA=20 /AA=40	0.3125	0.3418	0.4725
	rs2243639	T	C	TT=9/TC=32/CC=23	0.5	0.4761	0.7962
	rs721917	G	A	GG=12/GA=33/AA=19	0.5156	0.494	0.8045
CEU	rs72822457*	C	T	CC=1;CT=2;TT=96	0.0202	0.03959	0.03038
	rs3088308	T	A	TT=1/TA=13 /AA=85	0.1313	0.14	0.4361
	rs2243639	T	C	TT=20/TC=40/CC=39	0.404	0.4816	0.1424
	rs721917	G	A	GG=24/GA=40/AA=35	0.404	0.4938	0.06993
IBS	rs3088308	T	A	TT=0/TA=18 /AA=89	0.1682	0.1541	1
	rs2243639	T	C	TT=16/TC=45/CC=46	0.4206	0.4607	0.4017
	rs721917	G	A	GG=26/GA=45/AA=36	0.4206	0.4956	0.1209
JPT	rs3088308	T	A	TT=0/TA=6 /AA=98	0.05769	0.05603	1
	rs2243639	T	C	TT=7/TC=39/CC=58	0.375	0.3798	1
	rs721917	A	G	AA=17/AG=49/GG=38	0.4712	0.4796	0.84
CHS	rs3088308	T	A	TT=0/TA=1 /AA=104	0.009524	0.009478	1
	rs2243639	T	C	TT=2/TC=36/CC=67	0.3429	0.3084	0.3525
	rs721917	A	G	AA=14/AG=53/GG=38	0.5048	0.4739	0.6799
LWK	rs3088308	T	A	TT=0/TA=2 /AA=97	0.0202	0.02	1
	rs2243639	T	C	TT=0/TC=2/CC=97	0.0202	0.02	1
	rs721917	G	A	GG=21/GA=44/AA=34	0.4444	0.4914	0.4123

*HET(Obs): Heterocigocidad observada, HET(Esp): Heterocigocidad esperada. * =déficit de heterocigotos en poblaciones que se desvían del equilibrio genético (valores significativos al 95% de nivel de confianza). En gris: SNPs que presentaron p-valores significativos.*

Puno=Aymaras de Puno, SUR=población amerindia del sur de Chile, PEL= Peruanos de Lima, MXL= Mexicanos-Americanos de Los Angeles California, IBS= Población Ibérica de España, CEU= residentes de Utah con ancestría del Norte y Oeste de Europa, LWK= Luhya de Webuye Kenia , JPT= Japoneses de Tokyo y CHS= Población Han del Sur de China.

5. b Frecuencias alélicas

5.b.1 Frecuencias SNPs exónicos

A partir de los datos contenidos en los archivos binarios, se estimaron las frecuencias génicas de los SNPs exónicos de interés en cada muestra (ver Figura 4).

El alelo "A" del SNP rs3088308, asociado a mayores niveles de surfactante pulmonar, es el más frecuente en todas las poblaciones. Está prácticamente fijado en LWK, en las poblaciones asiáticas y en las poblaciones europeas. Alcanza una frecuencia de 0.89 en Puno y de 0.88 en PEL. Si bien este alelo también es el más frecuente en la muestra SUR (A=0.73), su frecuencia es más baja respecto a las otras poblaciones. También es en esta última población donde el alelo T alcanza su frecuencia más alta.

No se pudieron estimar las frecuencias génicas de rs2243639 en Puno por carecer de los datos. Sin embargo, a partir del panorama general se puede concluir que el alelo más frecuente en casi todas las poblaciones es "C", el cual está fijado en LWK. La única población donde la frecuencia del alelo "T" supera a la del alelo "C" es en la representada por la muestra SUR (T=0.54).

Finalmente, el alelo "A" de rs721917 es el más frecuente en todas las poblaciones, excepto en las poblaciones asiática, donde la frecuencia de "G" es notablemente más alta (JPT=0.60 y CHS=0.61). El alelo "A" alcanza su frecuencia más alta en Puno (A=0.75) y PEL (A=0.72), es un poco más baja en SUR (A=0.68) y no supera el 0.57 en el resto de las poblaciones (IBS=0.55, CEU=0.56, MXL=0.55, LWK=0.57).

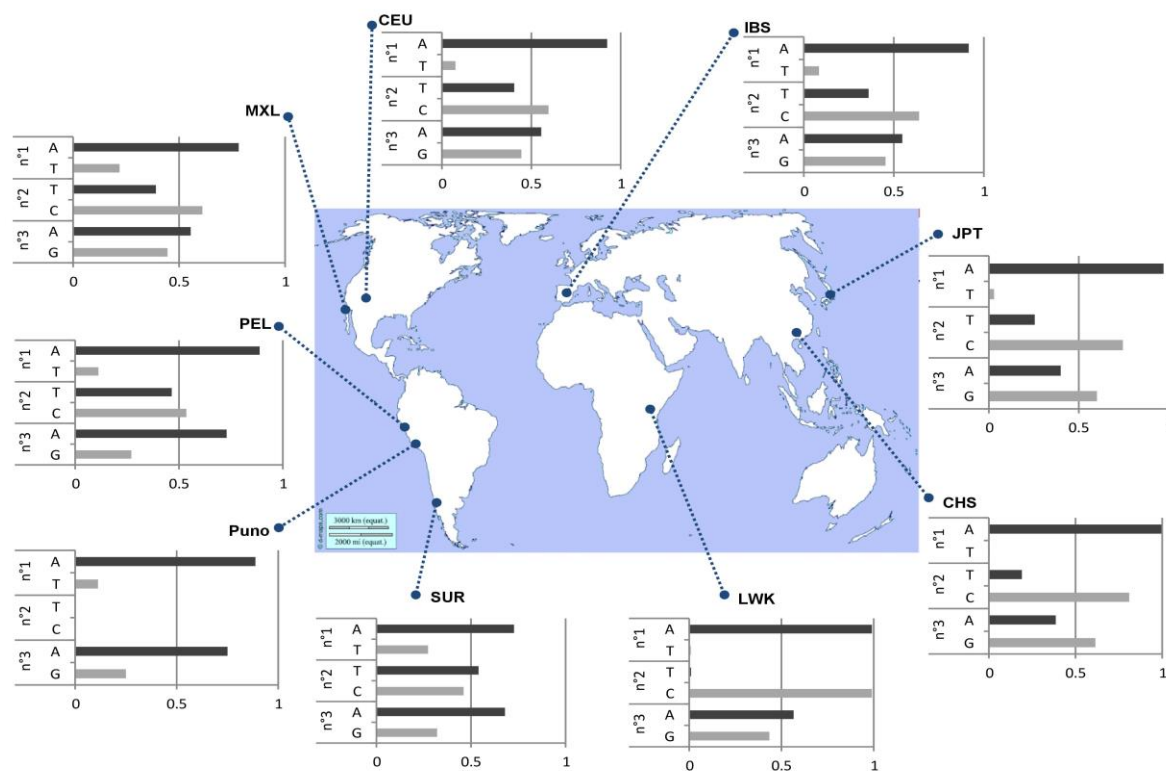


Figura 4: Frecuencias alélicas de los SNPs exónicos asociados a cambio aminoacídico n°1=rs3088308, n°2=rs2243639, n°3=rs721917

5.b.2 Diferencias significativas: Prueba exacta de Fisher

Para determinar la presencia de diferencias significativas entre las frecuencias alélicas de la muestra de Puno y las de las poblaciones de referencia, se utilizó la prueba exacta de Fisher al 95% de confianza utilizando el programa PLINK 1.09. También se utilizó este test para comparar las frecuencias alélicas encontradas en SUR con las de las poblaciones de 1000 Genomes, con el fin de obtener información adicional para interpretar los resultados. Los resultados se resumen en la **Tabla 7**.

Tabla 7: Diferencias significativas entre pares de poblaciones para las frecuencias alélicas de los 23 polimorfismos de *SFTPD* (valores significativos de p-valor de la prueba exacta de Fisher).

SNP	Puno x							SUR x							
	PEL	SUR	MXL	IBS	CEU	JPT	CHS	LWK	PEL	MXL	IBS	CEU	JPT	CHS	LWK
rs4469829*								4.126e-008							0.006156
rs1051246		9.735e-005				0.0013	1.038e-006	0.0117					4.566e-007	1.89e-01	
rs3088308*	0.0213	0.01577				0.0021	1.851e-006	2.464e-005	0.01404		0.001226	0.00069	1.686e-006	2.082e-009	2.283e-008
rs12241521						0.003	0.002897	4.559e-008					0.02329		2.451e-008
rs911887		0.01226									0.003882	0.03286		0.04731	
rs2243639	/	/	/	/	/	/	/	/	0.04251		0.006807		1.449e-005	3.434e-008	3.162e-025
rs17885228								0.0001474							0.0008371
rs75409925*															
rs17883490*								3.342e-007							0.0102
rs721917		0.0007	6.168e-005	0.00016	1.101e-011	1.462e-012	0.000262			0.04509			3.192e-005	1.014e-005	
rs12219080	0.0136	0.0069	0.011	0.002572	3.757e-026	9.974e-028	1.48e-011						7.284e-008	1.033e-008	0.03209
rs114811126*							5.119e-009								0.002091
rs77010722*			0.0015								0.01374				
rs75448566*	0.033						0.004738		0.005707	0.01127	0.003201	0.0039	0.003442	0.003359	
rs75904699*							2.742e-006								0.01732
rs1885550	0.039	0.0087			1.609e-013	7.49e-015	0.0001608		0.02398			0.04147	0.000531	0.0001189	
rs116773696															
rs72822457*							4.48e-005								0.01044
rs79947284	0.013	0.006	0.0205	0.0024	2.545e-026	1.969e-027	0.009117		0.01236				7.284e-008	1.912e-008	
rs2264953	0.006	0.0031	8.7e-009	2.9e-008	0.01904	0.004638	1.922e-013		3.04e-005	5.039e-008	1.315e-013	3.79e-013			1.945e-017
rs76239460		0.024			0.0013	1.038e-006	0.008344		0.02056		0.00199	0.00075	4.566e-007	1.89e-010	5.792e-006
rs12247705	0.013	0.006	0.02054	0.0024	1.044e-026	1.969e-027	6.275e-011		0.03942				3.88e-008	1.912e-008	0.06641
rs4255480						0.00769					0.02329		0.0008954	5.183e-005	0.0004858

*= SNPs con datos perdidos, "e-": notación científica

Respecto a los SNPs exónicos de interés, las frecuencias alélicas de la muestra de Puno para rs721917 fueron significativamente diferentes respecto a todas las poblaciones, con excepción de PEL y SUR. Las frecuencias génicas de rs3088308 en Puno sólo fueron significativamente diferentes respecto a SUR, MXL, LWK y con respecto a las de las poblaciones asiáticas (JPT, CHS). El polimorfismo rs2264953 no se pudo evaluar, pues no se cuenta con datos de este en la muestra de Puno.

Al comparar las frecuencias alélicas de los SNPs exónicos de interés entre SUR y las poblaciones de 1000Genomes, las frecuencias de rs3088308 presentaron diferencias respecto todas las poblaciones, excepto MXL. Las diferencias en las frecuencias de rs2243639 sólo no fueron significativas respecto a PEL y a CEU. Respecto a rs721917,

este sólo presentó diferencias significativas respecto a IBS y respecto a las poblaciones asiáticas.

Al comparar las frecuencias génicas de Puno y SUR, siete polimorfismos presentaron diferencias significativas: rs3088308 (polimorfismos exónico), rs75448566, rs1885550, rs12219080, rs79947284, rs2264953 y rs12247705. Los SNPs rs3088308 y rs75448566 presentaron un 45% y un 52% de datos perdidos respectivamente en la muestra SUR, por lo que estos resultados deben ser tomados con precaución. Todos los demás polimorfismos mencionados corresponden a SNPs intrónicos de *SFTPD* sin una cantidad relevante de datos perdidos.

Las frecuencias alélicas de estos polimorfismos se resumen en las **Figuras 5.A-5.G**. Todos los SNPs que presentaron diferencias significativas entre Puno y SUR también las presentaron al comparar SUR y PEL.

La variación interpoblacional de rs3088308 fue previamente discutida en el apartado 5.b.1. En el caso de rs75448566, el alelo "G" parece estar fijado en todas las poblaciones, excepto en SUR y en LWK, donde el alelo "A" se puede encontrar en baja frecuencia. Respecto al SNP rs1885550, el alelo más frecuente es "C" en todas las poblaciones, excepto en las poblaciones asiáticas, donde el alelo T tiene predominancia (JPT=0,55; CHS=0,57). Las diferencias son significativas al contrastar Puno con SUR, MXL, JPT, CHS y LWK, pero no al contrastar Puno con las poblaciones europeas (IBS y CEU).

Los polimorfismos rs12219080, rs79917284, rs2264953 y rs12247705 presentaron además diferencias significativas al contrastar la muestra de Puno con todas las de 1000Genomes. Al contrastar SUR con las poblaciones de 1000Genomes las diferencias son significativas respecto a las poblaciones asiáticas y respecto a LWK en tres de los cuatro polimorfismos mencionados (rs12219080, rs79947284, rs12247705). En el caso de estos tres SNPs, el alelo más frecuente en todas las poblaciones es "C", pero su frecuencia es igual a 1 solamente en Puno (**Ver Figuras 5.B, G y E**). Para todos estos polimorfismos el alelo "T" presenta una frecuencia inferior a 0.2 en todas las poblaciones, a excepción de las poblaciones asiáticas JPT (rs12219080 (T)= 0.41; rs79947284 (T)= 0.41; rs12247705 (T)=0.41) y CHS (rs12219080 (T)=0.42; rs79947284 (T)=0.42; rs12247705 (T)=0.42).

Finalmente, en el caso de rs2264953 el alelo "T" presenta una frecuencia superior al 50% en todas las poblaciones, pero sólo en SUR su frecuencia es igual a 1 (**Ver Fig5.F**).

Fig 5. A

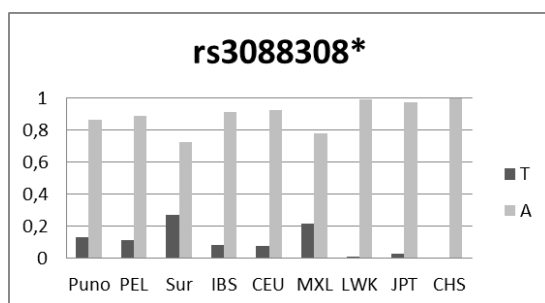


Fig5.B

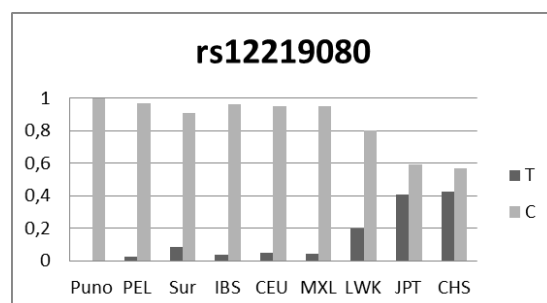


Fig5.C

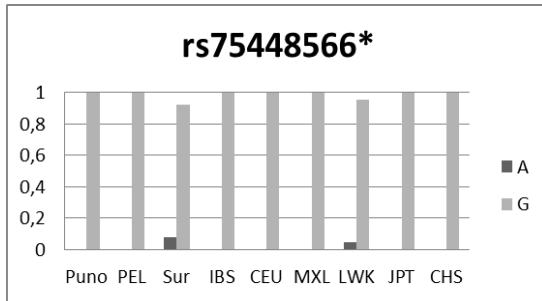


Fig5.D

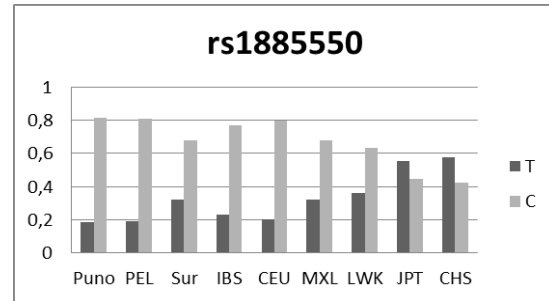


Fig5.E

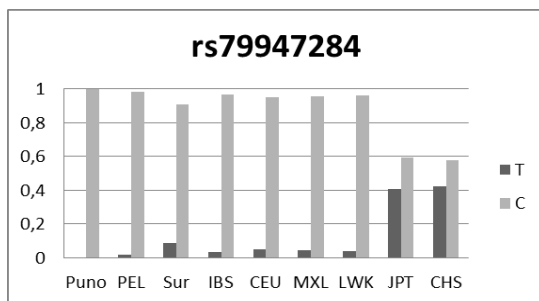


Fig5.F

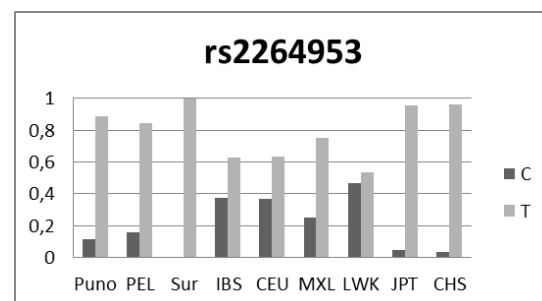
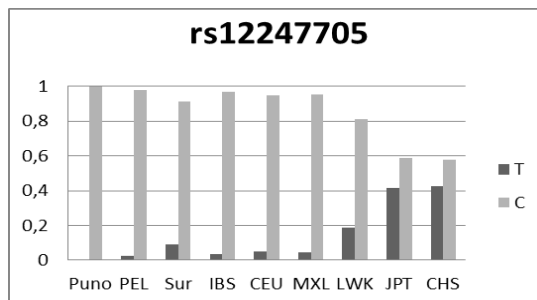


Fig5.G



Figuras 5(A, B, C, D, E, F, G): Gráficos de las frecuencias alélicas de los SNPs que presentaron diferencias significativas entre las poblaciones amerindias.

*=presenta datos perdidos.

5. c Frecuencias genotípicas

Se compararon las frecuencias genotípicas de los siete polimorfismos de *SFTPD* que en el análisis anterior presentaron diferencias significativas en sus frecuencias alélicas entre Puno y SUR (Ver **Tabla 7**). Se compararon los genotipos de la submuestra Puno99% (n=39) y SUR(n=39), utilizando la extensión Freeman-Halton de la prueba exacta de Fisher para tablas de contingencia de hasta tres filas por tres columnas (Freeman&Halton.,1951) a un 95% de confianza (prueba de dos colas), implementado

en la herramienta online disponible en vassarstats.net. Los resultados de este análisis se resumen en la **Tabla 8**.

El único polimorfismo exónico del grupo (rs3088308) no presentó diferencias significativas en sus frecuencias genotípicas, mientras que de los SNPs intrónicos mencionados anteriormente cinco presentaron diferencias significativas (rs12219080, rs79947284, rs12247705, rs75448566 y rs2264953). En el caso de rs12219080, rs79947284, rs12247705 y rs75448566, un alelo está fijado sólo en Puno, mientras que en el caso de rs2264953 un alelo está fijado sólo en SUR. En ambas situaciones, al haber un alelo fijado en una población hay un único genotipo posible, por lo que las diferencias respecto a la otra población aparecen como significativas, aunque cuantitativamente sean de sólo unos pocos sujetos.

Tabla 8: Resultados de la prueba exacta de Fisher, utilizando las frecuencias absolutas de los genotipos de los SNPs que presentaron diferencias significativas en sus frecuencias alélicas entre Puno y SUR.

SNP	Genotipo	Puno99%	SUR	Fisher (p-valor)
rs3088308*	TT	2	3	0.065
	TA	4	6	
	AA	33	13	
rs2264953	CC	0	0	0.005
	CT	8	0	
	TT	31	39	
rs12219080*	TT	0	0	0.011
	TC	0	7	
	CC	38	32	
rs1885550	TT	2	5	0.109
	TC	9	15	
	CC	28	19	
rs79947284	TT	0	0	0.011
	TC	0	7	
	CC	39	32	
rs12247705	TT	0	0	0.011
	TC	0	7	
	CC	39	32	
rs75448566*	AA	0	0	0.031
	AG	0	3	
	GG	39	16	

*=snps con datos perdidos en SUR, en negrita = p-valor<0.05.

5.d Desequilibrio de ligamiento

5.d.1 Haplotipos SNPs exónicos

Los haplotipos conformados por los tres SNPs exónicos y sus frecuencias en todas las poblaciones se resumen en la **Tabla 9**.

Respecto a los SNPs exónicos rs3088308 y rs721917, estos no presentan un valor de r^2 particularmente alto en Puno ($r^2=0.38$), siendo este valor mayor en SUR ($r^2=0.79$). Los haplotipos que portan el alelo “A” de ambos SNPs están relacionados a mayores niveles de *SFTPD*. En ambas poblaciones amerindias, el “haplotipo AA” es el más frecuente (Puno=75%; SUR=67.5%). En el caso de Puno, AG es el segundo más frecuente (13.8%), mientras que en el caso de SUR sólo tiene una frecuencia de 5%, siendo el haplotipo TG (27.5%) el segundo más frecuente. Este último tiene una frecuencia de sólo 11.2% en Puno.

