Universidad de Chile
Facultad de Ciencias Sociales
Departamento de Antropología

Investigación genómica de señales de selección positiva reciente en Aymaras, Pehuenches y Huilliches

Memoria para optar al título de
Antropóloga física

Paloma S. Contreras Z.
Profesor guía: Dr. Ricardo A. Verdugo S.
Profesor tutor: Dr. Mauricio Moraga.

2015

Agradecimientos

Este trabajo fue posible gracias al financiamiento de los proyectos FONDEF D10I1007 “Genómica de la Población Chilena: Obtención de Perfiles Genéticos Necesarios en Investigación Clínica, Salud Pública y Medicina Forense" y CONICYT USA2013-0015, titulado “Genomic investigation of the human biodiversity in the precolumbian and contemporary Chilean Patagonia”. Agradecemos también al Dr. A. Moreno, investigador de la Universidad de Stanford, por facilitar el acceso a las muestras de Puno, Perú.

Gracias al Dr. R. Verdugo, por el tiempo depositado en este trabajo, por la oportunidad de investigar este interesantísimo tema con datos y metodologías de exquisito valor, y por creer que en algún momento seremos colegas.

A M. Moraga, D. Véliz, S. Flores, S. Krapivka, y N. Montalva, por los orientadores comentarios recibidos en las distintas etapas de este trabajo.

Agradezco a los integrantes del Laboratorio GENOMED y al de Genética de Poblaciones y Evolución Humana, por la paciencia, los comentarios académicos, y el humor negro. Gracias a Karen, por compartir sus aprendizajes y mostrarme que nuestro trabajo también es un arte. A Alejandro B., por resolver los problemas más complejos, por los “dale, que ya estás lista”, gracias por el optimismo. A Alejandro S., por ayudarme incluso en los temas que no conocía, por los “tips” informáticos infinitamente útiles, y por el matonaje inmaduro –pero necesario- del día a día. A Felipe, Pamela y Larry, que hicieron de esta etapa algo que recordaré con cariño. A todos, por alimentarme de vez en cuando.

Gracias a quienes influyeron positivamente en mi formación: a L. Flores y S. Flores -nuevamente-, por inspirar mi interés en esta área de investigación, y por entregarme las herramientas necesarias para seguir aprendiendo, generando y difundiendo conocimiento.

A Paulina y Coté, fundamentales en mi paso por esta Universidad en la que nos hicimos un espacio y a final de cuentas, aprendimos bastante. Por las reflexiones, por ser tan distintas, por aguantar mis impulsos autoritarios, mis descuidos, los chistes fomes y por tantas, tantas calorías consumidas. También agradezco a Rocío, Sandra y Conipi, que aportaron a este trabajo y a mi sobrevivencia en la mención.

De forma especial, agradezco a Constanza, por insistir en que trabajara enérgicamente en esto, por las reflexiones científicas y los consejos inteligentes, por apoyarme intensamente desde otro hemisferio. Por alimentarme de comida, de afecto, de sabiduría y de las experiencias más bonitas que recordaré de este periodo.

Finalmente, gracias a mis padres por su paciencia, por el apoyo expresado en comidita al llegar tarde a casa, en las preguntas cotidianas, y en el asegurar todas las condiciones necesarias para que nos eduquemos. Por respetar que estudiara antropología sin saber qué significaba. Por la formación ética, ideológica y moral que promovieron en mi (no pueden quejarse ahora), y por confiarme desde pequeña a la educación pública (o lo que queda de ella): es lo mejor que pudieron haber hecho con mi destino.

Tabla de contenidos

Resumen

Los primeros habitantes del sur Sudamericano llegaron hace más de 14.000 años, enfrentando presiones selectivas como la altura, tóxicos ambientales, cambios en la dieta, exposición a patógenos, entre otras. Estas condiciones ambientales promovieron el desarrollo de adaptaciones biológicas, detectables a nivel genómico. Usando datos de genotipificación (700.000 *SNPs*), estudiamos elementos de la historia microevolutiva de los Aymara, Pehuenche y Huilliche; grupos de ancestría amerindia que descienden de pueblos prehispánicos del altiplano andino y del centro-sur de Chile. Al analizar su estructura poblacional con el test Fst, encontramos una baja diferenciación entre los grupos. Usando ADMIXTURE, no detectamos estructura genética entre Pehuenches y Huilliches. Con los test iHS y XPEHH, identificamos regiones genómicas con señales de selección positiva reciente (<30.000 años) en Amerindios, no presentes en asiáticos, europeos, ni africanos. Destacamos una región del cromosoma 6 (52-52,6 Mb), con alta evidencia de selección en los tres grupos. Las regiones más diferenciadas entre Aymaras y el grupo de Pehuenches y Huilliches (5% de los puntajes más altos de Fst y XPEHH) contienen genes asociados al metabolismo de lípidos y al sistema inmune. Estos resultados sugieren presiones selectivas en estos mecanismos biológicos que podrían haber existido recientemente en la historia de estas poblaciones.