No se pudo estimar el parámetro r^2 para rs2243639 en la muestra de Puno, por carecer de los datos necesarios. Su valor con los otros SNPs exónicos no es muy alto en la muestra SUR (rs2243639-rs3088308=0.42, rs2243639-rs721917=0.25). Los haplotipos que presentan el alelo T de rs2243639 y el alelo A de los otros dos SNPs son los que presentan las frecuencias más altas en SUR (rs3088308-rs2243639=AT=52.5%; rs721917-rs2243639=AT=48%). En PEL (la población que presenta menos diferencias con Puno) el desequilibrio de ligamiento de este polimorfismo con los otros SNPs exónicos tampoco es muy fuerte (r^2 rs2243639-rs3088308=0.11, r^2 rs2243639-rs721917=0.28). Al igual que en la muestra SUR, en PEL los haplotipos que presentan el alelo T de rs2243639 y el alelo A de los otros dos SNPs son los que presentan las frecuencias más altas (rs3088308-rs2243639=AT=46.5%; rs721917-rs2243639=AT=45.7%). Los valores de r^2 entre los SNPs exónicos rs3088308 y rs721917 en PEL son semejantes a los de Puno ($r^2=0.34$) y las frecuencias de sus haplotipos también (AA=72.9%, AG=15.9%, TG=11.2%).

De las posibles combinaciones de alelos, el haplotipo de mayor interés es “ATA”, pues combina los alelos que los antecedentes bibliográficos relacionan a mayores niveles de proteína surfactante D (A, T y A de los SNPs rs3088308, rs2243639 y rs721917 respectivamente). Este haplotipo alcanza su máxima frecuencia en PEL y SUR (0.46), disminuye en MXL (0.36) y es más bajo en las poblaciones europeas (CEU=0.40, IBS=0.46). En las poblaciones asiáticas es incluso más raro (CHS=0.20, JPT=0.26) y está ausente en LWK. Al no contar con datos de rs2243639 en Puno, el haplotipo “ATA” no pudo observarse, pero la frecuencia del haplotipo “A?A” es más alta en Puno que en cualquier otra población (Puno=0.75). La frecuencia del haplotipo “ATA” es igual en PEL y en SUR, y las diferencias en las frecuencias de los otros dos haplotipos que comparten estas poblaciones (“ACA” y “TGC”) no son significativas al 95% de nivel de confianza según la prueba exacta de Fisher (p-valor= 0.133). Considerando la similitud entre Puno y PEL, resultados análogos serían esperables en Puno.

Tabla 9: Haplotipos conformados por los tres SNPs exónicos y sus frecuencias

SNPs exónicos	Poblaciones								
	Puno	PEL	SUR	MXL	CEU	IBS	LWK	CHS	JPT
rs3088308/rs2243639/rs721917	A?A=0.75	ATA=0.46	ATA=0.46	ATA=0.39	ATA=0.40	ATA=0.36		ATA=0.20	ATA=0.26
		ACA=0.27	ACA=0.22	ACA=0.16	ACA=0.15	ACA=0.19	ACA=0.57	ACA=0.20	ACA=0.14
	A?G=0.14	ACG=0.15	ATG=0.06	ACG=0.23	ACG=0.36	ACG=0.37	ACG=0.41	ACG=0.61	ACG=0.57
	T?G=0.11	TCG=0.11	TCG=0.26	TCG=0.22	TCG=0.076	TCG=0.08	TCG=0.01		TCG=0.03
						ATG=0.01			

Gris= haplotipo asociado a mayores niveles de SFTPD.

5. d. 2 Bloques haplotípicos

Utilizando el data set completo (23 SNPs de *SFTPD* y 142 SNPs de las áreas adyacentes) se estimaron los patrones de bloques haplotípicos de cada población siguiendo el método de Gabriel et al., 2002.

Se utilizó el software Haploview 4.2 para estimar el patrón de bloques (Ver **Figuras 6.a – 6.k**) y se usó el “esquema de colores de intervalos de confianza” correspondiente al método mencionado para resumir los resultados del desequilibrio de ligamiento entre pares de polimorfismos. Los cuadrados gris oscuro indican que los pares de polimorfismos que representan poseen en el límite superior del intervalo de confianza de D' un valor sobre 0.98, y en el límite inferior un valor sobre 0.75, razón por la que se considera que estos polimorfismos están en un fuerte desequilibrio de ligamiento. Estos cuadrados tienden a presentar los valores de r^2 más altos, generalmente sobre 0.5. Por su parte, los cuadrados blancos representan pares de polimorfismos cuyos valores en el límite superior del intervalo de confianza de D' están bajo 0.9, por lo que se considera que presentan evidencia de recombinación (Gabriel et al., 2002), además tienden a tener valores de r^2 cercanos a 0. Los cuadrados gris claro representan pares de polimorfismos cuya relación no puede clasificarse dentro de ninguna de las dos categorías previas, por lo que se consideran “no informativos” para identificar patrones dentro del genoma. El desequilibrio de ligamiento entre bloques haplotípicos de una misma población se estimó utilizando el parámetro D' para múltiples *loci* (“ D' multilocus”)

Si el gen *SFTPD* estuviese bajo un barrido selectivo, se esperaría un gran bloque haplotípico en su región genómica, con altos valores de r^2 y D' (cuadrados gris oscuro) en comparación con las áreas adyacentes.

Se encontraron 11 bloques en la muestra de Puno, pero ninguno de ellos incluía la totalidad de los SNPs del *SFTPD*.

El bloque 4 (11kb) es el único que contiene tres de los SNPs de *SFTPD* (rs1051246, rs3088308 y rs911887). Los otros SNPs que conforman el bloque pertenecen al pseudogen *BMS1P21* (rs1570328, rs2758555, rs2758554) y al gen no caracterizado LOC105371611 (rs7535771).

Este bloque es pequeño y, aunque algunos pares de polimorfismos presentan valores de r^2 sobre o alrededor de 0.5, en general está compuesto por pares con valores de r^2 menores o iguales a 0.1, considerados poco informativos (Ver **figura 6.a y 6.b**).

En el caso de SUR, se identificaron 9 bloques (Ver **Figura 6.c y 6.d**), dos de los cuales incluyen SNPs de *SFTPD*: bloque 3 y bloque 4 (D' multilocus =0.87).

El bloque 3 incluye únicamente al SNP rs1051246 del gen *SFTPD* y a rs17883671 (no asociado a gen alguno). Este bloque se relaciona moderadamente con el bloque 2 (D' multilocus =0.6), asociado a su vez con el bloque 1 (D' multilocus =0.97). Este conjunto de bloques comprenden los SNPs del *BMS1P21* encontrados en Puno: rs1570328 (bloque 2), rs2758554 (bloque 2) y rs7072441 (bloque 1).

El bloque 4 incluye dos SNPs del gen *SFTPD* (rs76239460 y rs12247705, $r^2=0.025$), tres del gen TMEM254-AS1 (rs12242565, rs17098169 y rs61859013), uno de TMEM254 (rs1046707) y tres variantes no asociadas a ningún gen (rs10788345, rs1932571 y rs2342606). El bloque 3 posee una relación más fuerte con este bloque (D' multilocus =0.87) que con el bloque 2.

Respecto a las poblaciones de 1000 Genomes, en PEL se identificaron 11 bloques (Ver **Figura 6.e**). El bloque 3 y el bloque 4 presentaron SNPs de *SFTPD*, y una notable relación entre sí (D' multilocus =0.87).

El bloque 3 de PEL presentó los mismos SNPs de *SFTPD* que el bloque 4 de Puno (rs1051246, rs3088308 y rs911887), dos de los tres SNPs de *BMS1P21* (rs2758555, rs2758554) del mismo bloque y el mismo SNP de *LOC105371611* (rs7535771), además de una variante no asociada a gen alguno (rs17883671). Sus valores internos de r^2 fueron semejantes a los del bloque 4 de Puno. Este bloque presenta una fuerte relación con el bloque 2 (D' multilocus=0.98), compuesto por los mismos SNPs de *BMS1P21* del bloque 2 de Puno y por cuatro variantes no pertenecientes a ningún gen (rs61862196, rs2244701, rs2493719 y rs12261801). El bloque 4 de PEL está compuesto únicamente por tres SNPs de *SFTPD* (rs2243639, rs721917 y rs1885550), dos de los cuales son SNPs exónicos asociados a cambio aminoacídico. Sus valores internos de r^2 son bajos o intermedios, estando entre 0.2 y 0.63. Este bloque presenta una fuerte relación con el bloque 2 (D' multilocus =1), compuesto por dos SNPs de *BMS1P21* (rs7072441, rs41298231) y una variante no asociada a gen (rs12261801).

MXL presento también 11 bloques (Ver **Figura 6.f**), dos de los cuales presentan SNPs de *SFTPD*: bloque 4 y bloque 5 (D' multilocus =0.88).

El bloque 4 de MXL presentó los mismos SNPs de *SFTPD* que el bloque 4 de Puno y que el bloque 3 de PEL (rs1051246, rs3088308 y rs911887), dos SNPs de *BMS1P21* (rs2758554 y rs17879395), dos variantes no asociadas a gen alguno (rs17886252 y rs17883671) y el mismo SNP de *LOC105371611* del bloque 4 de Puno (rs7535771). Los valores de r^2 de los polimorfismos de *SFTPD* involucrados en este bloque son bajos o medios, entre 0.09 y 0.64. Este bloque presenta una fuerte asociación con el bloque 3 (D' multilocus=0.95), que incluye únicamente los dos SNPs del pseudogen *BMS1P21* (rs1570328, rs2758555), pertenecientes también a bloques ya mencionados en otras poblaciones. El bloque 5 posee los mismos SNPs que el bloque 4 de PEL, presentando valores de r^2 entre 0.31 y 0.59.

En ninguna de las poblaciones amerindias se encontró un bloque haplotípico extenso con altos valores de r^2 en la región del *SFTPD*, el cual habría sido esperable en caso de que alguna fuerza selectiva estuviese actuando en la región. Los bloques en las poblaciones americanas comparten muchos SNPs y asociaciones, siendo muy semejantes entre sí.

En CEU se encontraron 14 bloques (Ver **Figura 6.g**). Los bloques 5, 6 y 7 están conformados casi exclusivamente por SNPs de *SFTPD*. El valor de D' multilocus entre los bloques 5 y 6 es 1.0, y entre los bloques 6 y 7 es 0.76. EL bloque 5 está compuesto por rs1051246 y rs3088308 ($r^2=0.44$), y por la variante rs17883671 (sin asociación a gen alguno). El bloque 6 incluye solamente los SNPs de *SFTPD* rs911887 y rs2243639 ($r^2=0.4$), y el bloque 7 lo conforman los SNPs rs721917, rs12219080 y rs1885550. Los valores de r^2 al interior del bloque 7 son bajos, entre 0.07 y 0.32. Varios de los SNPs de *BMS1P21* descritos en las otras poblaciones se encuentran en bloques haplotípicos cercanos de poca extensión (Ej. rs1570328, rs278555 y rs2758554 en el bloque 4), pero no se encontraron bloques mixtos con SNPs de *SFTPD* y de otros genes.

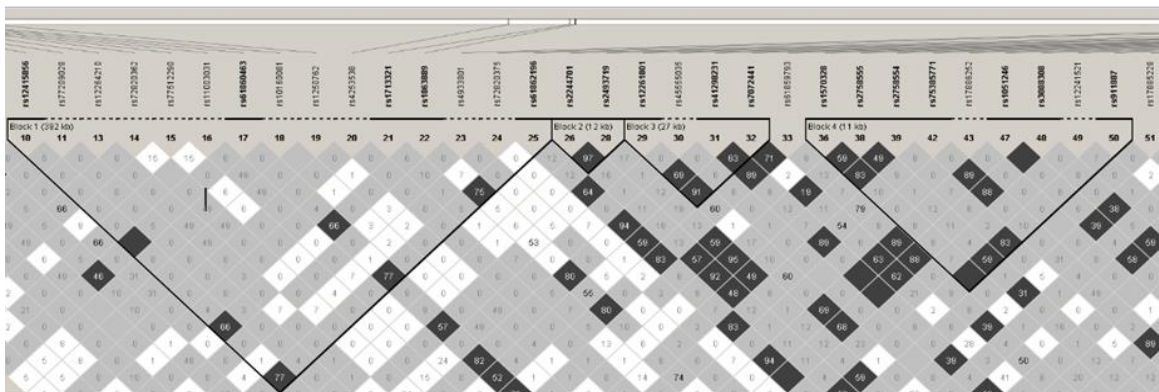
En IBS se encontraron 15 bloques (Ver **Figura 6.h**), de los cuales sólo dos pequeños bloques incluyen SNPs de *SFTPD*: el bloque 6 y el bloque 7 (D' multilocus=0.57). EL bloque 6 posee los mismos SNPs de *SFTPD* que el bloque 5 de CEU ($r^2=0.47$) y la misma variante no asociada a gen alguno. El bloque 7 sólo incluye los SNPs de *SFTPD* rs2243639 y rs721917 ($r^2=0.46$). Al igual que en el caso de CEU, se encontraron SNPs de *BMS1P21* descritos previamente en bloques haplotípicos cercanos de reducido tamaño (Ej. rs1570328, rs278555 y rs2758554 en el bloque 5), pero no conforman bloques con SNPs de *SFTPD*.

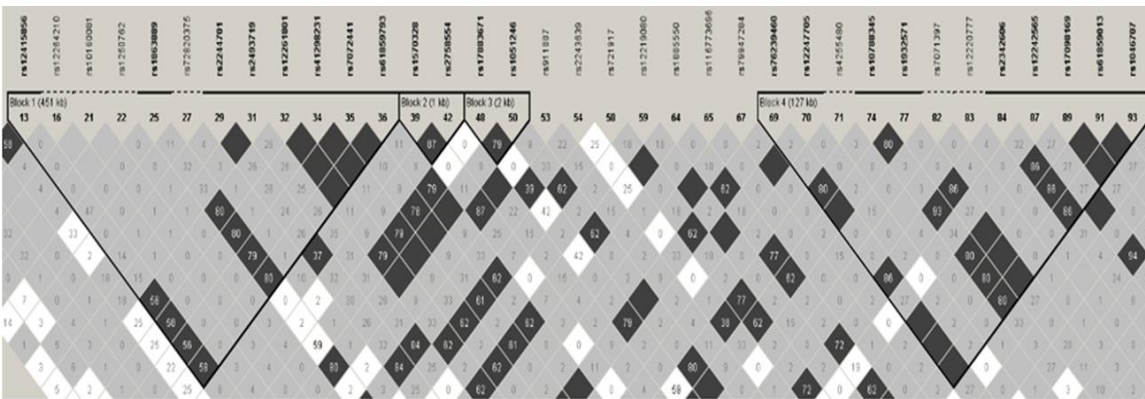
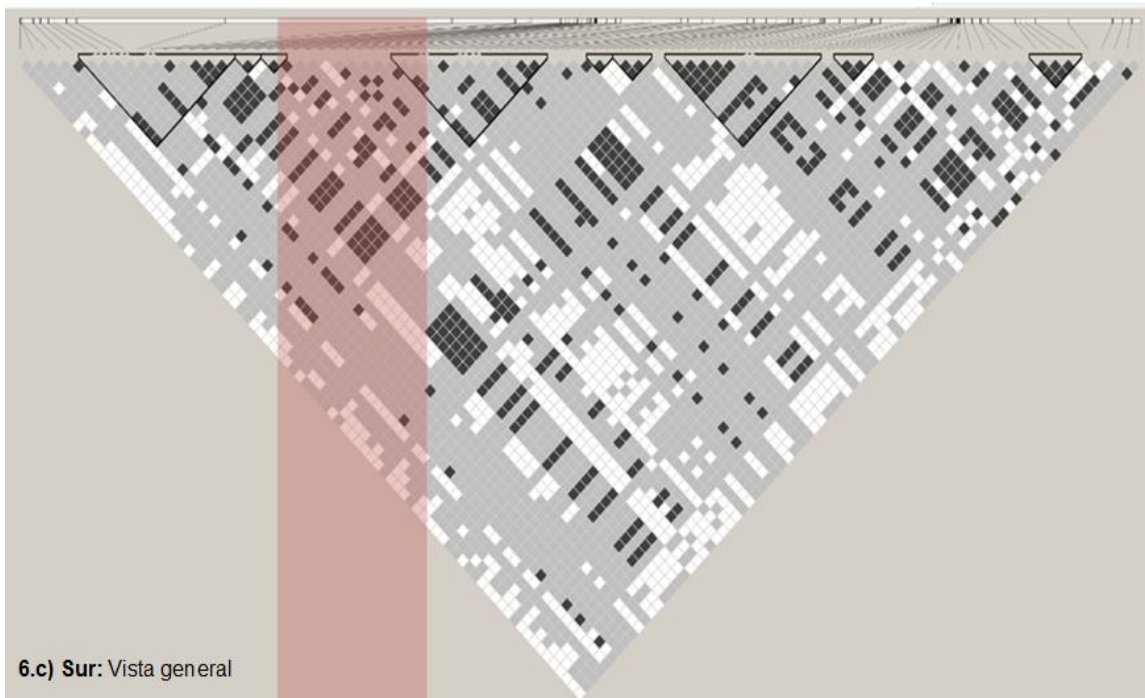
En el caso de las poblaciones asiáticas, los bloques haplotípicos encontrados fueron más extensos y con valores internos de r^2 más altos (Ver **Figuras 6.i y 6.j**).

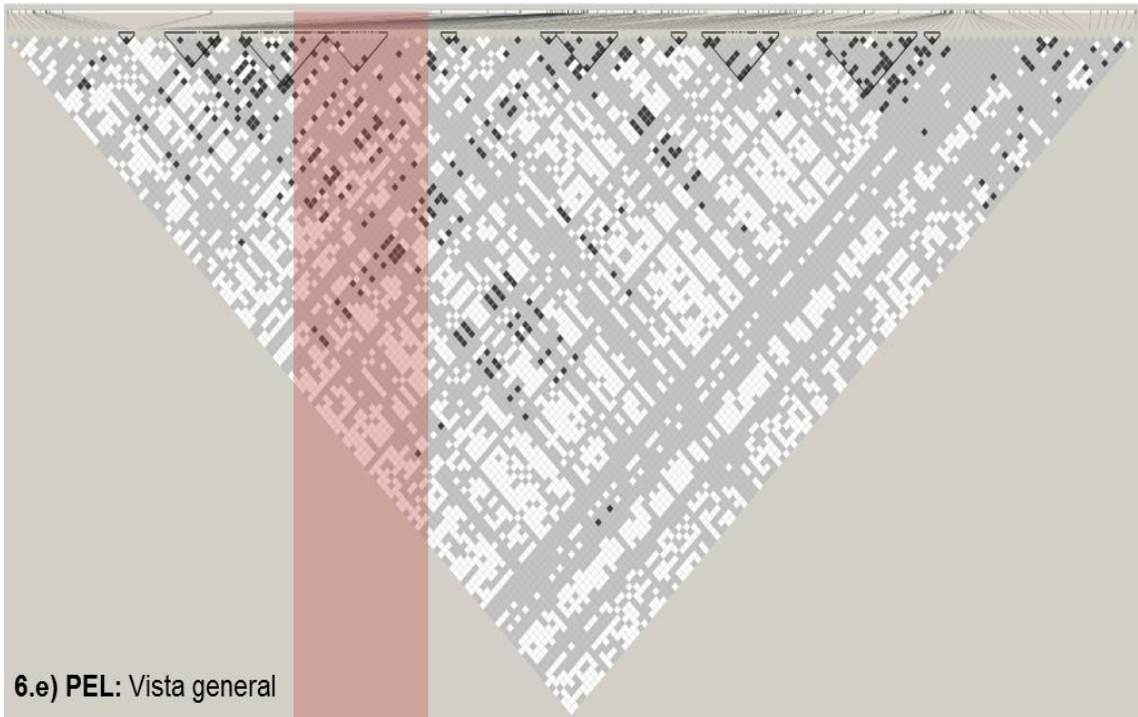
CHS y JPT presentaron 14 bloques cada uno. En CHS se encontraron dos bloques que incluyen SNPs de *SFTPD*: bloque 4 y bloque 5 (D' multilocus=0.82). El bloque 4 (58kb) incluye dos SNPs de *SFTPD* (rs911887 y rs2243639, $r^2=0.13$), tres SNPs de *BMS1P21* (rs2758554, rs278555 y rs1570328) y 2 variantes sin asociación a gen alguno (rs2493719 y rs22444701). El bloque 5 (48kb) está compuesto por 6 SNPs de *SFTPD* (rs721917, rs12219080, rs1885550, rs79947284, rs12247705, rs4255480) y una variante sin asociación a gen (rs10788345). Los valores de r^2 de este bloque en su mayoría están sobre 0.5 y en el caso de rs79947284-rs12247705 y rs79947283-rs1885550 llegan a 1. En JPT también se encontraron dos bloques que incluyen SNPs de *SFTPD*: bloque 3 y bloque 4, con un D' multilocus de 0.89. Estos bloques contienen los mismos SNPs y son de la misma extensión que los bloques 4 y 5 de CHS respectivamente, presentando valores de r^2 similares a los de sus contrapartes en CHS.

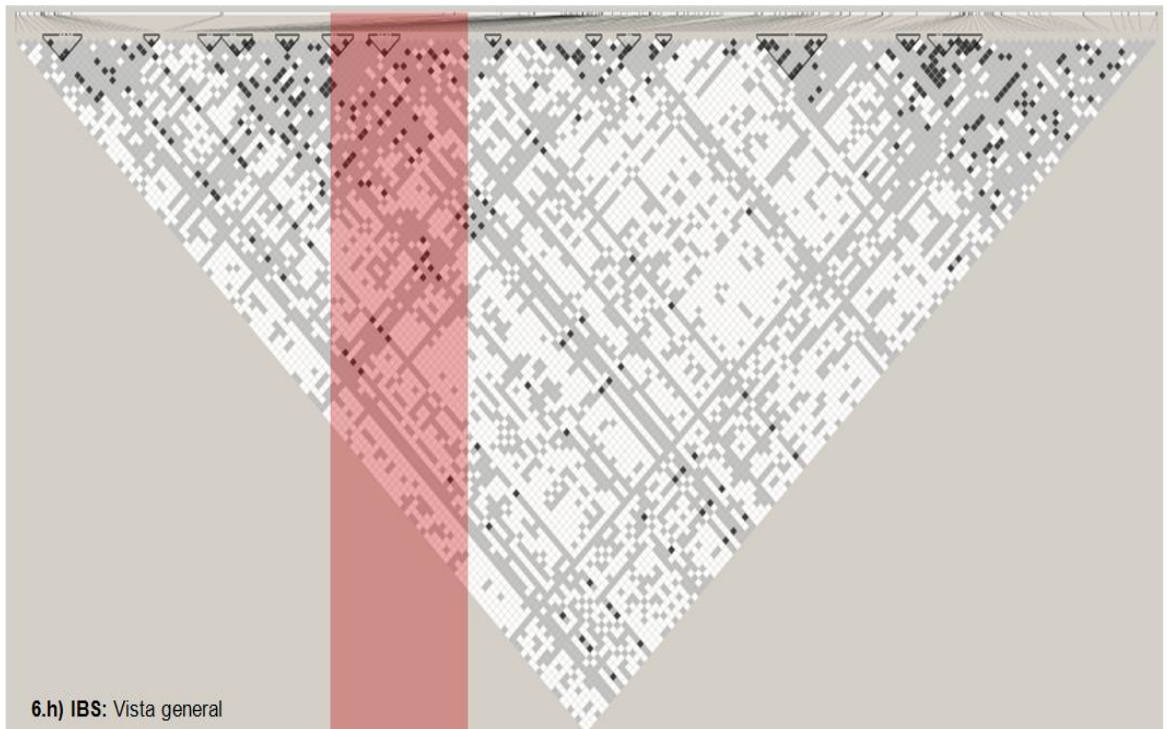
Finalmente, LWK presentó 11 bloques de poca extensión. De estos, el bloque 1 y el bloque 2 están compuestos únicamente por SNPs de *SFTPD* (D' multilocus=0.59). El bloque 1 lo conforman los SNPs rs1051246, rs12241521 y rs911887, mientras que el

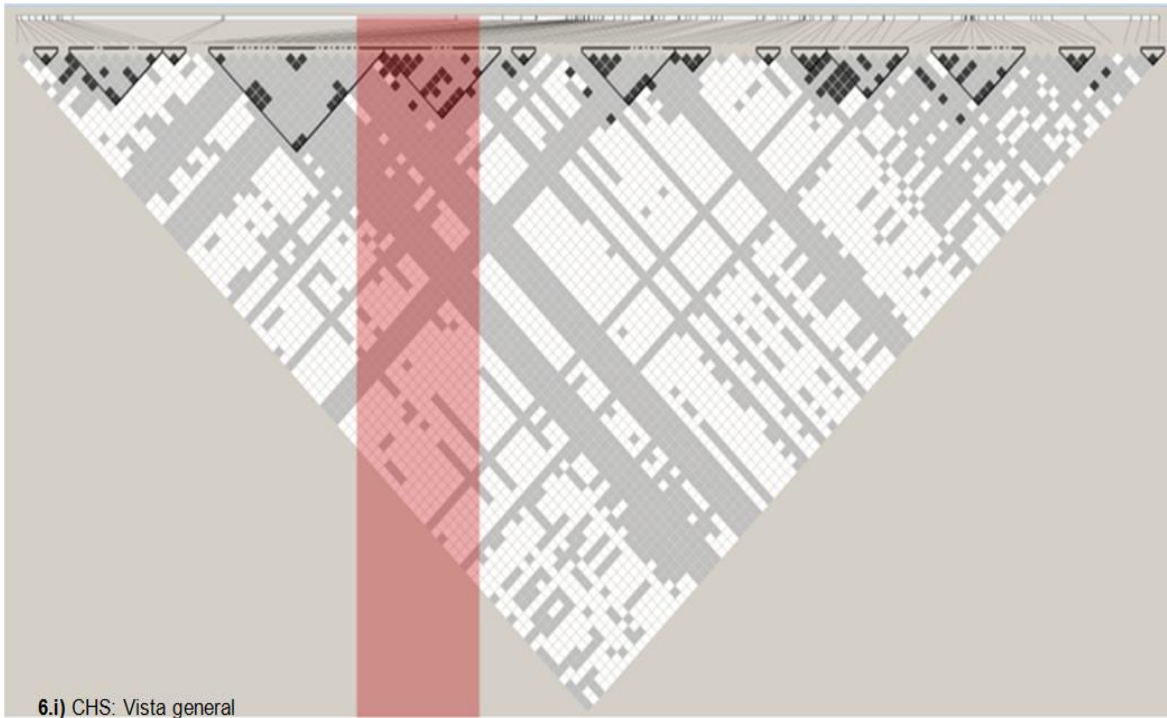
bloque 2 lo conforman siete SNPs (rs12219080, rs114811126, rs75904699, rs1885550, rs72822457, rs2264953 y rs12247705). Aunque este último bloque presenta la mayor cantidad de SNPs de *SFTPD*, sus valores internos de r^2 son más bajos que los de los bloques asiáticos, representados por una mayor cantidad de cuadrados blancos y grises claros en el esquema (Ver **Figura 6.k**). A diferencia de otras poblaciones, en este caso los SNPs de *BMS1P21* no forman bloques ni individualmente ni en conjunto con polimorfismos de *SFTPD*.



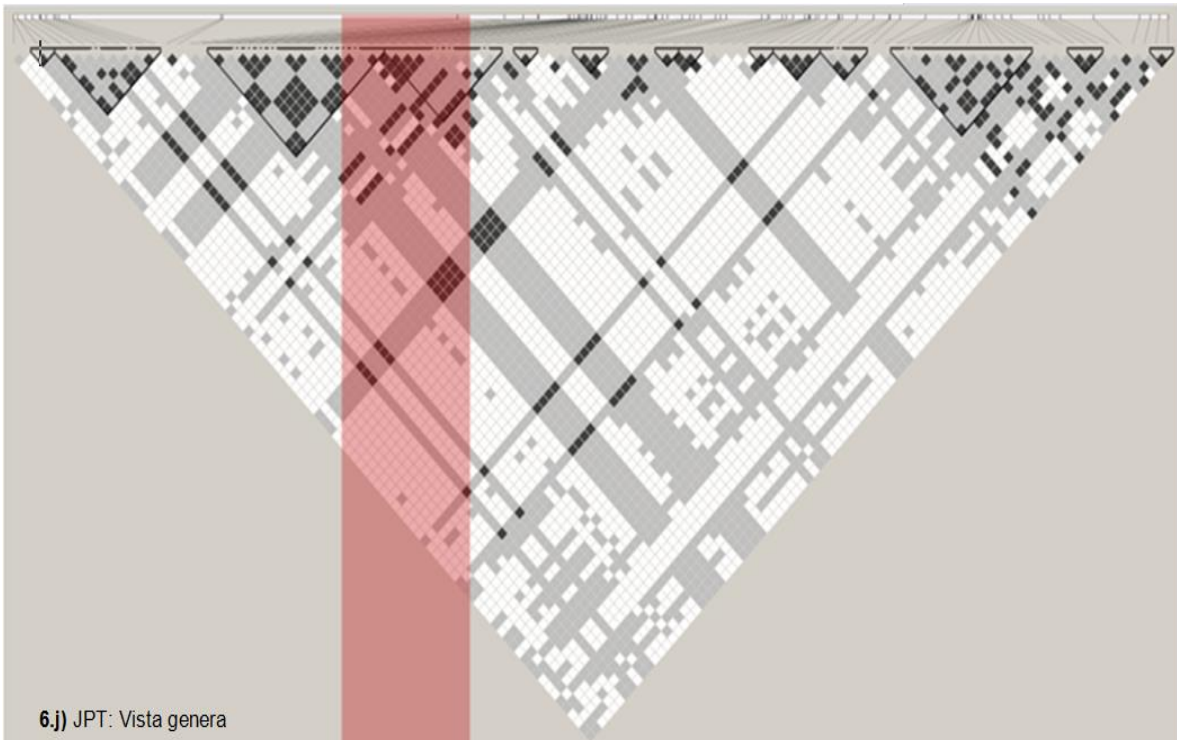




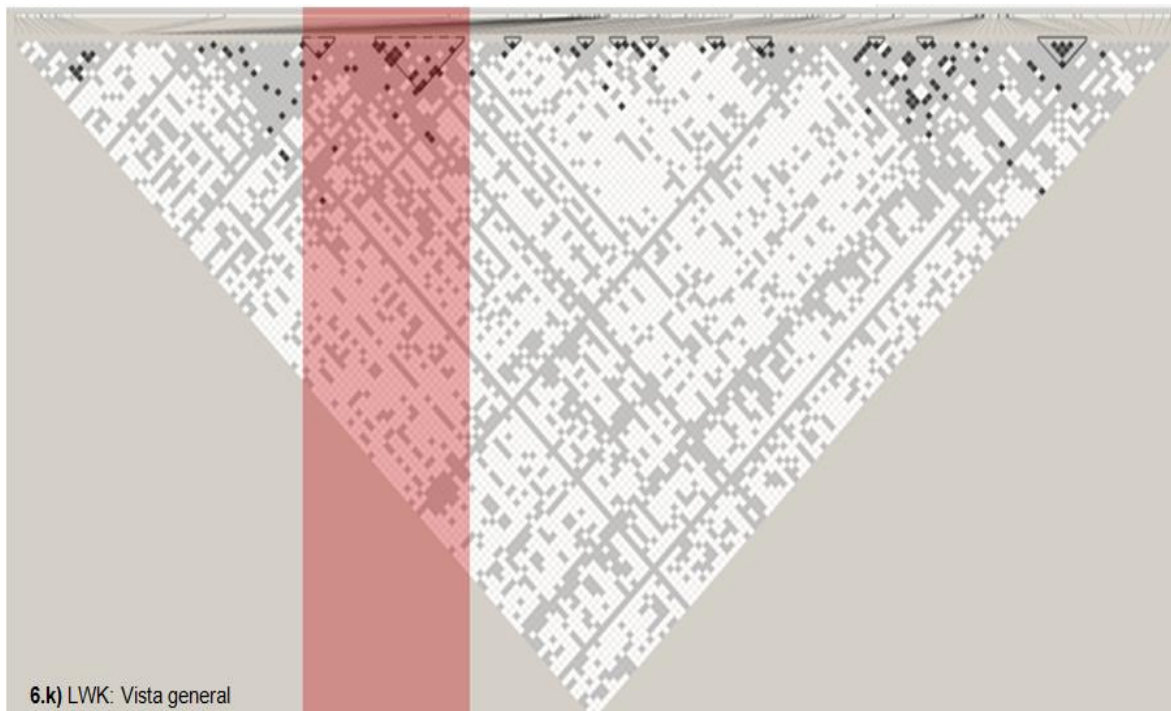




6.i) CHS: Vista general



6.j) JPT: Vista genera



Figuras 6.a – 6k: LD-plots que describen el desequilibrio de ligamiento al interior de la región genómica estudiada y con respecto a las regiones circundantes para cada población.

Rojo= región gen SFTPD, números: valores r^2 , cuadrado negro= desequilibrio de ligamiento fuerte, cuadrado gris=no informativo, cuadrado blanco=evidencia de recombinación, línea punteada=SNP excluido del bloque

5.e Valores de Fst y Neighbour-Joining

Los resultados de los análisis de Fst para los SNPs de *SFTPD* y para la región putativamente neutral se resumen en las **Tabla 10 y 11** respectivamente. El SNP rs 2243629 no fue utilizado en estos análisis debido a su ausencia en la muestra de Puno. Para reducir el número de falsos positivos y disminuir las probabilidades de error tipo I, los p-valores obtenidos en el análisis fueron corregidos mediante FDR³(Benjamini&Hochberg,1995).

Los valores de Fst obtenidos de los SNPs de *SFTPD* son significativos al 95% de nivel de confianza en todos los casos, excepto entre poblaciones que son muy próximas geográficamente y/o que son cercanas en términos de ancestría continental (PELxPuno, IBSxCEU y JPTxCHS) y entre MXLxSUR. En la región putativamente neutral no fueron significativos los valores de Fst obtenidos de nueve parejas: PELxPuno, IBSxCEU, MXLxPuno, MXLxPEL, MXLxCHS,JPTxPuno,JPTxPEL,JPTxMXL y JPTxSUR.

Respecto a los valores de Fst obtenidos de los SNPs de *SFTPD*, en general estos indican escasa o nula diferenciación genética, con valores de Fst ente 0 y 0.05 o bordeando el límite superior de este intervalo. La excepción a esto son las poblaciones

³ “false discovery rate” por su sigla en inglés.

asiáticas. CHS y JPT presentan una gran diferenciación genética respecto al resto de las poblaciones, presentando en todas las comparaciones valores de F_{st} mayores a 0.15, a excepción de cuando se las compara entre sí ($F_{st}=0.006$). El valor más alto de F_{st} se da al comparar JPT y Puno ($F_{st}=0.27$).

Puno presenta poca o ninguna diferenciación respecto a las otras poblaciones, con valores de F_{st} bajo 0.05, exceptuando a las poblaciones asiáticas y a LWK ($F_{st}=0.11$), con quien presenta una diferenciación moderada. Al comparar Puno con MXL, el valor de F_{st} también es bajo ($F_{st}=0.056$), pero ligeramente más alto que si se compara la muestra de Puno con las otras poblaciones americanas y europeas. Se obtuvieron valores de F_{st} equivalente a diferenciación genética moderada (entre 0.05 y 0.15) al contrastar SUR con las poblaciones europeas, y al comparar LWK con PEL, MXL y CHS.

Respecto a los valores de F_{st} obtenidos de la región putativamente neutral del cromosoma 10, LWK presenta los valores más altos al compararla con las otras poblaciones. Los valores de F_{st} entre LWK y las poblaciones americanas del cono Sur (Puno, PEL y SUR), están sobre 0.15, indicando gran diferenciación genética. En el resto de los casos, el valor de F_{st} indica una diferenciación más bien moderada.

Los valores de F_{st} obtenidos de la comparación de las poblaciones Americanas entre sí (MXL, Puno, PEL y SUR) y de la comparación entre poblaciones Americanas y Asiáticas están bajo 0.05, indicando escasa o nula diferenciación, a excepción de la comparación entre SUR y CHS. Los valores de F_{st} también son inferiores a 0.05 entre las poblaciones europeas y al comparar estas con MXL y CHS.

Tabla 10⁴: Valores de F_{st} entre poblaciones y p-valor correspondiente (SNPs *SFTPD*).

FDR p-valores: valores de p-value de F_{st} ajustados con FDR

Fst	FDR p-valores								
	Puno	SUR	IBS	PEL	CEU	MXL	CHS	JPT	LWK
Puno	*	0.0205560000	0.0063900000	0.9889100000	0.0104400000	0.0059890909	0.0000000000	0.0000000000	0.0000000000
SUR	0.04495	*	0.0004000000	0.0320748387	0.0010114286	0.0845672727	0.0000000000	0.0000000000	0.0063900000
IBS	0.03551	0.06774	*	0.0104400000	0.9672685714	0.0102876923	0.0000000000	0.0000000000	0.0009000000
PEL	-0.01114	0.03201	0.02845	*	0.0120413793	0.0095472000	0.0000000000	0.0000000000	0.0000000000
CEU	0.04191	0.07549	-0.00933	0.03562	*	0.0450000000	0.0000000000	0.0000000000	0.0007578947
MXL	0.05690	0.01877	0.03295	0.04593	0.02482	*	0.0000000000	0.0000000000	0.0004000000
CHS	0.21090	0.17255	0.15147	0.19033	0.17186	0.20303	*	0.2057717647	0.0000000000
JPT	0.27349	0.20224	0.20343	0.24918	0.22139	0.23186	0.00619	*	0.0000000000
LWK	0.11248	0.04636	0.05030	0.09796	0.05948	0.06901	0.12236	0.14739	*

Tabla 11: Valores de F_{st} entre poblaciones y p-valor correspondiente (región neutral)

Fst	FDR p-valores								
	Puno	SUR	IBS	PEL	CEU	MXL	CHS	JPT	LWK
Puno	*	0.4665176471	0.0000000000	0.9490100000	0.0000000000	0.2398180645	0.0000000000	0.2296924138	0.0000000000
SUR	0.01617	*	0.0000000000	0.0102450000	0.0000000000	0.0044921739	0.0000000000	0.0732666667	0.0000000000
IBS	0.09378	0.14104	*	0.0000000000	0.5275748571	0.0111168000	0.0027490909	0.0002000000	0.0000000000
PEL	-0.00954	0.02679	0.07962	*	0.0000000000	0.2806650000	0.0010620000	0.2398180645	0.0000000000
CEU	0.12044	0.16409	-0.00128	0.10351	*	0.0005684211	0.0018685714	0.0000000000	0.0000000000
MXL	0.00643	0.03612	0.03262	0.00445	0.04925	*	0.1602514286	0.4622400000	0.0000000000
CHS	0.04862	0.08743	0.03630	0.04605	0.03990	0.00902	*	0.0481153846	0.0000000000
JPT	0.00488	0.01261	0.06491	0.00399	0.07989	0.00033	0.01515	*	0.0000000000
LWK	0.17844	0.24491	0.08226	0.17319	0.12645	0.11508	0.11953	0.15200	*

⁴ Los valores menores a 1 son resultado del algoritmo utilizado por *Arlequin* para estimar valores de F_{st} deben ser interpretados como "0".

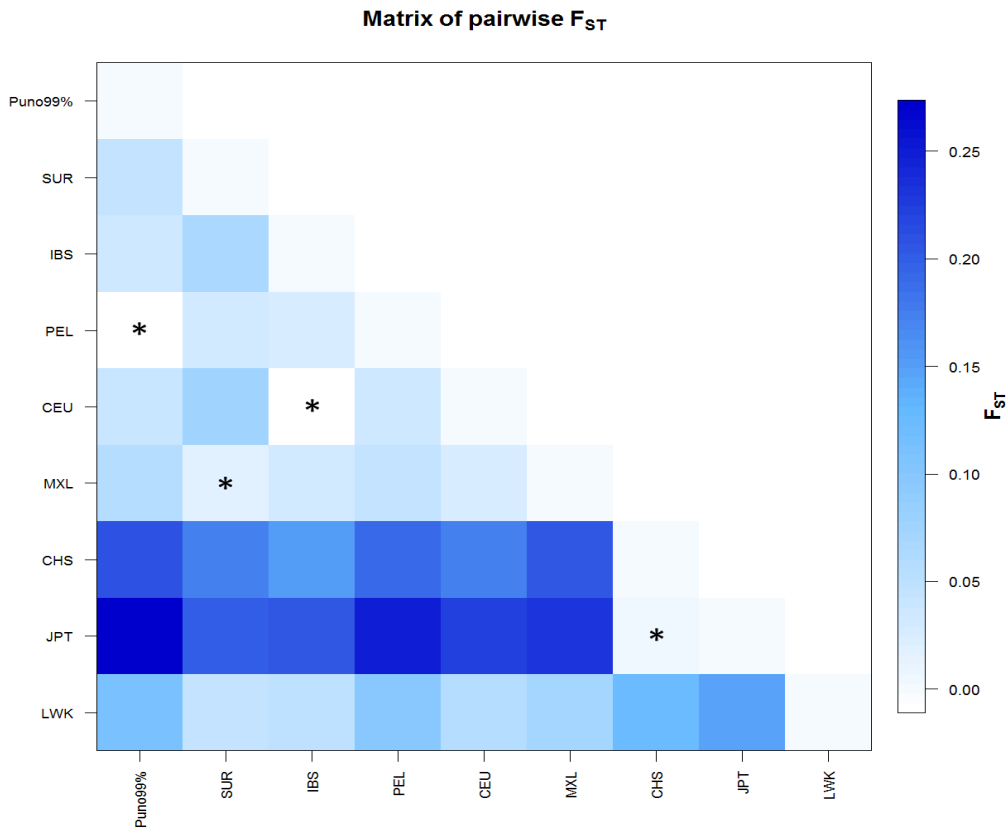


Figura 7.a: Valores de F_{ST} entre poblaciones (SNPs SFTPD). * =valores no significativos

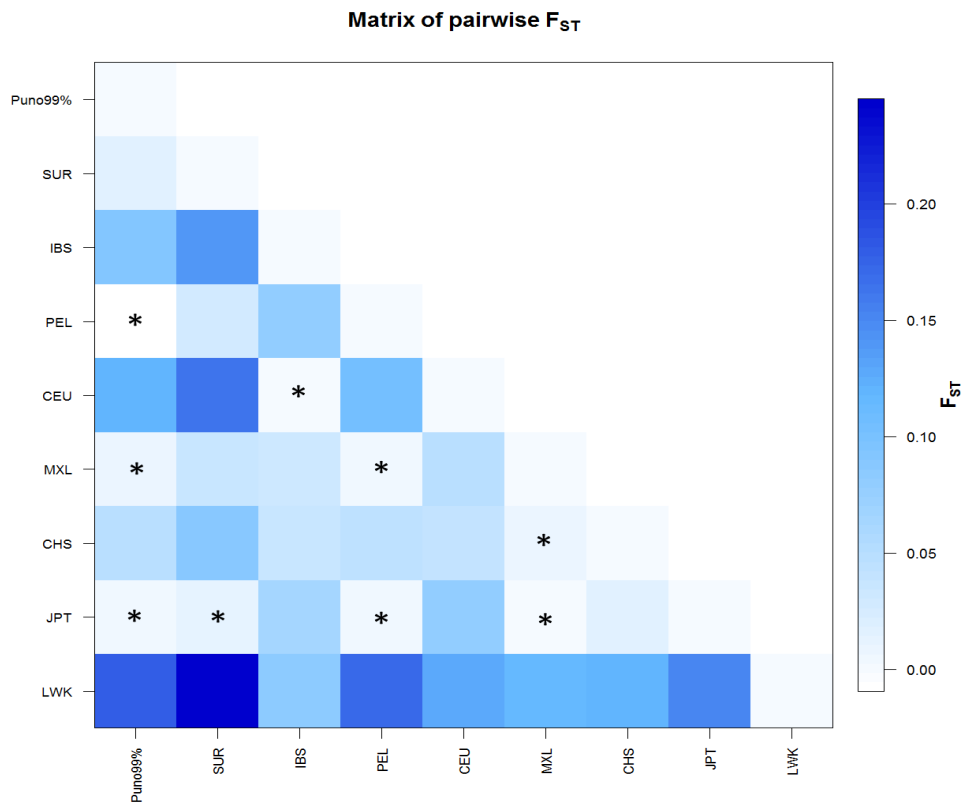


Figura 7.b: Valores de F_{ST} entre poblaciones (región neutral). * =valores no significativos

A partir de la matriz de valores de F_{st} de los SNPs de *SFTPD* se construyó un dendrograma de Neighbor-Joining utilizando el programa *MEGA 7*, con el fin de visualizar las distancias genéticas (**Ver Figura 8**).