Palabras clave: Microevolución, Adaptación biológica, Selección positiva, Altiplano andino, Centro-sur de Chile.

Introducción

Luego de que el ser humano moderno iniciara sus primeras migraciones dentro y fuera de África hace aproximadamente 50.000-100.000 años atrás (Scally & Durbin, 2012), este ha conseguido habitar una gran cantidad de biomas (regiones biogeográficas homogéneas). Algunos de estos presentan condiciones altamente adversas, tales como escasa disponibilidad de alimentos, temperaturas extremas, alta radiación solar, hipoxia, entre otros factores estresores. Estos elementos ambientales desencadenan respuestas que pueden ser modificaciones biológicas temporales (aclimatación), ajustes biológicos durante el desarrollo del individuo (que luego son irreversibles), ajustes regulatorios en el comportamiento asociados con la configuración cultural y social, o adaptaciones evolutivas o genéticas (heredables) (Moran, 2008).

El estudio de la adaptación humana se ha realizado utilizando herramientas provenientes tanto desde las ciencias sociales como biológicas, aunque se ha intensificado gracias al aporte de disciplinas como la biología evolutiva y la genética desde la segunda mitad del siglo veinte (Schutkowski, 2006), confirmando algunas hipótesis antropológicas. Análisis recientes de la variación genética humana revelan que cientos de genes habrían sido sujetos a selección positiva reciente, a menudo en respuesta a actividades humanas, lo cual ha generado consenso en torno a la idea de que la evolución humana reciente ha sido moldeada –en distintos grados- por la interacción gen-cultura (Laland, Odling-Smee, & Myles, 2010).

Presiones ambientales como la dieta, el clima, y los patógenos varían notoriamente entre distintos territorios y promueven respuestas adaptativas locales en las distintas poblaciones humanas (Li et al., 2008). En Sudamérica se han investigado características adaptativas en las poblaciones nativas a través de análisis genéticos y fenotípicos. Estos estudios se han centrado en descubrir adaptaciones asociadas a elementos ambientales específicos como la altura, en habitantes del centro sur Andino (A. W. Bigham et al., 2013; Eichstaedt et al., 2014; Francisco Rothhammer, Fuentes Guajardo, Chakraborty, Lorenzo Bermejo, & Dittmar, 2015; Valverde et al., 2015), o al consumo de arsénico, en San Antonio de los Cobres y el Valle de Camarones (Apata & Moraga, 2015; Eichstaedt et al., 2015; Schlebusch, Gattepaille, Engström, Vahter, & Broberg, 2015).

Los pueblos Aymara, Pehuenche y Huilliche, son grupos étnicos descendientes de los habitantes del continente americano que existieron en el altiplano andino y centro-sur de Chile desde antes de la conquista española. Existe evidencia de que esas zonas fueron habitadas por humanos hace más de 12.000 años antes del presente; tiempo suficiente para que una población se diferencie de otras y desarrolle adaptaciones al entorno local. El estudio genómico de estos grupos ofrece la posibilidad de identificar estructura genética entre ellos y de encontrar señales de selección positiva reciente, que permitirán conocer mejor la historia evolutiva humana y las características genéticas de los nativos sudamericanos.

Antecedentes

1. Origen de las poblaciones en estudio

Los pueblos nativos de Sudamérica que ocupan el continente desde tiempos prehispánicos comparten una historia migratoria: existe consenso general en que grupos de ancestría asiática que colonizaron el noreste de Siberia durante el último máximo glacial habrían cruzado hacia el continente americano en una, dos o tres oleadas, según lo propuesto por distintos investigadores (Perego et al., 2010; Francisco Rothhammer & Dillehay, 2009). Estudios recientes que incorporan información genética y arqueológica sugieren que el continente americano habría sido poblado por una sola oleada migratoria hace no más de 23.000 años atrás, que luego de un posible periodo de incubación en Siberia o Beringia, se expandió rápidamente hacia el sur (Raghavan et al., 2015).

El ingreso a América del Sur a través de Panamá, fue probablemente hace 15.000-13.500 años atrás, como lo prueban sitios arqueológicos con evidencia de ocupación humana temprana como Monte Verde, datado en 14.800 años cal. AP. (Dillehay et al., 2008). En consideración de las barreras geográficas (como el Río Amazonas, y la Cordillera de los Andes) y de la evidencia arqueológica, genética, y paleoclimática existente en esta región, se han propuesto distintas rutas migratorias luego del ingreso de humanos a Sudamérica por su extremo Noroeste. Sin embargo, no existe un modelo consensual sobre las rutas seguidas por estos primeros grupos (Figura 1), aunque varios trabajos apoyan la hipótesis de una separación en la zona norte del continente, entre un grupo que continúa hacia el sur por la costa del Pacífico y al oeste de la Cordillera de los Andes, y otro grupo que avanza hacia el este del continente (Bodner et al., 2012; F Rothhammer, Llop, Carvallo, & Moraga, 2001; Schurr, 2004).