En el dendrograma se observa claramente que las poblaciones asiáticas (estrechamente relacionadas) se separan de las demás conformando un grupo aparte. El resto de las poblaciones se posicionan a una distancia variable dentro de un mismo grupo, de forma relativamente congruente con su distribución geográfica y antecedentes de mestizaje.

PEL y Puno son las poblaciones geográficamente más próximas y presentan una estrecha cercanía. CEU e IBS también se agrupan juntas a una distancia cercana de PEL y Puno, probablemente debido a la presencia de sujetos mestizos en la muestra de PEL. Este factor también explicaría la proximidad de MXL a las poblaciones europeas. SUR se ubica más distante, posiblemente debido un mayor aislamiento y menor mestizaje con poblaciones europeas en comparación con las otras poblaciones americanas en los grupos Huilliche y Pehuenche que la conforman.

Finalmente, a una distancia notablemente mayor pero dentro de la misma rama se localiza LWK.

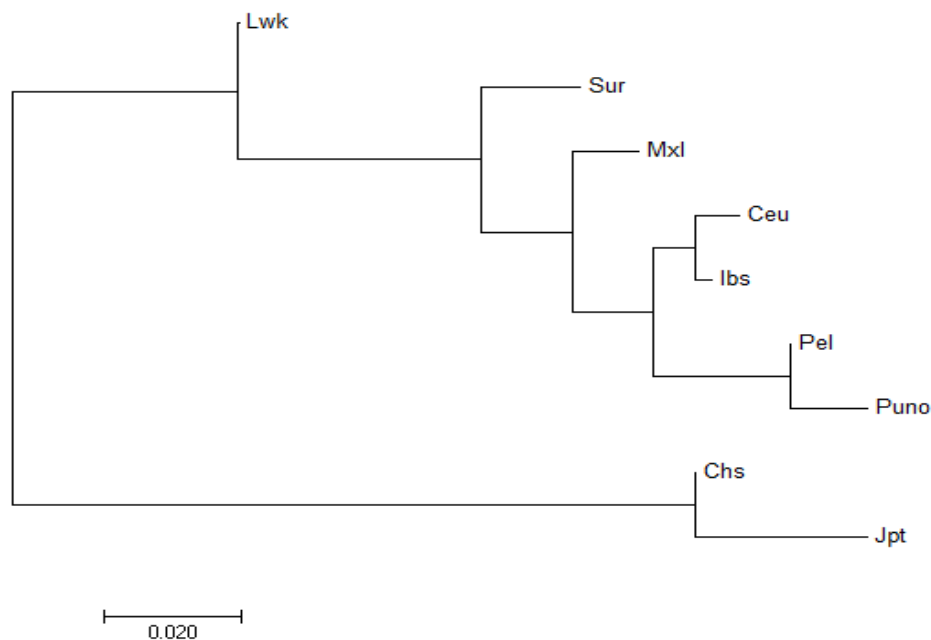


Figura 8: Dendrograma de Neighbor-Joining construido a partir de las frecuencias alélicas de los SNPs de *SFTPD*.

Adicionalmente, se evaluaron los niveles de diferenciación genética entre las poblaciones para los distintos SNPs y su desviación respecto a lo esperado bajo un modelo neutral mediante análisis de F_{st} Outlier, utilizando el software LOSITAN (Antao et al., 2008). Este método permite detectar *loci* con niveles de diferenciación genética inusualmente altos o bajos al contrastarlos con un modelo de distribución neutral generado mediante simulaciones (Beaumont & Nichols, 1996). Si un locus estuviese

siendo afectado por selección balanceadora, se esperaría un incremento de la heterocigosidad en las cercanías del *locus*. Por el contrario, se esperaría una disminución local de la heterocigosidad si el *locus* estuviese bajo selección positiva (Nielsen & Slatkin, 2013)

Los análisis de Fst Outlier se realizaron con los SNPs de *SFTPD*, utilizando inicialmente 50.000 simulaciones (50ksims) a un intervalo de confianza del 95%, con un FDR de 5%. En los casos en los que se detectaron outliers se incrementó el número de simulaciones para verificar confiabilidad del análisis. Los resultados se resumen en las **Figuras 9 y 10a-b**.

En el análisis general construido utilizando todas las poblaciones (**Ver Figura 9**) se encontraron 2 outliers (LOCUS 5 y LOCUS 11), que corresponden a los SNPs rs911887 y rs114811126 respectivamente. El primero fue detectado como un *locus* candidato para selección balanceadora, mientras que el segundo aparece como candidato para selección positiva. Sin embargo, ninguno de los dos SNPs fue clasificado como outlier mediante el método de FDR incorporado en el análisis, y rs114811126 pierde esta categoría si se aumenta el número de simulaciones. Además, dado que el alelo "A" de rs114811126 está fijado en todas las poblaciones excepto en LWK (G=15%, A=85%), es posible que la calidad de outlier de este SNP esté dada más bien por su MAF ("Minor Allele Frequency") inferior a 0.05 en todo el mundo excepto en África. En el caso de rs911887, si este realmente estuviese bajo selección balanceadora, entonces este SNP debería aparecer como outlier en las comparaciones entre pares de poblaciones y debiese observarse un aumento local de la heterocigosidad debido a la cercanía de los *loci* entre sí, cosa que no sucede. Por estos motivos, los resultados de este análisis deben ser tomados con precaución.

Al contrastar Puno con las otras ocho poblaciones (**Ver ANEXO 2**), sólo se encontraron outliers en PunoxCHS y PunoxLWK, clasificados como candidatos a selección positiva (**Ver Figuras 10a y 10b**). En el caso de PunoxCHS se observaron 3 outliers consistentes al correr 55ksims o más: LOCUS 10, LOCUS 18 y LOCUS 21. Los mismos *loci* aparecen como outliers en la comparación PunoxJPT a partir de las 65ksims, aunque este resultado es sensible al número de simulaciones. Estos *loci* corresponden a los SNPs rs12219080, rs79947284 y rs12247705, los cuales se encuentran al interior del bloque haplotípico 5 de CHS (**Ver Figura 6.i**), presentando altos valores de desequilibrio de ligamiento. Si bien la frecuencia del alelo "T" (alelo menor a nivel mundial) de estos tres SNPs es inusualmente alta en las poblaciones asiáticas si se las compara con el resto (**Ver Tabla 12**), su clasificación como outlier debe ser tomada con precaución, ya que la ausencia de este alelo en Puno puede estar sesgando el análisis. Finalmente, al comparar PunoxLWK se encontraron 6 outliers: LOCUS 3, LOCUS 6, LOCUS 8, LOCUS 13, LOCUS 14 y LOCUS 17. Estos corresponden a los SNPs rs3088308, rs17885228, rs17883490, rs75448566, rs75904699 y rs72822457. Al igual que en el caso del análisis general, la mayoría de estos *loci* son SNPs con frecuencias anómalas y alelos raros presentes únicamente en LWK (**Ver Tabla 12**), por lo se descarta que su calidad de outlier tenga alguna relación con una presión selectiva en Puno.

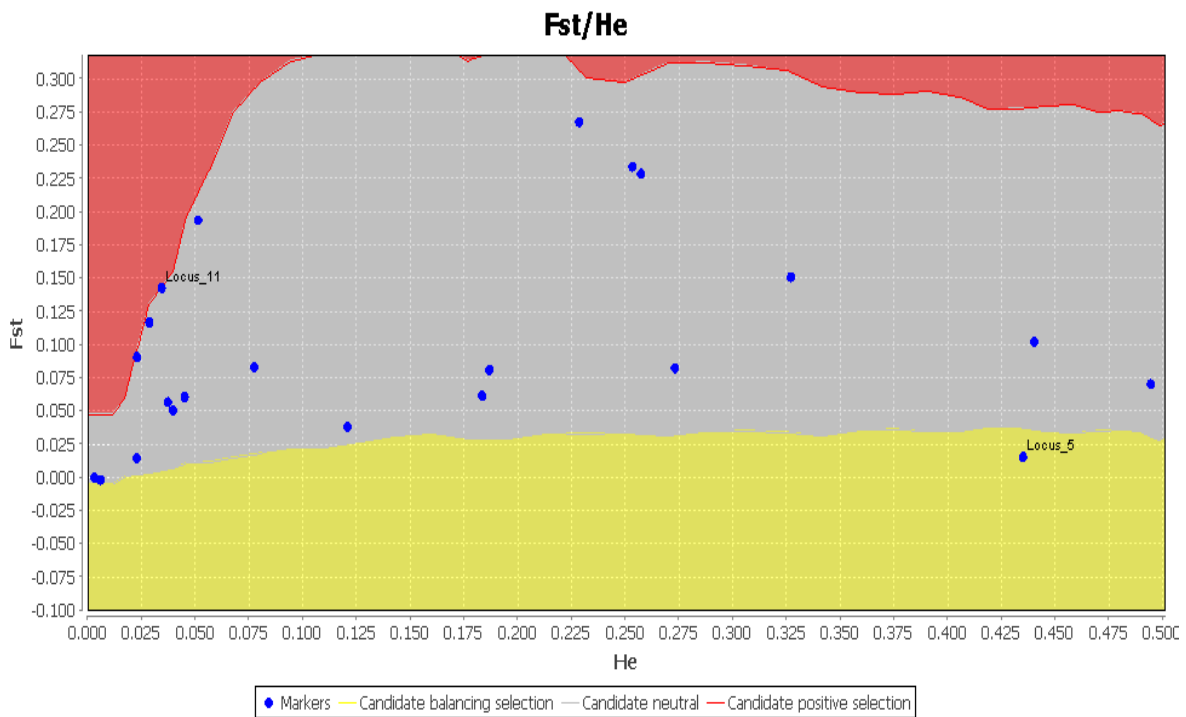


Figura 9: Gráfico de Fst vs Heterocigosidad utilizando como contraste una distribución neutral generada bajo un modelo genético simple (modelo de “migración de isla”). *Análisis construido utilizando las nueve poblaciones (50ksims)*. Área roja: candidatos a loci bajo selección positiva, área amarilla: candidatos a loci bajo selección balanceadora, área gris: loci neutrales.

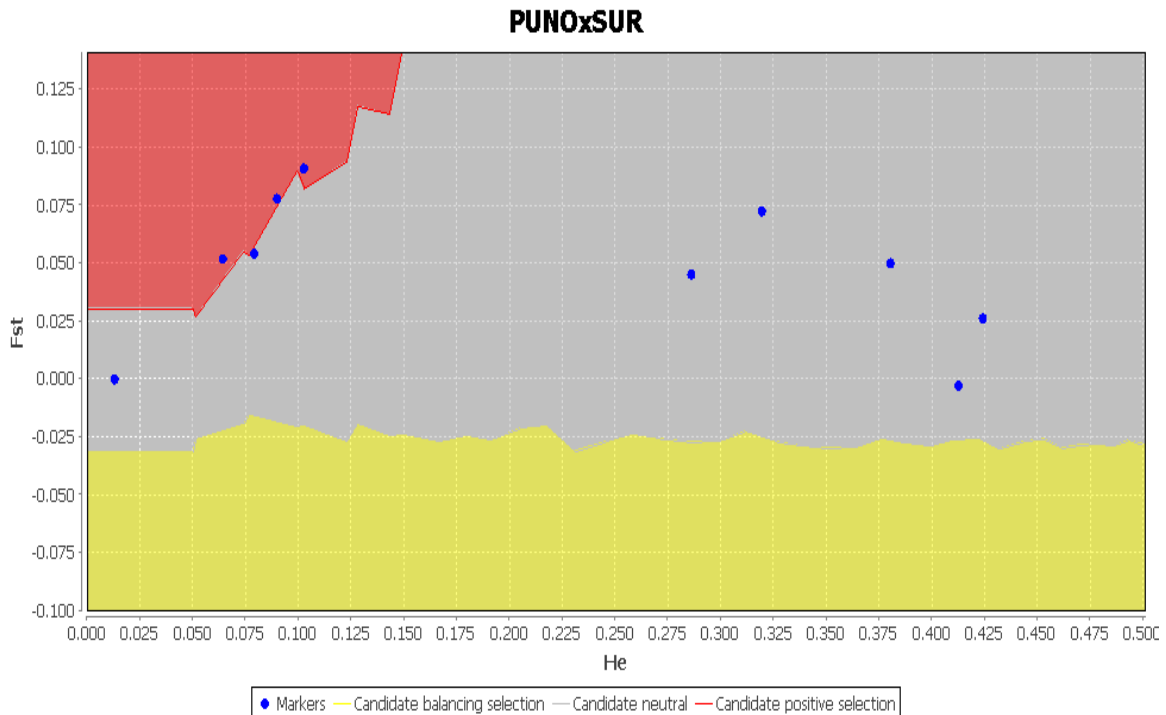


Figura 10a: PunoxSUR (60ksims)⁵.

⁵ Aunque el LOCUS 22 (rs4255480) presenta la cualidad de outlier a 50ksims, al aumentar el número de simulaciones pierde este estatus. Esto sumado a la escasa diferencia en la MAF de este SNP entre Puno y SUR (0.06 y 0 respectivamente), sugieren que su posición en el gráfico y su cualidad de outlier se deben más bien a un resultado artefactual debido al modelo de coalescencia.

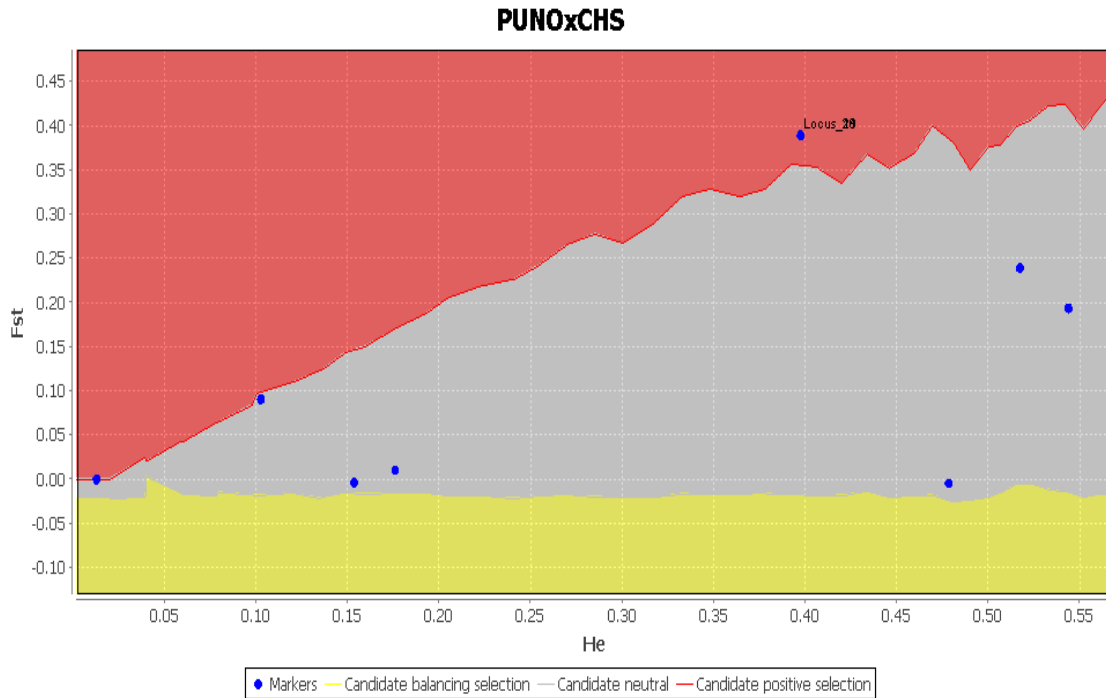


Figura 10.b: PunoxCHS (55ksims). Outliers: LOCUS 10 (rs12219080), LOCUS 18 (rs79947284) y LOCUS 21(rs12247705)

Figura 10.a-b: Gráficos de Fst vs Heterocigidad contrastando dos poblaciones

5.f Análisis de Componentes Principales

5.f.1 PCA : Frecuencias alélicas

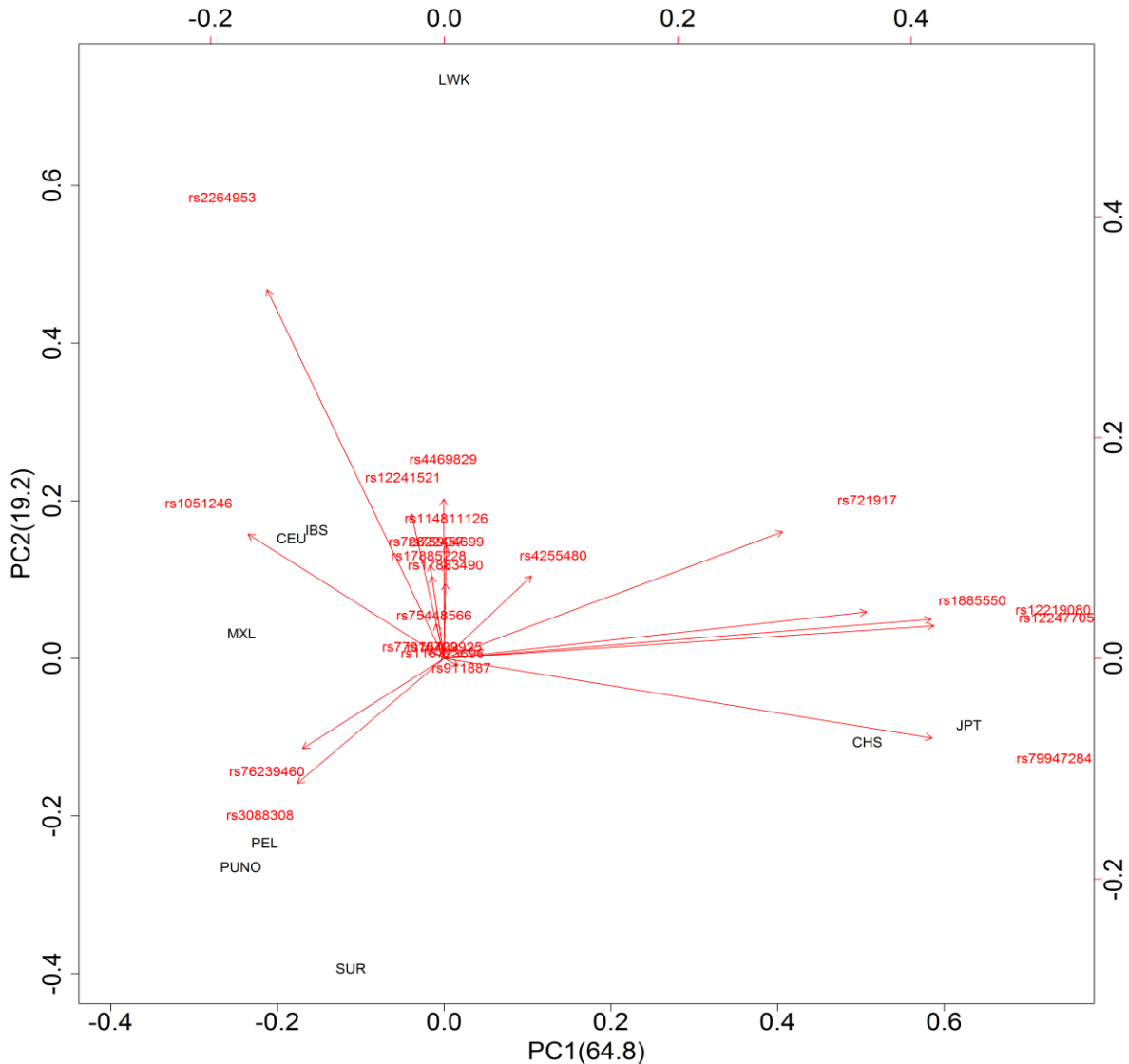
El primer PCA se construyó a partir de la frecuencia del alelo menor de los 22 SNPs de *SFTPD* presentes en todas las poblaciones y permitió identificar SNPs relacionados a los agrupamientos observados.

El componente principal 1 y el componente principal 2 (PC1 y PC2) explican en conjunto el 84,07% de la proporción de varianza (64,84% y 19,23% respectivamente), por lo que se utilizaron estos componentes para construir un gráfico que resume las distancias genéticas entre las poblaciones y que incluye coeficientes de correlación de los SNPs en relación a cada componente (**Ver Figura 11**).

En el primer componente, las poblaciones se distinguen claramente y se concentran principalmente hacia los valores negativos del eje X, mientras que en el segundo componente se agrupan hacia los valores centrales y negativos del eje Y. La población de LWK se ubica en el extremo superior del gráfico, alejada de todas las demás poblaciones en el PC2, pero cercana a las poblaciones europeas y americanas en PC1. Las poblaciones europeas se posicionan cerca y MXL se sitúa próxima a ellas, en un punto intermedio entre estas y las poblaciones americanas. Puno, PEL y Sur se agrupan juntas hacia los valores negativos de PC1. Finalmente, las poblaciones asiáticas se localizan cercanas entre sí hacia los valores positivos del PC1 y alejadas de todas las demás por la mayor de todas las distancias observadas. En el PC2 se sitúan más

próximas a las poblaciones americanas. Estos resultados son coherentes con lo observado en el dendrograma de Neighbor-Joining.

Figura 11: Análisis de Componentes Principales, Biplot PC1/PC2.



A mayor coeficiente de correlación entre el SNP y el componente, más largo el vector en la dirección del eje del componente.

Los coeficientes de correlación positiva más altos en PC1 son los que presenta con rs12247705 (coef=0.46), rs79947284 (coef=0.459), rs12219080 (coef=0.458), rs721917 (coef=0.318) y rs1885550 (coef 0.397). Estos SNPs conforman los bloques haplotípicos 5 (en CHS) y 4 (en JPT) previamente descritos (Ver **Figuras 6.i y 6.j**). Estos bloques son idénticos en su composición y están presentes sólo en las poblaciones asiáticas. Todos los SNPs de *SFTPD* de estos bloques presentaron diferencias significativas en sus frecuencias alélicas al comparar poblaciones asiáticas y amerindias utilizando la prueba exacta de Fisher (Ver **Tabla 7**). Se observa también que alelos menos frecuentes en las

poblaciones occidentales poseen frecuencias notablemente más altas en las poblaciones asiáticas (Ver **Tabla 12**).

Tabla 12: MAF de los SNPs que presentaron correlaciones importantes con los componentes. *Gris= SNPs de los bloques haplotípicos asiáticos*

Estos antecedentes señalan que la separación de las poblaciones asiáticas observada

SNPs	Poblaciones								
	Puno	SUR	IBS	PEL	CEU	MXL	CHS	JPT	LWK
rs12247705	0	0.08974	0.05128	0.02564	0.03846	0.02564	0.3974	0.4744	0.1923
rs79947284	0	0.08974	0.05128	0.02564	0.03846	0.02564	0.3974	0.4744	0.03846
rs12219080	0	0.08974	0.05128	0.03846	0.03846	0.02564	0.3974	0.4744	0.2051
rs721917	0.2564	0.3205	0.4487	0.2436	0.4744	0.3974	0.5897	0.6667	0.4615
rs1885550	0.1667	0.3205	0.2179	0.1795	0.1923	0.2692	0.5256	0.6026	0.3974
rs1051246	0.1026	0.2308	0.141	0.1282	0.1667	0.3462	0	0.03846	0.2949
rs3088308	0.1026	0.2727	0.07692	0.1026	0.1026	0.2308	0	0.03846	0
rs2264953	0.1026	0	0.359	0.1154	0.3846	0.2949	0.0641	0.0641	0.4231
rs76239460	0.1026	0.2308	0.07692	0.1026	0.1026	0.2308	0	0.03846	0.02564
rs4469829	0	0	0	0	0	0.01282	0	0	0.2179
rs12241521	0.01282	0	0.05128	0.02564	0.02564	0.05128	0	0	0.1923
rs114811126	0	0	0	0	0	0	0	0	0.1538
rs75904699	0	0	0	0	0	0	0	0	0.1282
rs72822457	0	0	0.0641	0.01282	0.01282	0	0	0	0.1154
rs4255480	0.0641	0	0.03846	0.05128	0.02564	0.01282	0.1282	0.08974	0.1667
rs17885228	0	0	0.05128	0.01282	0.01282	0	0	0	0.1026
rs17883490	0	0	0	0	0	0	0	0	0.1026

está dada por la presencia de uno o más haplotipos que involucran los alelos menos frecuentes (a nivel global) de los SNPs de *SFTPD* localizados en los bloques previamente descritos.

PC1 presenta correlación negativa con rs1051246 (coef=-0.184), rs3088308 (coef:-0.135), rs2264953 (coef=-0.167) y rs76239460 (coef=-0.13). Todos estos SNPs, a excepción de rs2264953, también presentaron diferencias significativas en sus frecuencias alélicas al comparar poblaciones asiáticas y amerindias utilizando la prueba .exacta de Fisher. A diferencia del resto de las poblaciones, la frecuencia del alelo menor de estos SNPs está bajo 0.05 en CHS y JPT (Ver **Tabla 12**). Las correlaciones del resto de los SNPs con el PC1 son casi nulas.

PC2 presenta correlación positiva principalmente con SNPs cuya MAF es más alta en LWK, tales como rs2264953 (coef=0.67), rs4469829 (coef=0.29), rs12241521 (coef=0.27), rs114811126 (coef=0.20), rs75904699 (coef=0.17), rs72822457 (coef=0.17), rs4255480 (coef=0.15), rs17885228 (coef=0.15) y rs17883490 (coef=0.14). Los SNPs rs721917 (coef=0.23) y rs1051246 (coef=0.22) también están correlacionados positivamente a este componente. PC2 se correlaciona negativamente con rs3088308 (coef=-0.22), rs79947284 (coef=-0.15) y rs7623946 (coef=-0.164). Estos SNPs también se correlacionan con el PC1, por lo que puede que señalen cercanía respecto a las poblaciones asiáticas. Los valores de los coeficientes de correlación del resto de los SNPs con el PC2 son cercanos a 0.

5. f.2 PCA: Frecuencias haplotípicas

En total se identificaron 28 haplotipos que incluyen los 23 SNPs de *SFTPD* (Ver Tabla 13), los cuales fueron nombrados según su número. Los haplotipos 18 a 28 están presentes sólo en LWK y en muy baja frecuencia (inferior al 8%). Esto se debe a que SNPs de *SFTPD* que en las otras poblaciones poseen un alelo fijado, en LWK presentan ambos alelos.

Debido a la falta del SNP rs2243639 en la población de Puno, para determinar cuáles de los haplotipos identificados a nivel global estaban presentes en esta muestra se recurrió a un método indirecto. Primero, se estimaron haplotipos mediante el algoritmo EM utilizando los 22 SNPs restantes, luego se observó a que haplotipos presentes en las otras poblaciones podían corresponder. Descontando el SNP faltante, el haplotipo más frecuente identificado en Puno (47.6%) posee la misma combinación de alelos que el “Haplotipo 1”, el cual es el más frecuente en SUR, PEL, CEU, IBS y MXL, y el segundo más frecuente en CHS y JPT. El segundo haplotipo identificado en Puno (21.8%) coincide en sus SNPs restantes con el “Haplotipo 3”, identificado en SUR, PEL, MXL, CHS y JPT. Siguiendo esta lógica, el tercer, cuarto, quinto y sexto haplotipo presentes en Puno corresponden a los “Haplotipos 2, 7, 12 y 10” identificados en las otras poblaciones. Aunque se trata de un método indirecto, la frecuencia de los haplotipos en las otras poblaciones (en especial aquellas cercanas a Puno como PEL y SUR) y la correspondencia exacta de la combinación de alelos de los 22 SNPs restantes fundamentan la asignación realizada en Puno.

Tabla 13: Haplotipos *SFTPD*: Combinaciones de alelos de los 23 SNPs de *SFTPD* encontrados en las poblaciones estudiadas. *El primer SNP de cada haplotipo corresponde a rs4469829 y el último a rs4255480.*

Nombre	Combinación de alelos
Haplotipo1	CAACTTCTGACATGACCTCTCCT
Haplotipo2	CGTCTCCTGGCATGATCTCTTCT
Haplotipo3	CAACCCCTGACATGACCTCTCCT
Haplotipo4	CAACTTCTGACATAACCTCTCCT
Haplotipo5	CAACTTCTGGTATGATCTTTCTT
Haplotipo6	CAACCTCTGACATGACCTCTCCT
Haplotipo7	CAACCCCTGGCATGACCTCCCCT
Haplotipo8	CGACTCCTGACATGACCTCCCCT
Haplotipo9	CAACTCCTGGTATGATCTTTCTT
Haplotipo10	CAAACCCCTGACATGACCTCCCCT
Haplotipo11	CAACCCCTGGCATGATCTCTCCT
Haplotipo12	CAACCCCTGGCATGATCTCTCCC
Haplotipo13	CAACCCCTGGCACGACCTCCCCT
Haplotipo14	CAACTCTTGACATGACCCCTCCT
Haplotipo15	CAACCCCTGACATGACCTCCCCT
Haplotipo16	CAACTCCTGACATGACCTCTCCT
Haplotipo17	CAACTCCTGGCATGATCTCTCCC
Haplotipo18	TGACTCCTGGTGTGGTCTCTCTT
Haplotipo19	CAACTCCTAGCATGACCTCCCCT
Haplotipo20	TGACTCCTGGCATGATCTCTCCC
Haplotipo21	CAACTCCTGGCATAATCTCTCCC
Haplotipo22	CAACTCCTGGTGTGATCTCTCTT
Haplotipo23	CAACTCCTAGCATGATCTCTTCT
Haplotipo24	CAACTCCTGGTGTGGTCTCTCTT
Haplotipo25	CAACTCCTGACATGACCTCCCCT
Haplotipo26	CAACTCCTAACATGACCTCCCCT
Haplotipo27	CAACCCCTGACATGACTCCTCCT
Haplotipo28	CAACTCCTGGTGTGGTCTCTCCT

A partir de las frecuencias haplotípicas en cada población (Ver **Tabla 14**), se realizó un segundo PCA, con el objetivo de identificar aquellos haplotipos que permitieran diferenciar a las poblaciones.

Tabla 14: Frecuencias Haplotípicas por población.

Haplotipo	PUNO	SUR	IBS	PEL	CEU	MXL	CHS	JPT	LWK
HAP_1	0.476	0.345	0.36	0.459	0.394	0.391	0.188	0.255	0.010
HAP_2	0.116	0.273	0.084	0.112	0.076	0.219	0	0.029	0
HAP_3	0.218	0.182	0	0.176	0	0.023	0.174	0.135	0
HAP_4	0	0.086	0	0	0	0	0	0	0
HAP_5	0	0.068	0	0	0	0	0	0	0
HAP_6	0	0.045	0	0	0	0	0	0	0
HAP_7	0.063	0	0.154	0.065	0.194	0.109	0.038	0.048	0
HAP_8	0	0	0.075	0.035	0.079	0.078	0	0	0.060
HAP_9	0	0	0.033	0.018	0.051	0.047	0.419	0.404	0.030
HAP_10	0.041	0	0.051	0.024	0.045	0.031	0	0	0.232
HAP_11	0	0	0.047	0	0.040	0.031	0	0	0
HAP_12	0.060	0	0.056	0.053	0.035	0.016	0.129	0.106	0.025
HAP_13	0	0	0.070	0.012	0.035	0.016	0	0	0
HAP_14	0	0	0.028	0	0.020	0	0	0	0.096
HAP_15	0	0	0.023	0	0	0.016	0	0	0.065
HAP_16	0	0	0	0.012	0	0	0.014	0	0.040
HAP_17	0	0	0	0	0	0	0.014	0	0
HAP_18	0	0	0	0	0	0	0	0	0.075
HAP_19	0	0	0	0	0	0	0	0	0.061
HAP_20	0	0	0	0	0	0	0	0	0.050
HAP_21	0	0	0	0	0	0	0	0	0.045
HAP_22	0	0	0	0	0	0	0	0	0.040
HAP_23	0	0	0	0	0	0	0	0	0.035
HAP_24	0	0	0	0	0	0	0	0	0.021
HAP_25	0	0	0	0	0	0	0	0	0.020
HAP_26	0	0	0	0	0	0	0	0	0.015
HAP_27	0	0	0	0	0	0	0	0	0.015
HAP_28	0	0	0	0	0	0	0	0	0.010

En este segundo PCA, el Componente Principal 1 y el Componente Principal 2 (PC1 y PC2) explican en conjunto el 79,86% de la varianza (48,92% y 30,94% respectivamente), dando como resultado la distribución observada en la **Figura 12**. Al igual que en el análisis anterior, las poblaciones Asiáticas se distancian claramente de las demás poblaciones, agrupadas en su mayoría hacia los valores negativos del primer componente. Las poblaciones europeas y americanas nuevamente forman grupos separados y MXL se posiciona como un punto intermedio entre ambas. Finalmente, LWK aparece aislada de las demás tanto en el PC1 como en el PC2.

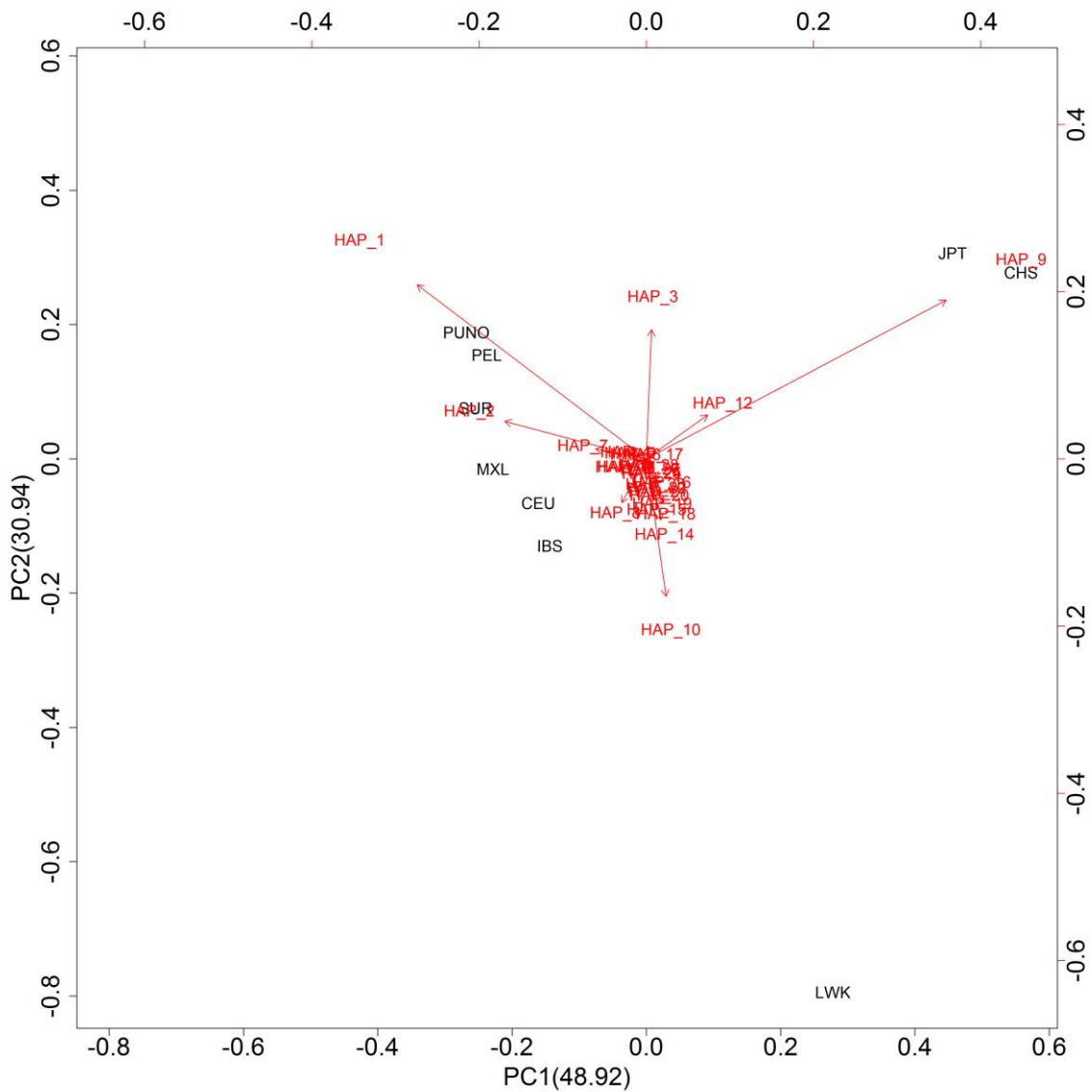


Figura 12: Análisis de Componentes Principales, Biplot PC1/PC2 (Haplotipos).

A mayor coeficiente de correlación entre el SNP y el componente, más largo su vector en la dirección del eje del componente.

Respecto a los coeficientes de correlación (Ver **Figura 12**), PC1 presenta correlación positiva principalmente con el “Haplotipo 9” (coef=0.72) y en menor medida con el “Haplotipo 12” (coef= 0.14). El “Haplotipo 9” es el más frecuente en las poblaciones asiáticas. Se encuentra en un 41,9% y en un 40,4% en CHS y JPT respectivamente, mientras que en todas las otras poblaciones no supera el 6%, estando ausente tanto en Puno como en SUR. La característica distintiva del “Haplotipo 9” es que los SNPs de los bloques haplotípicos 4 y 5 (en CHS), y 3 y 4 (en JPT) antes mencionados, presentan en este haplotipo alelos que son poco frecuentes a nivel global. En el “Haplotipo 9” el SNP rs12247705 presenta una “T” en vez de una “C”, rs79947284 presenta “T” en vez de “C”, rs12219080 presenta “T” en vez de “C”, rs721917 presenta “G” en vez de “A” y rs1885550 presenta “T” en vez de “C”. Además, en este haplotipo el SNP rs2243639 presenta el alelo “C”, siendo que en los haplotipos más frecuentes en las otras poblaciones presenta el alelo “T”. Así, la combinación de alelos que resume al “Haplotipo 9” y que lo distingue inequívocamente de cualquiera de los otros 28 haplotipos

identificados es: C-G-T-T-T-T (rs2243639, rs721917, rs12219080, rs1885550, rs79947284 y rs12247705 en ese orden).

Respecto al “Haplotipo 12”, este presenta también una frecuencia mayor en CHS y JPT (12,9% y 10% respectivamente), mientras que en las otras poblaciones no supera el 6%. Se caracteriza por la combinación de alelos C-C-G-C-T-C-C (rs911887, rs2243639, rs721917, rs12219080, rs1885550, rs79947284 y rs12247705 en ese orden).

PC1 presenta correlación negativa principalmente con el “Haplotipo 1” (coef= -0.55) y en menor medida con el “Haplotipo 2” (coef= -0.34). El “Haplotipo 1” es el más frecuente en todas las poblaciones no Asiáticas consideradas, a excepción de LWK (donde sólo se encuentra en un 1%). Sus frecuencias más altas están en Puno (47,6%) y PEL (45,9%), seguidos en orden decreciente por CEU (39,1%), MXL (39,1%), IBS (36%) y SUR (34,5%), poblaciones que efectivamente se agrupan hacia los valores negativos del PC1. Es también el segundo haplotipo más frecuente en CHS y JPT, aunque con una frecuencia considerablemente más baja si se lo compara con el “Haplotipo 9” (18,8% y 25,5% en CHB y JPT respectivamente). La combinación de alelos que caracteriza al “Haplotipo 1” y lo distingue de los demás, es opuesta a la del “Haplotipo 9” respecto a los SNPs antes mencionados: T-A-C-C-C-C (rs2243639, rs721917, rs12219080, rs1885550, rs79947284 y rs12247705 en ese orden).

Por su parte, el “Haplotipo 2” alcanza su frecuencia más alta en SUR (27,3%), seguido en orden decreciente por MXL (21,9%), Puno (11,6%), PEL (11,2%), IBS (8,4%) y CEU (7,6%). En CHS, JPT y LWK este haplotipo no supera el 3%. Es uno de los pocos haplotipos que contiene el alelo “G” del SNP rs1051246 y el único que contiene el alelo “T” de rs3088308 y de rs76239460, por lo que puede ser fácilmente identificado por la combinación de alelos: G-T-T (rs1051246, rs3088308 y rs76239460 en ese orden). El alelo “T” de rs3088308 y de rs76239460 sólo alcanzan una frecuencia mayor a 0.20 en SUR y MXL, siendo más alta en SUR (rs3088308=0.27 y rs76239460= 0.23).

PC2 presenta correlación positiva principalmente con el “Haplotipo 1” (coef=0.53), con el “Haplotipo 9” (coef=0.48) y con el “Haplotipo 3” (coef=0.39). El “Haplotipo 3” sólo se encontró en las poblaciones americanas y asiáticas, alcanzando su frecuencia más alta en Puno (21,8%), seguido en orden decreciente por SUR (18,2%), PEL (17,6%), CHS (17,4%), JPT (13,5%) y MXL (2,3%). En conjunto, las frecuencias de estos haplotipos estarían determinando la proximidad de las diferentes poblaciones hacia las esquinas superiores de la **Figura 12**, encabezadas por las poblaciones Asiáticas (que poseen las frecuencias más altas del “Haplotipo 9”) y Puno (que posee la frecuencia más alta de los “Haplotipos 1 y 3”).

PC2 presenta el coeficiente de correlación negativa más alto con el “Haplotipo 10” (coef=-0.41). Este es el haplotipo más frecuente en LWK (23,2%), pero en ninguna de las otras poblaciones supera el 6%, siendo las más cercanas IBS (5,1%) y CEU (4,5%). Esto sugiere que la posición distanciada de LWK está dada por la frecuencia de este haplotipo. El “Haplotipo 10” es el único que presenta el alelo “A” del SNP rs12241521, el

cual es muy raro en las otras poblaciones (**Ver Tabla 12**) y ya había sido relacionado a cercanía con LWK en el PCA anterior.

Además de esta primera diferenciación entre las poblaciones, si se grafica el componente principal 1 y el componente principal 3 (PC1 y PC3) es posible encontrar divisiones al interior de las poblaciones americanas que están dadas por la frecuencia de ciertos haplotipos (**Ver Figura 13**).

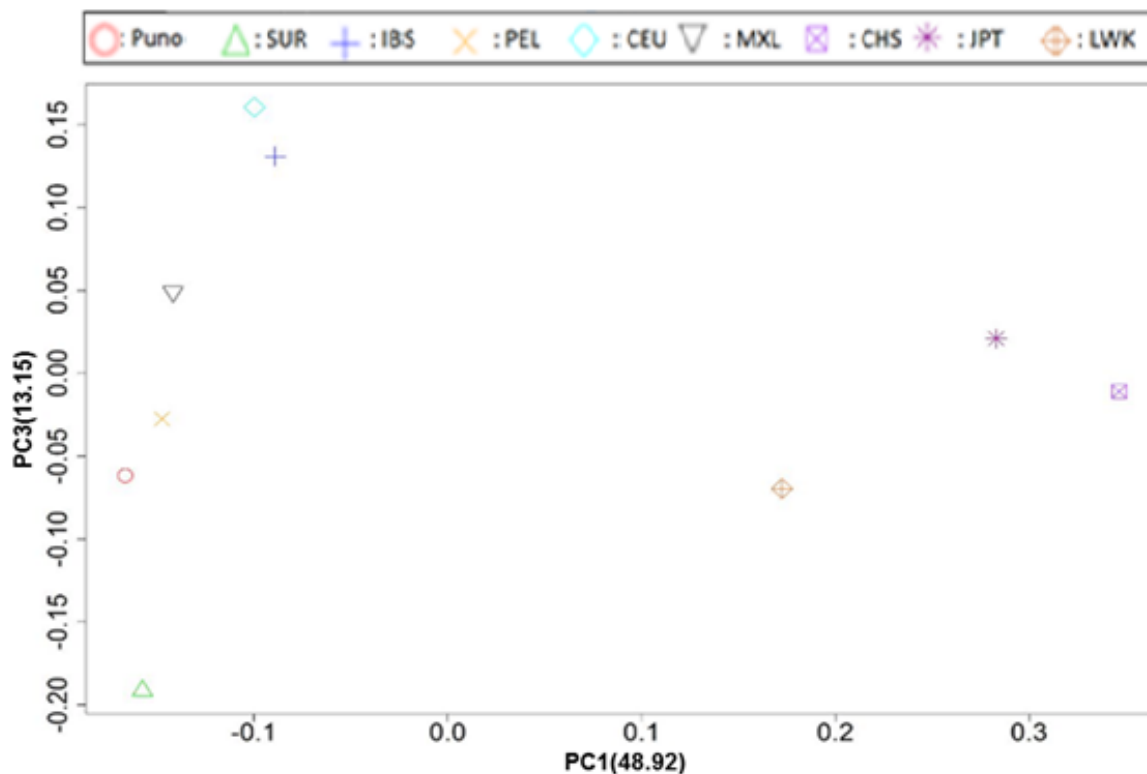


Figura 13: Análisis de Componentes Principales, Biplot PC1/PC3 (Haplotipos)

PC3 explica el 13,15% de la varianza y presenta correlación positiva principalmente con el “Haplotipo 7” (coef=0.56). Este es el segundo haplotipo más frecuente en las poblaciones europeas estudiadas (19% en CEU y 15% en IBS), con una frecuencia inferior al 7% en las otras poblaciones, a excepción de MXL (10,9%). La presencia de haplotipos raros que son más frecuentes en las poblaciones europeas y en MXL, tales como el “Haplotipo 7”, el “Haplotipo 11” (coef=0.15) y el “Haplotipo 13” (coef=0.17) explicarían el agrupamiento de MXL con las poblaciones europeas.

PC3 presenta correlación negativa principalmente con el “Haplotipo 3” (coef=-0.56), el “Haplotipo 2” (coef=-0.30), el “Haplotipo 4” (coef=-0.18) y el “Haplotipo 5” (coef=-0.14). Como se mencionó anteriormente, el “Haplotipo 3” sólo se encontró en las poblaciones americanas y asiáticas, siendo más frecuente en Puno, SUR y PEL, mientras que el “Haplotipo 2” alcanza sus frecuencias más altas en SUR, MXL, Puno y PEL. Los “Haplotipos 4 y 5” sólo se encontraron en SUR (8,6% y 6,8%). El “Haplotipo 4” posee el alelo “A” del SNP rs75448566. Como se vio anteriormente en el apartado de frecuencias

génicas, este alelo sólo se encuentra en SUR y en LWK. Sólo se encontraron dos haplotipos que incluyen el alelo “A” de este SNP: el “Haplotipo 4” (presente sólo en SUR) y el “Haplotipo 21” (presente sólo en LWK). Por su parte, el “Haplotipo 5” posee casi la misma combinación de alelos que el “Haplotipo 9”, con la diferencia de que presenta el alelo “T” del rs2243639 en vez del alelo “C”. Esta información sugiere que los “Haplotipos 3 y 4” serían los responsables de la agrupación de las poblaciones americanas hacia los valores negativos del PC1, y hacia los valores centrales y negativos del PC3 en la **Figura 13**. La presencia del “Haplotipo 7”, así como de otros haplotipos compartidos por las poblaciones europeas y MXL explican porque esta población se sitúa en un punto medio entre las poblaciones europeas y americanas, probablemente debido a la presencia de sujetos con una mayor proporción de ancestría europea en la muestra de MXL. Finalmente, los “Haplogrupos 4 y 5” exclusivos de SUR, serían la razón de la separación de esta población respecto al resto de América.

6. Discusión

6. 1 *SFTPD* y el patrón adaptativo de las poblaciones altiplánicas

Las frecuencias de los alelos, genotipos y haplotipos de los SNPs rs721917, rs3088308 y rs2243639 asociados a mayores niveles de proteína surfactante D, son altas en todas las poblaciones occidentales, presentando escasa variabilidad interpoblacional.

Aunque la frecuencia del alelo “A” de rs721917 y de rs3088308 (relacionados a mayores concentraciones de SP-D en suero) fue levemente más alta en la población Aymara de Puno que en la población amerindia del centro-sur de Chile (muestra SUR) y que en la población mestiza de Lima (Ver **Figura 4**), estas diferencias no son significativas o están al borde de la significación (Ver **Tabla 7**). Las tres poblaciones no presentan diferencias significativas ni en las frecuencias alélicas de la mayoría de los polimorfismos del gen *SFTPD* considerados, ni en sus frecuencias haplotípicas (Ver **Tabla 9**), por lo que parece poco probable que las pequeñas variaciones observadas estén explicadas por un evento selectivo en la población altiplánica producto de las condiciones de hipoxia.

De los 23 SNPs de *SFTPD* considerados originalmente, ninguno es significativamente diferente en cuanto a frecuencias alélicas entre Puno y PEL, y sólo siete polimorfismos presentaron diferencias significativas entre Puno y SUR, según la prueba exacta de Fisher (Ver **Tabla 7**). Puno se encuentra en equilibrio genético para todos estos SNPs y sólo en tres ocasiones se observaron desviación respecto al equilibrio de Hardy-Weinberg en otras poblaciones, las cuales están asociadas a un déficit de heterocigotos respecto a lo esperado (Ver **Tabla 6**). A excepción de rs3088308 en Puno, este déficit de heterocigosidad observada respecto a la esperada no parece estar relacionado a datos perdidos. Diversos fenómenos tienen como resultado una disminución en la proporción de individuos heterocigotos respecto a lo esperado bajo equilibrio de Hardy-Weinberg, entre ellos la selección positiva, un mayor nivel de endogamia en la población o bien efecto Wahlund debido a estructuración poblacional al interior de cada grupo (Nielsen & Slatkin, 2013). Puesto que la selección positiva no parece una causa probable para este déficit, es posible que en poblaciones pequeñas y aisladas (como SUR) la falta de

heterocigotos se deba a un mayor nivel de emparejamientos entre sujetos que comparten ancestros en común, mientras que en poblaciones más grandes (como CEU) esto se deba a efecto Wahlund.

A pesar de que existe cierta variabilidad propia de cada población, los bloques haplotípicos son muy similares en cuanto a composición entre las poblaciones amerindias, y prácticamente idénticos entre Puno y PEL (**Figuras 6.a – 6e**). Los valores de F_{st} obtenidos de las frecuencias alélicas de estos SNPs tampoco muestran una diferenciación importante al comparar SUR, Puno y PEL (**Tabla 10 y 11**), y los análisis de PCA apoyan el patrón observado en el dendrograma de Neighbor-Joining (Ver **Figuras 8, 11 y 12**). Paralelamente, los análisis de F_{st} Outlier indican que los SNPs de *SFTPD* no poseen valores de F_{st} o de H_e fuera de lo esperado bajo un modelo neutral (Ver **Figura 9 y 10a-b**), y que aquellos clasificados como candidatos para selección positiva en las comparaciones entre pares de poblaciones tienen esta condición más bien por frecuencias anómalas en las poblaciones de CHS y LWK que por variaciones propias de Puno.

En definitiva, los resultados de esta memoria no apoyan la señal de selección encontrada por Valverde et al., 2015, debido a que las variantes encontradas en Puno son frecuentes en varias poblaciones a nivel global. Una relación directa entre el gen *SFTPD* y el patrón adaptativo particular de las poblaciones que habitan el altiplano andino no parece probable.

Desde una perspectiva biológico-evolutiva, no encontrar variabilidad genética o fisiológica vinculada a barridos selectivos al comparar población altiplánica y poblaciones de baja altitud puede tener múltiples explicaciones. Puede que el conjunto de *loci* estudiado esté relacionado de manera demasiado remota al fenotipo de adaptación a la hipoxia de las poblaciones andinas como para apreciar diferencias significativas (Ferrell, Bertin, Barton, Rothhammer, & Schull, 1980). La adaptación a la altura implica fenotipos poligenéticos, y una sumatoria de pequeños cambios que vistos individualmente pueden no aportar mucha información, pero que en conjunto pueden dar cuenta de patrones adaptativos complejos. Es posible que el leve aumento en la frecuencia de alelos y haplotipos relacionados a mayores niveles de SP-D en sangre observados en la población de Puno por sí sólo no constituya una ventaja determinante para la supervivencia en la altura, pero su efecto en conjunto con el de otras variaciones genéticas puede dar cuenta de un fenotipo ventajoso en las poblaciones altiplánicas.

También se ha planteado la posibilidad de que las respuestas fisiológicas frente a la hipoxia observadas en los habitantes del altiplano en realidad estén dentro del rango normal de respuesta de las poblaciones humanas (Ferrell et al., 1980), sin involucrar necesariamente cambios a nivel genético. Esto es factible si consideramos las alturas mínimas de esta región. El altiplano posee un rango de altura de entre 3000 a 4500 msnm. En el límite inferior del intervalo, la aclimatación puede permitir que un mayor número de individuos no adaptados sobreviva, sin que intervenga en esto la selección natural. Más aún, el movimiento entre pisos ecológicos de diferentes altitudes característico de las poblaciones andinas podría haber disminuido los efectos de la

presión selectiva de la hipoxia en estos grupos humanos, pues descender esporádicamente a zonas de menor altitud puede ayudar a paliar los efectos dañinos de la falta de oxígeno en el organismo.

El tiempo es también un factor a considerar, pues no se tiene claridad respecto a la relación que tienen las actuales poblaciones andinas con los primeros pobladores del altiplano. Si las actuales poblaciones Aymara y Quechua efectivamente descienden de los primeros cazadores-recolectores que habitaron los Andes hace 12000 años, es probable que bajo condiciones de barrido selectivo fuerte la selección natural haya dejado huellas distinguibles en el genoma de estas poblaciones. Sin embargo, si estos grupos en realidad descienden de poblaciones de baja altitud que se asentaron en el altiplano tras la caída de Tiawanaku unos 1500 años atrás, entonces cabe dudar que la selección natural haya tenido suficiente tiempo como para reducir su diversidad genética y producir cambios significativos.

Estas últimas dos explicaciones deben ser tomadas con precaución, debido a que diferentes estudios genéticos independientes sugieren la presencia de señales de selección natural en el genoma de las actuales poblaciones del altiplano (Beall., 2014). Comprobar cualquiera de los escenarios anteriormente mencionados implicaría nuevas investigaciones cuyos alcances van más allá de los objetivos de este trabajo.

6.2 Análisis genómicos y métodos para detectar selección natural

Los resultados de esta memoria invitan a evaluar críticamente las regiones candidatas propuestas mediante estudios genómicos y remarcan la importancia de contrastar la información obtenida de poblaciones locales con el panorama global.

En el caso del gen *SFTPD*, es necesario discutir los factores que pudieron influenciar la detección de una señal de selección en esta región. A parte de esta señal, el estudio de Valverde et al., 2015 propuso también otras regiones candidatas con SNPs bajo posible selección natural, por lo que no debe descartarse que la señal encontrada en el *SFTPD* en la población altiplánica esté dada en realidad por otro gen en un área vecina. Un candidato muy destacado fue la región del cromosoma 10 entre las posiciones 82.0 y 82.3. Esta región obtuvo el puntaje más alto en varios análisis genómicos del mismo estudio, contiene varios posibles genes candidatos (como *FAM213A* y *ANXA11*) y se encuentra relativamente cerca de la región del *SFTPD*. Puede ser que la señal encontrada en este gen sea sólo un efecto de la señal en esa otra región.

Otras posibles explicaciones implican factores de confusión en los análisis genómicos, los que incluyen la tasa de recombinación de la región genómica, errores metodológicos, la asignación de valores de significación estadística y la estructuración poblacional debido a eventos demográficos (Nielsen, Hellmann, Hubisz, Bustamante, & Andrew, 2007). En el caso del estudio de Valverde et al., 2015, los factores de confusión más probables son la tasa de recombinación de la región y la estructuración poblacional.

La señal de selección hallada en la región del *SFTPD* fue encontrada utilizando análisis de REHH y se sustenta principalmente en la presencia de un haplotipo con una alta

frecuencia (52,6%) en la población altiplánica, cuya homocigosidad se extiende a través de una larga región cromosómica antes de decaer. La tasa de recombinación de una región genómica puede influir en los resultados de este tipo de análisis, pues regiones con baja recombinación implican extensas áreas con alto desequilibrio de ligamiento y poca diversidad que pueden producir falsos positivos en la detección de señales de selección (Nielsen et al., 2007). Sin embargo, en los mapas de desequilibrio de ligamiento del genoma humano consultados (Kong et al., 2002; Myers, 2005), la región del gen *SFTPD* no presenta indicios de recombinación particularmente baja.

La estructuración poblacional también puede producir falsos positivos en las señales de selección y eventos como los cuellos de botella pueden reducir la diversidad genética, generando patrones similares a la selección natural (Nielsen et al., 2007). El estudio de Valverde et al., 2015 realizó simulaciones computacionales bajo diferentes modelos demográficos, considerando modelos neutrales de cuello de botella y de tamaños poblacionales constantes en las poblaciones amerindias. Se concluyó que los valores de diferenciación genética observados en la región estudiada (cromosoma 10: 81.1 - 82.6) entre la población altiplánica y la población Guaraní eran demasiado altos para ser alcanzados por deriva (F_{st} sobre 0.5). Sin embargo, los valores de F_{st} encontrados en dicha región por Valverde et al., 2015 distan mucho de los valores estimados en esta memoria para la región particular del gen *SFTPD* al comparar las poblaciones amerindias, por lo que es posible que los resultados de las simulaciones no estén reflejando la situación particular de este gen. También puede cuestionarse si los resultados de estas simulaciones serían los mismos si se consideraran escenarios más complejos, como múltiples cuellos de botella en poblaciones que se desplazan geográficamente.

Independientemente de lo indicado en las simulaciones, algunos de los resultados del estudio de Valverde et al., 2015 sí parecen relacionarse con cuellos de botella. Las diferencias significativas observadas entre las dos poblaciones amerindias respecto a las frecuencias de las tres mutaciones no sinónimas (rs3088308, rs2243639 y rs721917) en ese estudio están dadas más por frecuencias anómalas en la población Guaraní que por patrones particulares en las poblaciones altiplánicas. Son las frecuencias alélicas de la población Guaraní (que no habita la altura) las que difieren de lo esperado si consideramos el panorama mundial. Puesto que el mismo estudio señaló no encontrar señales de selección coherentes en la población Guaraní, es posible que estas diferencias sí se deban a cuellos de botella sufridos ya sea por los grupos ancestrales que poblaron las regiones tropicales de Sudamérica o como efecto de reducciones post-contacto con los colonizadores europeos.

Algunas observaciones también pueden hacerse respecto a los enfoques que buscan detectar señales de selección a través del desequilibrio de ligamiento (como el REHH), pues la presencia de haplotipos extendidos y de bloques haplotípicos largos no necesariamente son indicativos de selección natural. El desequilibrio de ligamiento en una misma región puede variar entre diferentes poblaciones solamente por los procesos de mestizaje que les dieron origen.

La **Figura 14**, construida a partir de *loci* putativamente neutrales utilizando las poblaciones de 1000Genomes ejemplifica esta situación. Se observa que los valores medios de r^2 en una misma ventana pueden ser muy distintos inclusive entre poblaciones cercanas o de una misma región geográfica. Por ejemplo, los valores promedio de r^2 son considerablemente más bajos en la muestra de Colombianos de Medellín (CLM) que en la de Peruanos de Lima (PEL). También, se muestra que el decaimiento del desequilibrio de ligamiento a medida que incrementa la distancia entre los *loci* difiere entre poblaciones. Por ello, no sería extraño encontrar bloques haplotípicos más extensos y valores de desequilibrio de ligamiento más altos en PEL que en CLM al comparar la misma región cromosómica

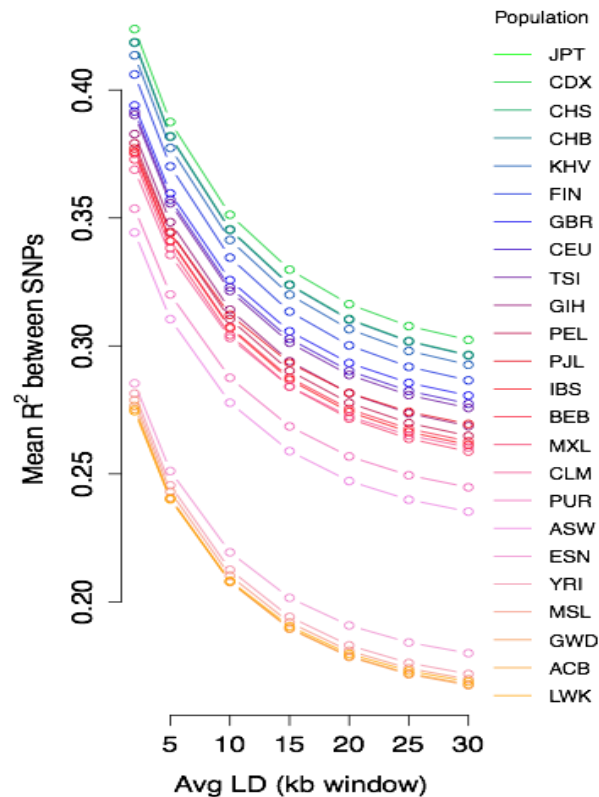


Figura 14: Desequilibrio de ligamiento en diferentes poblaciones. Imagen no publicada producida por el 1000Genomes Consortia (phase 3 data) y facilitada por la Dr. Shaila Musharoff (Stanford University). ***JPT**: Japoneses en Tokyo (Japón), **CDX**: Dai en Xishuangbanna (China), **CHS**: Han del Sur de China, **CHB**: Han en Beijing (China), **KHV**: Kinh en Ho Chi Minh City (Vietnam), **FIN**: Finlandeses en Finlandia, **GBR**: Británicos en Inglaterra y Escocia (UK), **CEU**: Residentes de Utah (EEUU) con ancestría del Norte y Oeste de Europa, **TSI**: Toscanos en Italia, **GIH**: Guajarati de Houston Texas (EEUU), **PEL**: Peruanos de Lima, **PJI**: Punjabi de Lahore (Pakistán), **IBS**: Población Ibérica de España, **BEB**: Bengali de Bangladesh, **MXL**: Mexicanos-Americanos de Los Angeles California (EEUU), **CLM**: Colombianos de Medellín (Colombia), **PUR**: Puerto Riqueños de Puerto Rico, **ASW**: Americanos de ancestría africana en SW (USA), **ESN**: Esan en Nigeria, **YRI**: Yoruba en Ibadan (Nigeria), **MSL**: Mende en Sierra Leona, **GWD**: Gambiaños en las divisiones occidentales de Gambia, **ACB**: Caribeños africanos en Barbados, **LWK**: Luhya de Webuye Kenia.

6.3 Diversidad haplotípica y diferenciación genética entre poblaciones.

Para la región estudiada, los bloques haplotípicos en las poblaciones americanas comparten muchos SNPs y poseen valores de desequilibrio de ligamiento parecidos, siendo muy similares entre sí (Ver **Figura 6.a-6.f**). En todas las poblaciones americanas, los SNPs de *SFTPD* incluidos en bloques de ligamiento son escasos, y en ninguna de las poblaciones amerindias se encontró un bloque de haplotipos extenso con altos valores de r^2 en la región del *SFTPD*. Este tipo de bloque habría sido esperable en Puno si el gen estuviese bajo alguna presión selectiva relacionada a la altura.

El patrón de bloques observado en las otras poblaciones difiere según el grupo continental. En las poblaciones europeas, estas presentaron bloques más pequeños y fragmentados, pero que incluyen varios de los polimorfismos encontrados en las poblaciones americanas (Ver **Figura 6.g-6.h**). En el caso de las poblaciones asiáticas, los bloques haplotípicos son más extensos y presentan valores internos de r^2 y de intervalos de D' más altos. El bloque 3 de JPT (58kb) es el que más se asemeja de entre todos al tipo de bloque haplotípico que cabría esperar si los SNPs de *SFTPD* estuviesen bajo selección natural, pero incluso en este bloque hay pares de SNPs con valores de intervalos de D' y r^2 demasiado bajos como para que sea el caso (Ver **Figura 6.j**). Finalmente, en LWK se encontró un patrón fragmentado, compuesto por bloques de poca extensión. Este patrón fragmentado es característico de bloques haplotípicos más antiguos cuyos valores de desequilibrio de ligamiento interno han sido afectados por la recombinación, desarmando parcialmente los bloques.

Lo mencionado en el párrafo anterior sugiere que existen ciertas tendencias en cuanto a haplotipos en el conjunto de poblaciones estudiado. Exceptuando a LWK (que posee un gran número de haplotipos propios no encontrados en ningún otro lugar), el resto de las poblaciones pueden dividirse en dos grandes grupos: aquellas donde el haplotipo más frecuente es el "Haplotipo 1" y aquellas donde el más frecuente es el "Haplotipo 9". El primer grupo incluye a las poblaciones americanas y europeas, mientras que el segundo incluye a las dos poblaciones asiáticas consideradas. Estos dos haplotipos son los principales determinantes de la agrupación de las poblaciones en el PCA de frecuencias haplotípicas (Ver **Figura 12**). En el primer grupo, Puno efectivamente presenta la frecuencia más alta del Haplotipo 1 (Ver **Tabla 14**), pero su diferencia respecto a SUR y a las otras poblaciones es demasiado baja como para plantear una posible presión selectiva sobre él. Observando el PC3 en el PCA de frecuencias haplotípicas en relación a este grupo (Ver **Figura 13**), se observó que el distanciamiento de la muestra SUR respecto a las otras poblaciones del Cono Sur de América está dada por haplotipos que aparecen sólo en esta población ("Haplotipos 4 y 5") y que la proximidad de MXL a las poblaciones europeas está relacionado al aumento de haplotipos más frecuentes en Europa (Ej: "Haplotipo 7") en MXL, posiblemente debido a una mayor proporción de ancestría europea en los Mexicanos-Americanos de Los Angeles California, que en las otras poblaciones americanas estudiadas. Estos antecedentes sugieren que el PC3 está

reflejando mestizaje, posicionando a las poblaciones europeas en un extremo y a SUR (la más aislada de las poblaciones amerindias consideradas) del otro.

El segundo grupo está compuesto por las poblaciones asiáticas, donde el “Haplotipo 9” es el más frecuente. Este se encuentra sobre el 40% en CHS y JPT, y está bajo el 6% en todas las demás (Ver **Tabla 14**). A diferencia del grupo anterior, los polimorfismos de *SFTPD* en las dos poblaciones asiáticas consideradas se agrupan en bloques haplotípicos con desequilibrio de ligamiento alto (Ver **Figura 6.i y 6.j**) y presentan altos valores de *Fst* al compararlos con otras poblaciones. Estos patrones no están presentes en todo el genoma, como se pudo comprobar a partir de la comparación con los valores de *Fst* de la zona neutral (Ver **Figuras 7.a y 7.b**). Este resultado es interesante, pues el “Haplotipo 9” incluye mutaciones de *SFTPD* que están en baja frecuencia en otras poblaciones, tres de las cuales fueron señaladas como datos atípicos en el análisis de *Fst* Outlier entre Puno y CHS (Ver **Figura 10b**). Este haplotipo incluye alelos de las tres mutaciones no sinónimas mencionadas por Valverde et al., 2015 relacionados a niveles significativamente más bajos de proteína surfactante D (Horimasu et al., 2014; Ishii, Hagiwara, Ikeda, et al., 2012; Ishii, Hagiwara, Kamio, et al., 2012; Ou, Chen, Hsiue, Lin, & Wang, 2015). Estudios médicos recientes han encontrado diferencias significativas en los niveles de proteína surfactante D al comparar muestras de población alemana y japonesa, siendo estos niveles más bajos en la población de Japón (Horimasu et al., 2014). Los genotipos relacionados a niveles más altos de esta proteína están en mayor frecuencia en la población Alemana, mientras que en la población Japonesa son más frecuentes el alelo codificante para Treonina de rs721917 y el alelo codificante para Alanina de rs2243639, relacionados a niveles más bajos de la proteína (Horimasu et al., 2014).

El patrón genético encontrado en las poblaciones asiáticas (alto desequilibrio de ligamiento y elevados valores de *Fst* en una región codificante del genoma, junto con diferencias significativas en frecuencias alélicas y haplotípicas en comparación con otras poblaciones) puede ser confundido con señales de selección natural, pero considerando la información antes mencionada, los antecedentes históricos de estas poblaciones y los hallazgos recientes del Asian Diversity Project (Liu et al., 2017), es más probable que se trate de un fenómeno demográfico conocido como “allelic surfing” (“surf alélico”).

Estudios genómicos recientes utilizando algoritmos altamente sensibles han encontrado señales de selección positiva compartidas por múltiples poblaciones Asiáticas. Estos resultados incluyen señales compartidas entre grupos Han y Japoneses, tales como las encontradas en el gen *ABCC11* que determina la composición de la cera de oído, entre muchos otros candidatos (Liu et al., 2017). Sin embargo, no se detectaron señales de selección positiva en el gen *SFTPD*.

Aunque el haplotipo predominante en la población asiática está relacionado a un cambio fenotípico (menores niveles de proteína surfactante D), todavía existe polémica respecto de si tener menores niveles de proteína surfactante D se trata de un rasgo perjudicial o beneficioso. Por ejemplo, algunas investigaciones médicas indican que menores niveles

de SP-D debido a la variante que codifica para Treonina del rs721917 aumentan la susceptibilidad a sufrir enfisema y cáncer de pulmón (Ishii et al., 2012), mientras que otros estudios sugieren que menores niveles de SP-D debido a la misma mutación podrían reducir el riesgo de enfermedad pulmonar obstructiva crónica y facilitar la recuperación (Ou et al., 2015). Es posible que ambos escenarios sean ciertos y que variaciones en los niveles de SP-D sean beneficiosas o perjudiciales dependiendo del tipo de patologías que afecten a las poblaciones, pudiendo esta característica ser una ventaja selectiva en ciertos ambientes. Sin embargo, hasta que no se disponga de un estudio detallado sobre como las enfermedades respiratorias han afectado la historia evolutiva de las poblaciones del este de Asia, proponer una posible presión selectiva relacionada a este fenotipo carece de antecedentes.

Por otro lado, eventos estocásticos como el “allelic surfing” pueden dejar patrones en el genoma que presentan muchas de las características atribuibles a los barridos selectivos, tales como baja diversidad haplotípica y alta frecuencia de variantes raras. El “allelic surfing” es un fenómeno demográfico producido por una población en expansión que se desplaza geográficamente a una velocidad constante, causando una “ola de avance” que se mueve en la dirección de la expansión (Edmonds, Lillie, & Cavalli-Sforza, 2004). Se genera un pronunciado efecto fundador en las mutaciones que aumentan su frecuencia en las áreas recientemente colonizadas, al ser difundidas por la “ola de avance” (Edmonds et al., 2004). Esta combinación de eventos demográficos permite que alelos raros en la población original puedan aumentar rápidamente su frecuencia en las nuevas áreas colonizadas, “surfeando” en el frente de la “ola de avance” que produce la población al expandirse territorialmente (Excoffier & Ray, 2008; Hofer, Ray, Wegmann, & Excoffier, 2009). En comparación, las mutaciones que permanecen en su mismo lugar de origen o cerca de él tienden a tener bajas frecuencias (Edmonds et al., 2004). La propagación de estos alelos raros es posible debido a que (comparativamente) los grupos en el frente de la “ola de avance” son más pequeños que la población total, permitiendo que eventos estocásticos produzcan grandes cambios en sus frecuencias alélicas (Excoffier & Ray, 2008). Es la composición genética de estos grupos en el “frente de la ola” lo que determinará la diversidad genética propagada por los colonizadores, la cual puede ser muy diferente de la composición general de la población original (Excoffier & Ray, 2008).

Los patrones producidos en la estructura genética de las poblaciones por una sola de estas expansiones territoriales pueden ser muy complejos e incluir muchos elementos típicamente atribuidos a la selección natural: la diversidad genética inicial disminuye debido a un conjunto de cuellos de botella y sucesivos eventos fundadores ocurridos durante la expansión, a lo que se suma el movimiento de la población. Esto genera un aumento en la distancia genética respecto a la población ancestral (Excoffier & Ray, 2008; Klopstein, Currat, & Excoffier, 2006) El “allelic surfing” no se restringe a las mutaciones nuevas. Si una mutación previamente existente en la población se encuentra lo suficientemente próxima al frente de la oleada migratoria, entonces se propagará de la misma manera que aquellas que aparecieron durante la oleada (Klopstein et al., 2006). Otras características incluyen la aparición de varias clinas (“clines”) en la dirección de la

migración y la alta frecuencia de unas cuantas mutaciones (aquellas que tuvieron éxito durante el “surfing”) en un área geográfica colonizada (Klopfstein et al., 2006).

El “allelic surfing” es capaz de afectar otros aspectos de la diversidad genética además de las frecuencias, tales como el desequilibrio de ligamiento y la homocigosidad (Hofer et al., 2009), debido a que las variables raras favorecidas por la oleada pueden ser capturadas en haplotipos extensos sólo por efecto de deriva a medida que la población se expande geográficamente (Nielsen et al., 2007). Las mutaciones favorecidas por la oleada pueden ubicarse tanto en regiones génicas como en regiones putativamente neutrales y, al igual que la selección natural, este fenómeno no afecta al genoma completo, sino que sólo a algunos *loci* (Hofer et al., 2009). Tampoco está limitado a poblaciones pequeñas, pudiendo ocurrir en grupos migrantes con una capacidad de carga de más de 500 individuos, lo que en poblaciones humanas indica que este fenómeno puede darse tanto en pequeños grupos cazadores-recolectores como en poblaciones agrícolas más extensas (Klopfstein et al., 2006).

Puesto que no se conocen métodos que permitan distinguir con precisión los patrones dejados en el genoma por los eventos selectivos de aquellos dejados por el surfing (Nielsen, 2005), resulta difícil determinar si el aumento en la frecuencia de variantes específicas de mutaciones no sinónimas en una población se debe a barridos selectivos o procesos demográficos complejos. Un buen ejemplo de la similitud en los patrones dejados por estos dos fenómenos es el caso del gen humano *MCPH1* (Excoffier & Ray, 2008). Este posee una mutación con muy alta frecuencia en las poblaciones no africanas, lo cual se ha interpretado tradicionalmente como efecto de selección positiva sobre el gen y sobre el tamaño cerebral posterior a la salida de los seres humanos modernos de África (Evans et al., 2005). Mientras que las investigaciones biológicas han sido incapaces de encontrar evidencia de ventajas selectivas asociadas a esta mutación (Mekel-Bobrov et al., 2007; Timpson, Heron, Smith, & Enard, 2007), las simulaciones computacionales han demostrado que el efecto puramente azaroso del “allelic surfing” durante la expansión poblacional a la salida de África es capaz de dejar un patrón muy semejante (Currat et al., 2006). Situaciones como la anterior llevan a cuestionar si algunas supuestas señales de selección encontradas mediante métodos genómicos no sean en realidad productos del “allelic surfing”. Tampoco debe descartarse que estos fenómenos puedan ocurrir simultáneamente, y que alelos beneficiosos en un momento y ambiente determinados puedan esparcirse hacia nuevas regiones producto de procesos de difusión démica. Quizás la mayor diferencia entre estos fenómenos sea que el surfing no se relaciona directamente a presiones ambientales, sino más bien a comportamientos poblacionales.

Considerando la baja frecuencia del “Haplotipo 9” (bajo el 6% en las otras poblaciones y de sólo 3% en la población africana de LWK) y los antecedentes históricos de la población Han y Japonesa, es posible que su mayor frecuencia en las poblaciones asiáticas se deba a un fenómeno de “allelic surfing” posterior a la salida de África y a procesos migratorios al interior del continente asiático.

La “difusión démica” es la expansión geográfica de una población inicialmente pequeña debido a un crecimiento demográfico inicial (Edmonds et al., 2004), patrón que sumado al movimiento poblacional calza con los requerimientos para producir “allelic surfing”. Estudios genéticos utilizando marcadores uniparentales en la población Han apoyan la hipótesis de procesos de difusión démica en su historia temprana (Wen et al., 2004; Zhao et al., 2015).

Los ancestros de la población Han eran agricultores cuyos orígenes se remontan a los grupos étnicos Huaxia (XXI-VIII BC) de la Planicie Central del norte de China (Zhao et al., 2015). Hacia el 260 AC los grupos Han (o sus ancestros) se movilizaron hacia el norte y hacia el sur de China, contribuyendo fuertemente al pool genético en las regiones colonizadas (Zhao et al., 2015). Los patrones genéticos producto de la expansión hacia el Sur de la población Han apoyan la hipótesis de movimientos masivos de gente durante al menos los últimos dos milenios, encontrándose una gran similitud en los haplogrupos del cromosoma-Y entre los grupos Han del Norte y del Sur y un importante aporte de las poblaciones Han en el ADN mitocondrial de las poblaciones del Sur de China (Qu et al., 2012; Wen et al., 2004). A esto se suman registros históricos que señalan al menos tres oleadas migratorias masivas de grupos Han del Norte hacia el Sur debido a conflictos bélicos y hambruna: la primera se dio entre 265-316 DC e involucró 900.000 personas, la segunda ocurrió entre 618 y 907 DC e involucró un número similar de personas y la tercera tuvo lugar entre 1127 y 1279 DC e involucró 5 millones de personas (Wen et al., 2004). Además de estas tres oleadas, otras migraciones más pequeñas hacia el Sur ocurrieron en repetidas ocasiones durante al menos los últimos dos milenios (Wen et al., 2004)

Es posible que el “Haplotipo 9” del gen *SFTPD* se haya originado en las poblaciones agrícolas del continente (de las cuales las actuales poblaciones Han son descendientes) producto de estos movimientos demográficos u otros equivalentes, y que haya llegado al archipiélago japonés hace unos pocos milenios con grupos migrantes del continente.

La colonización de Japón se dio principalmente en dos momentos distintos por dos grupos humanos: 1) los grupos cazadores-recolectores Jomon provenientes del Sudeste Asiático que ingresaron al archipiélago entre el 12000-10000 A.P, durante el Último Máximo Glaciar y 2) las poblaciones agrícolas provenientes del Noreste Asiático conocidas como “cultura Yayoi”, que ingresaron a las islas desde el continente hace 1700-2800 años vía la península de Corea (Adachi, Shinoda, Umetsu, & Matsumura, 2009; Nakagome et al., 2015).

Hacia finales del Paleolítico, los deshielos causados por el aumento de las temperaturas habrían elevado el nivel del mar, aislando a los grupos Jomon en el archipiélago hasta la llegada de los Yayoi (Adachi et al., 2009). Los grupos agrícolas Yayoi se habrían mezclado con los grupos Jomon originales, especialmente en las islas principales del oeste del archipiélago, y habrían extendido su influencia gradualmente hacia el este, existiendo una gran afinidad entre las poblaciones del noreste asiático y las del oeste de Japón en un gran número de características biológicas, tales como grupos sanguíneos, estatura, índices cefálicos y polimorfismos nucleares (Hanihara, 1991; Nakagome et al.,

2015). Análisis morfológicos y genéticos sugieren que los grupos Yayoi y los grupos Han del Norte de China comparten una misma población ancestral (Hanihara, 1991; Nakagome et al., 2015), por lo que es posible que los eventos de expansión y migración poblacional que afectaron a los grupos Han hayan afectado también la composición genética de los grupos Yayoi. Estos grupos podrían haber difundido los haplotipos generados previamente por el “allelic surfing” durante su migración hacia el archipiélago japonés al colonizar las islas y mezclarse con la población original, entre los cuales podría haber estado el “Haplotipo 9”.

Si bien los tiempos entre las migraciones registradas de la población Han y el ingreso de los grupos Yayoi al archipiélago japonés se sobrelapan en algunos periodos, todavía existe discrepancia respecto al inicio de al menos el primero de estos procesos y no debe descartarse que los eventos de difusión démica que afectaron a las poblaciones asiáticas del continente se hayan dado en fechas más tempranas de las aquí presentadas. Además, la velocidad de la oleada expansiva y del “allelic surfing” está determinada en gran medida por la velocidad de la migración, la tasa de crecimiento logístico de la población y el tamaño poblacional de la misma, así como también la “cercanía” de las mutaciones respecto al frente de la “oleada de avance”, pudiendo variar sus efectos significativamente de acuerdo a estos factores (Klopfstein et al., 2006). El proceso de colonización será más rápido en poblaciones más grandes, pero las probabilidades de un alelo de fijarse en las áreas colonizadas en un menor número de generaciones serán mayores en grupos migratorios con menor capacidad de carga (Klopfstein et al., 2006).

Una posible alternativa para analizar este panorama sería realizar simulaciones computacionales utilizando parámetros poblacionales teóricos para las poblaciones Han del noreste asiático durante el periodo de las migraciones masivas más tempranas registradas. Esto permitiría analizar cuantas generaciones de “surfing” serían necesarias para obtener las frecuencias alélicas y haplotípicas observadas en las poblaciones asiáticas actuales, a diferentes velocidades de migración, tasas de crecimiento y capacidades de carga. Con ello se podría estudiar si existen escenarios donde los datos encontrados sean concordantes con los periodos históricos de las migraciones masivas mencionadas.

7. Conclusiones

Si bien se observó variabilidad interpoblacional en variantes del gen *SFTPD*, los resultados de esta memoria no apoyan la existencia de una relación entre este gen y el patrón adaptativo particular de las poblaciones que habitan el altiplano andino, al menos en lo que respecta a las mutaciones consideradas. Las frecuencias alélicas, genotípicas y haplotípicas entre la población altiplánica de Puno y las poblaciones de baja altitud estudiadas (tanto de Sudamérica como de otros continentes) presentan poca variación, y en muchos casos las diferencias no son significativas.

Los resultados de esta memoria resaltan la importancia de las comparaciones interpoblacionales con varios grupos control, pues características y señales que pueden parecer muy propias de una población particular cuando se la compara con un grupo control específico, pueden en realidad ser elementos comunes a varias poblaciones cuando se las inserta en un contexto más amplio.

En los estudios sobre selección natural es importante además tener en cuenta factores de confusión tales como la tasa de recombinación y la estructuración poblacional. Pues por simple azar estos elementos pueden generar patrones en el genoma con muchas de las características de los barridos selectivos, incluyendo extensos haplotipos en regiones génicas con altas frecuencias y altos niveles de diferenciación genética respecto a poblaciones ancestrales y emparentadas.

El modelo anterior permite proponer una explicación a los altos valores de diferenciación genética encontrados al comparar las poblaciones asiáticas (Han del Sur de China y Japonesa de Tokyo) con poblaciones de otras regiones geográficas. Posiblemente estas diferencias están dadas más por eventos estocásticos en las poblaciones asiáticas, tales como el “allelic surfing”, que por efecto de la selección natural. Este panorama está respaldado por estudios genéticos y antecedentes históricos que sugieren la existencia de expansiones territoriales y movimientos poblacionales masivos en la historia temprana de la población Han y durante la colonización prehistórica del archipiélago Japonés por parte de poblaciones del continente.

8. Referencias

- Adachi, N., Shinoda, K. I., Umetsu, K., & Matsumura, H. (2009). Mitochondrial DNA analysis of jomon skeletons from the funadomari site, hokkaido, and its implication for the origins of native american. *American Journal of Physical Anthropology*, 138(3), 255–265. <http://doi.org/10.1002/ajpa.20923>
- Alexander, D. H., Lange, K., Pritchard, J., Stephens, M., Donnelly, P., Tang, H., ... Nichols, R. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, 12(1), 246. <http://doi.org/10.1186/1471-2105-12-246>
- Amils, R., Ellis-Evans, C., & Hinghofer-Szalkay, H. (2007). *Life in extreme environments*. New York, NY: Springer.
- Antao, T., Lopes, A., Lopes, R. J., Beja-pereira, A., & Luikart, G. (2008). LOSITAN : A workbench to detect molecular adaptation based on a F st -outlier method. *BMC Bioinformatics*, 5, 1–5. <http://doi.org/10.1186/1471-2105-9-323>
- Beall, C. M., Baker, P. T., Baker, T. S. and Haas, J. D. (1977). The effects of high altitude on adolescent growth in southern Peruvian Amerindians. *Human Biol.* 49, 109–124
- Beall, C. M. (2007). Two routes to functional adaptation: Tibetan and Andean high-altitude natives. *Proceedings of the National Academy of Sciences*, 104(suppl 1), 8655-8660.
- Beall, C. M. (2014). Adaptation to High Altitude: Phenotypes and Genotypes. *Annual Review of Anthropology*, 43, 251–272. <http://doi.org/10.1146/annurev-anthro-102313-030000>
- Beaumont, M. A., & Nichols, R. A. (1996). Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B: Biological Sciences*, 263(1377), 1619-1626.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289–300. Retrieved from <http://www.jstor.org/stable/2346101>
- Bennett, W. C., & Bird, J. B. (1964). *Andean culture history* (No. 15). American Museum of Natural History.
- Bigham, A., Bauchet, M., Pinto, D., Mao, X., Akey, J. M., Mei, R., ... Shriver, M. D. (2010). Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. *PLoS Genetics*, 6(9). <http://doi.org/10.1371/journal.pgen.1001116>
- Bigham, A. W., Wilson, M. J., Julian, C. G., Kiyamu, M., Vargas, E., Leon-Velarde, F., ... & Brutsaert, T. D. (2013). Andean and Tibetan patterns of adaptation to high altitude. *American Journal of Human Biology*, 25(2), 190-197.
- Bigham, A. W., & Lee, F. S. (2014). Human high-altitude adaptation: forward genetics meets the HIF pathway. *Genes & Development*, 28(20), 2189–204.

<http://doi.org/10.1101/gad.250167.114>

- Budge, P. J., Griffin, M. R., Edwards, K. M., Williams, J. V., Verastegui, H., Hartinger, S. M., ... Grijalva, C. G. (2014). A Household-based Study of Acute Viral Respiratory Illnesses in Andean Children. *The Pediatric Infectious Disease Journal*, 33(5), 443–447. <http://doi.org/10.1097/INF.000000000000135>
- Capriles, J. M., & Albarracín-Jordan, J. (2013). The earliest human occupations in Bolivia : A review of the archaeological evidence. *Quaternary International*, 301, 46–59. <http://doi.org/10.1016/j.quaint.2012.06.012>
- Carnese, F. R., Mendisco, F., Keyser, C., Dejean, C. B., Dugoujon, J. M., Bravi, C. M., ... Crubézy, E. (2010). Paleogenetical study of pre-Columbian samples from Pampa Grande (Salta, Argentina). *American Journal of Physical Anthropology*, 141(3), 452–462. <http://doi.org/10.1002/ajpa.21165>
- Castro López, F. W., Labarrere Cruz, Y., González Hernández, G., & Barrios Rentería, Y. (2007). Factores de riesgo del Síndrome Dificultad Respiratoria de origen pulmonar en el recién nacido. *Revista Cubana de Enfermería*, 23(3). Retrieved from http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S086403192007000300005&lng=es&nrm=iso&tlng=es
- Chiodi, H. (1957). Respiratory adaptations to chronic high altitude hypoxia. *Journal of Applied Physiology*, 10(1), 81-87.
- Currat, M., Excoffier, L., Maddison, W., Otto, S. P., Ray, N., Whitlock, M. C., & Yeaman, S. (2006). Comment on “Ongoing Adaptive Evolution of ASPM, a Brain Size Determinant in Homo sapiens” and “Microcephalin, a Gene Regulating Brain Size, Continues to Evolve Adaptively in Humans.” *Science*, 313(5784), 172a–172a. <http://doi.org/10.1126/science.1122712>
- Eckhardt, R. B. and Melton, T. W. (1992). Population Studies on Human Adaptation and Evolution in the Peruvian Andes. *Occasional Papers in Anthropology at Penn State*, (14).
- Edmonds, C. a, Lillie, A. S., & Cavalli-Sforza, L. L. (2004). Mutations arising in the wave front of an expanding population. *Proceedings of the National Academy of Sciences of the United States of America*, 101(4), 975–979. <http://doi.org/10.1073/pnas.0308064100>
- Evans, P. D., Gilbert, S. L., Mekel-Bobrov, N., Vallender, E. J., Anderson, J. R., Vaez-Azizi, L. M., ... Lahn, B. T. (2005). Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. *Science (New York, N.Y.)*, 309(5741), 1717–1720. <http://doi.org/10.1126/science.1113722>
- Excoffier, L., & Ray, N. (2008). Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology and Evolution*, 23(7), 347–351. <http://doi.org/10.1016/j.tree.2008.04.004>
- Fehren-Schmitz, L., Warnberg, O., Reindel, M., Seidenberg, V., Tomasto-Cagigao, E., Isla-Cuadrado, J., ... Herrmann, B. (2011). Diachronic Investigations of Mitochondrial and Y-Chromosomal Genetic Markers in Pre-Columbian Andean Highlanders from

- South Peru. *Annals of Human Genetics*, 75(2), 266–283. <http://doi.org/10.1111/j.1469-1809.2010.00620.x>
- Fehren-Schmitz, L., & Georges, L. (2016). Ancient DNA reveals selection acting on genes associated with hypoxia response in pre-Columbian Peruvian Highlanders in the last 8500 years. *Scientific Reports*, 6, 23485. <http://doi.org/10.1038/srep23485>
- Ferrell, R. E., Bertin, T., Barton, S. a, Rothhammer, F., & Schull, W. J. (1980). The Multinational Andean Genetic and Health Program. IX. Gene frequencies and rare variants of 20 serum proteins and erythrocyte enzymes in the Aymara of Chile. *American Journal of Human Genetics*, 32(1), 92–102.
- Freeman, G.H., Halton, J.H.,(1951).Note on exact treatment of contingency, goodness of fit and other problems of significance,*Biometrika*, 38(1/2),141-149.
- Futuyma, D. J. (2009). Natural selection and adaptation. *Evolution. Massachusetts, USA: Sinauer Associates, Inc. Sunderland*, 279-301.
- Gayà-Vidal, M., Dugoujon, J. M., Esteban, E., Athanasiadis, G., Rodríguez, A., Villena, M., ... Moral, P. (2010). Autosomal and X chromosome Alu insertions in Bolivian Aymaras and Quechuas: Two languages and one genetic pool. *American Journal of Human Biology*, 22(2), 154–162. <http://doi.org/10.1002/ajhb.20967>
- Gayà-Vidal, M., Moral, P., Saenz-Ruales, N., Gerbault, P., Tonasso, L., Villena, M., ... Dugoujon, J. M. (2011). MtDNA and Y-chromosome diversity in Aymaras and Quechuas from Bolivia: Different stories and special genetic traits of the Andean Altiplano populations. *American Journal of Physical Anthropology*, 145(2), 215–230. <http://doi.org/10.1002/ajpa.21487>
- Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., ... & Liu-Cordero, S. N. (2002). The structure of haplotype blocks in the human genome. *Science*, 296(5576), 2225-2229.
- Gartner, L & Hiatt, J. (2000). Sistema respiratorio. (Lippincott, Ed.) Color atlas of histology (2nd ed.). Philadelphia.
- Griese, M. (1999). Pulmonary surfactant in health and human lung diseases:state of the art. *European Respiratory Journal*, 13(6), 1455-1476.
- Hanihara, K. (1991). *Dual Structure Model for the Formation of the Japanese Population*. International Research Center for Japanese Studies.
- Hartl, D. L. (1988). *A Primer of Population Genetics*. Sinauer Associates, Inc
- Hofer, T., Ray, N., Wegmann, D., & Excoffier, L. (2009). Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection. *Annals of Human Genetics*, 73(1), 95–108. <http://doi.org/10.1111/j.1469-1809.2008.00489.x>
- Hoff, C. (1972). Altitudinal variations in the physical growth of Peruvian Quechua. *American Journal of Physical Anthropology*,37(3), 441.
- Hoffmann, T. J., Zhan, Y., Kvale, M. N., Hesselson, S. E., Gollub, J., Iribarren, C., ...

- Risch, N. (2011). Genomics Design and coverage of high throughput genotyping arrays optimized for individuals of East Asian , African American , and Latino race / ethnicity using imputation and a novel hybrid SNP selection algorithm. *Genomics*, 98(6), 422–430. <http://doi.org/10.1016/j.ygeno.2011.08.007>
- Horimasu, Y., Hattori, N., Ishikawa, N., Tanaka, S., Bonella, F., Ohshimo, S., ... Kohno, N. (2014). Differences in serum SP-D levels between German and Japanese subjects are associated with SFTPD gene polymorphisms. *BMC Medical Genetics*, 15(1), 4. <http://doi.org/10.1186/1471-2350-15-4>
- HPEE. (2016). *Environments, The Society for Human Performance in Extreme*. Recuperado el 6 de 05 de 2016, de <http://www.hpee.org/>.
- Hurtado, A. (1932). Respiratory adaptation in the Indian natives of the Peruvian Andes: Studies at high altitude. *American Journal of Physical Anthropology*.17(2), 137–165.
- Ishii, T., Hagiwara, K., Ikeda, S., Arai, T., Mieno, M. N., Kumasaka, T., ... Kida, K. (2012). Association Between Genetic Variations In Surfactant Protein D and Emphysema, Interstitial Pneumonia, and Lung Cancer in a Japanese Population. *Copd: Journal of Chronic Obstructive Pulmonary Disease*, 9(4), 409-416. <http://doi.org/10.3109/15412555.2012.676110>
- Ishii, T., Hagiwara, K., Kamio, K., Ikeda, S., Arai, T., Mieno, M. N., ... Kida, K. (2012). Involvement of surfactant protein D in emphysema revealed by genetic association study. *European Journal of Human Genetics*, 20(2), 230–235. <http://doi.org/10.1038/ejhg.2011.183>
- Jobling, M. A., Hollox, E., Hurles, M., Kivisild, T., & Tyler-Smith, C. (2014). Human Evolutionary Genetics. 2nd. *Garland Science*.
- Klopfstein, S., Currat, M., & Excoffier, L. (2006). The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution*, 23(3), 482–490. <http://doi.org/10.1093/molbev/msj057>
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., & 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.
- Kong, A., Gudbjartsson, D. F., Sainz, J., Jonsdottir, G. M., Gudjonsson, S. A., Richardsson, B., ... Stefansson, K. (2002). A high-resolution recombination map of the human genome. *Nature Genetics*, 31(3), 241–247. <http://doi.org/10.1038/ng917>
- Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33(7), msw054. <http://doi.org/10.1093/molbev/msw054>
- Lahti, M., Löfgren, J., Marttila, R., Renko, M., & Klaavuniemi, T. (2002). Surfactant Protein D Gene Polymorphism Associated with Severe Respiratory Syncytial Virus Infection. *Pediatric research*, 51(6), 696-699. <http://doi.org/10.1023/01.PDR.0000015911.65104.93>

- Lathrap D.W. (1970). *The Upper Amazon*. Southampton: Thames and Hudson.
- Leth-Larsen, R., Garred, P., Jensenius, H., Meschi, J., Hartshorn, K., Madsen, J., ... Holmskov, U. (2005). A common polymorphism in the SFTPD gene influences assembly, function, and concentration of surfactant protein D. *Journal of Immunology (Baltimore, Md. : 1950)*, *174*(3), 1532–8. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15661913>
- Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, *28*(2), 298–299. <http://doi.org/10.1093/bioinformatics/btr642>
- Liu, X., Lu, D., Saw, W.-Y., Shaw, P. J., Wangkumhang, P., Ngamphiw, C., ... Teo, Y.-Y. (2017). Characterising private and shared signatures of positive selection in 37 Asian populations. *European Journal of Human Genetics*, *25*(4), 499-508. <http://doi.org/10.1038/ejhg.2016.181>
- Maryland, U. o. (s.f.). Domain mapping of disease mutations. Recuperado el 13 de 4 de 2016, de DMDM: <http://bioinf.umbc.edu/dmdm/about.php>
- Mekel-Bobrov, N., Posthuma, D., Gilbert, S. L., Lind, P., Gosso, M. F., Luciano, M., ... Lahn, B. T. (2007). The ongoing adaptive evolution of ASPM and Microcephalin is not explained by increased intelligence. *Human Molecular Genetics*, *16*(6), 600–608. <http://doi.org/10.1093/hmg/ddl487>
- Melton, T. W. (1992). Comparison of growth and development in two Peruvian populations of high altitude ancestry. En *Population Studies on Human Adaptation and Evolution in the Peruvian Andes* (ed. R. B. Eckhardt and T. W. Melton, págs. 192-219). University Park: Pennsylvania State University
- Moore, L. G. (2001). Human Genetic Adaptation to High Altitude. *High Altitude Medicine & Biology*, *2*(2), 257–279. <http://doi.org/10.1089/152702901750265341>
- Moore, L. G., Charles, S. M., & Julian, C. G. (2011). Humans at high altitude: Hypoxia and fetal growth. *Respiratory Physiology and Neurobiology*, *178*(1), 181–190. <http://doi.org/10.1016/j.resp.2011.04.017>
- Moran, E. (2008). Human Adaptability to high altitudes. En *Human adaptability: an introduction to ecological anthropology* (E. Moran ,Third Edition ed., págs. 157-187). Boulder, CO : Westview Press.
- Myers, S. (2005). A Fine-Scale Map of Recombination Rates and Hotspots Across the Human Genome. *Science*, *310*(5746), 321–324. <http://doi.org/10.1126/science.1117196>
- Nakagome, S., Sato, T., Ishida, H., Hanihara, T., Yamaguchi, T., Kimura, R., ... Sawai, H. (2015). Model-based verification of hypotheses on the origin of modern Japanese revisited by Bayesian inference based on genome-wide SNP data. *Molecular Biology and Evolution*, *32*(6), 1533–1543. <http://doi.org/10.1093/molbev/msv045>
- Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics*, *39*, 197–218. <http://doi.org/10.1146/annurev.genet.39.073003.112420>

- Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C., & Andrew, G. (2007). Recent and ongoing selection in the human genome, *Nature reviews. Genetics*, 8(11), 857.-868. <http://doi.org/10.1038/nrg2187.Recent>
- Nielsen, R., & Slatkin, M. (2013). *An introduction to population genetics: theory and applications*. Sunderland: Sinauer Associates.
- Orgeig, S., Morrison, ., & Daniels, C. B. (2015). Evolution, Development, and Function of the Pulmonary Surfactant System in Normal and Perturbed Environments. *Comprehensive Physiology*, 6, 363–422. <http://doi.org/10.1002/cphy.c150003>
- Ou, C. Y., Chen, C. Z., Hsiue, T. R., Lin, S. H., & Wang, J. Y. (2015). Genetic variants of pulmonary SP-D predict disease outcome of COPD in a Chinese population. *Respirology*, 20(2), 296–303. <http://doi.org/10.1111/resp.12427>
- Peacock, a J. (1998). ABC of oxygen: oxygen at high altitude. *BMJ (Clinical Research Ed.)*, 317(7165), 1063–1066. <http://doi.org/10.1136/bmj.317.7165.1063>
- Pritchard, J. K., Pickrell, J. K., & Coop, G. (2010). The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation. *Current biology*, 20(4). <http://doi.org/10.1016/j.cub.2009.11.055>.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ & Sham PC (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, 81.
- Qu, H.-Q., Li, Q., Xu, S., McCormick, J. B., Fisher-Hoch, S. P., Xiong, M., ... Jin, L. (2012). Ancestry informative marker set for han chinese population. *G3: Genes, Genomes, Genetics*, 2(3), 339-341. <http://doi.org/10.1534/g3.112.001941>
- Rademaker, K., Hodgins, G., Moore, K., Zarrillo, S., Miller, C., Bromley, G. R., ... & Sandweiss, D. H. (2014). Paleoindian settlement of the high-altitude Peruvian Andes. *Science*, 346(6208), 466-469. <http://doi.org/10.1126/science.1258260>
- Rothhammer, F., Llop, E., Carvallo, P., & Moraga, M. (2001). Origin and evolutionary relationships of native Andean populations. *High Altitude Medicine & Biology*, 2(2), 227–233. <http://doi.org/10.1089/152702901750265323>
- Rupert, J. L., Devine, D. V., Monsalve, M., & Hochachka, P. W. (1999). β -fibrinogen allele frequencies in Peruvian Quechua, a high-altitude native population. *American journal of physical anthropology*, 109(2), 181-186.
- Rupert, J., & Hochachka, P. W. (2001). Genetic approaches to understanding human adaptation to altitude in the Andes. *The Journal of Experimental Biology*, 204, 3151–3160.
- Sarricolea Espinoza, P., & Romero Aravena, H. (2015). Variabilidad y cambios climáticos observados y esperados en el altiplano del norte de Chile. *Revista de Geografía Norte Grande*, 183(62), 169–183. <http://doi.org/10.4067/S0718-34022015000300010>

- Sorensen, G. L., Dahl, M., Tan, Q., Bendixen, C., Holmskov, U., & Husby, S. (2014). Surfactant Protein-D–Encoding Gene Variant Polymorphisms Are Linked to Respiratory Outcome in Premature Infants. *The Journal of pediatrics*, 165(4), 683–689. <http://doi.org/10.1016/j.jpeds.2014.05.042>
- Timpson, N., Heron, J., Smith, G. D., & Enard, W. (2007). Comment on papers by Evans et al. and Mekel-Bobrov et al. on Evidence for Positive Selection of MCPH1 and ASPM. *Science (New York, N.Y.)*, 317(5841), 1036; author reply 1036. <http://doi.org/10.1126/science.1141705>
- Valverde, G., Zhou, H., Lippold, S., de Filippo, C., Tang, K., López Herráez, D., ... Stoneking, M. (2015). A Novel Candidate Region for Genetic Adaptation to High Altitude in Andean Populations. *Plos One*, 10(5), e0125444. <http://doi.org/10.1371/journal.pone.0125444>
- Weinstein, K., (2014). Modern Human Biological Adaptations to High-Altitude Environments in the Andean Archaeological Record. En *High Altitude Primates* (N. Grow ,págs. 285–324). <http://doi.org/10.1007/978-1-4614-8175-1>
- Wen, B., Li, H., Lu, D., Song, X., Zhang, F., He, Y., ... Jin, L. (2004). Genetic evidence supports demic diffusion of Han culture. *Nature*, 431(7006), 302–305. <http://doi.org/10.1038/nature02878>
- Wiley, A. S. (2004). *An ecology of high-altitude infancy: A biocultural perspective* (Vol. 10). Cambridge University Press.
- Zhao, Y.-B., Zhang, Y., Zhang, Q.-C., Li, H.-J., Cui, Y.-Q., Xu, Z., ... Zhu, H. (2015). Ancient DNA reveals that the genetic structure of the northern han chinese was shaped prior to 3,000 years ago. *PloS ONE*, 10(5), e0125676. <http://doi.org/10.1371/journal.pone.0125676>
- Zhou, D., Udpa, N., Ronen, R., Stobdan, T., Liang, J., Appenzeller, O., ... Haddad, G. G. (2013). Whole-Genome sequencing uncovers the genetic basis of chronic mountain sickness in andean highlanders. *American Journal of Human Genetics*, 93(3), 452–462. <http://doi.org/10.1016/j.ajhg.2013.07.011>