Estudios de variación genética en microsatélites, ADN mitocondrial y cromosoma Y, revelan que existe una menor diversidad en el lado Este de Sudamérica, en contraste con el lado Oeste. Esto, junto a la evidencia de una alta semejanza entre el componente genético Andino y poblaciones Mesoamericanas, sugiere una colonización inicial de Sudamérica desde su lado Oeste, más numerosa y diversa, y que grupos de estas zonas habrían colonizado algunos sectores del territorio Este de forma posterior (de Saint Pierre et al., 2012; Wang et al., 2007). También se ha postulado que el flujo existente entre ambos lados de la cordillera fue bidireccional en varios puntos de Sudamérica a lo largo de la historia prehispánica (Bodner et al., 2012) o que hubo una migración que, desde el noroeste sudamericano, avanzó por los Andes hasta Perú, y luego pobló el Noroeste argentino, el Este Brasilero y las Pampas Argentinas, mientras que un segundo grupo migratorio, que avanzó primero hacia el este del continente, habría poblado el Amazonas brasilero para luego seguir avanzando hacia el sur y al oeste, alcanzando la vertiente oriental de los Andes hace 3.500 AP (F Rothhammer et al., 2001).

Figura 1: Posibles rutas de poblamiento Sudamericano, junto a algunos sitios arqueológicos con dataciones correspondientes al pleistoceno tardío (extraído de (Francisco Rothhammer & Dillehay, 2009)

Los primeros ocupantes del cono sur, por tanto, se encuentran en esta región desde hace varios miles de años, y luego de ocupar distintos ambientes habrían tenido poco flujo genético con otras regiones del continente. Desde la arqueología se ha observado que al menos hace 11.000-10.000 años atrás hubo un aumento en el uso de recursos costeros, concentración demográfica en torno a cuencas fluviales, y modificación de paisajes y de la distribución de plantas y animales, seguido de una diversificación tecnológica y aparición de evidencia de prácticas rituales. Todo esto sería huella de una complejización cultural que lejos de ser uniforme, habría sido altamente heterogénea dentro de Sudamérica, y

característica de cada región (Francisco Rothhammer & Dillehay, 2009), desarrollando distintas respuestas específicas a los variados hábitats que fueron ocupados.

Aymaras del Altiplano Andino

En la región del altiplano Andino se ha verificado la presencia de humanos desde al menos el pleistoceno terminal, en sitios especialmente altos como Pucuncho (4.355 msnm), datado en 12,8 – 11,5 miles de años AP, y Cuncaicha (4480 msnm), con fechados de hasta 12.4 miles de años AP. (Rademaker 2014), ambos en Perú. En Bolivia, en tanto, el sitio Cueva Bautista, ubicado a más de 3.800 msnm, fue ocupado por humanos hacia los 13.000 años cal. AP (Capriles & Albarracin-Jordan, 2013), lo cual prueba una temprana presencia humana en estas zonas altas. Los antecedentes paleoclimáticos en estas regiones indican que hacia el holoceno se produce una estabilización en las condiciones climáticas en comparación con el pleistoceno, con un periodo de aumento en las temperaturas entre 10.700 y 9.700 cal AP, y una alta aridización en el Holoceno medio, entre 6.200 y 2.300 cal AP. (Capriles y Albarracin-Jordan, 2013). Los fechados de sitios arqueológicos señalan que la ocupación de estas regiones Andinas habría sido relativamente continua, sin mayores vacíos en el registro (Gayo, Latorre, & Santoro, 2015). Los habitantes del altiplano andino se adaptaron a estos contextos, instalando en tiempos más tardíos complejos sistemas sociales: allí emergieron culturas como Chavin (2900-2200 AP), Tiwanaku (2100-800 AP), y Huari (1300-800 AP), así como una de las culturas de mayor influencia en los Andes: los Incas (Stanish, 2001). Actualmente, los grupos nativos que habitan estas zonas son mayoritariamente Quechuas y Aymaras. Sin embargo, la relación existente entre estos grupos modernos y las poblaciones que habitaron previamente el altiplano andino es desconocida, y por tanto no es posible suponer una continuidad genética o cultural entre ellos.

Los Aymaras son el segundo grupo lingüístico nativo más grande en los Andes (2.5 millones de hablantes, en Perú, Bolivia, Chile y Argentina), luego de los Quechuas (10 millones de hablantes). Se ha postulado que la lengua Aymara -o su predecesora- fue utilizada en las zonas de influencia Tiwanaku que, luego de su colapso, se redujeron a señoríos Aymaras fragmentados que persistieron hasta ser conquistados por el Imperio Inca (700-500 AP) (Gaya-Vidal et al. 2011). También se ha postulado que el origen de las poblaciones Aymaras no habría sido previo a los 1500 años atrás, sino que su aparición sería posterior a la desintegración del imperio Tiwanaku. A partir de análisis lingüísticos, tecnológicos, arqueológicos y genéticos, se ha propuesto que los habitantes nativos de los Andes tienen su origen en poblaciones de las zonas de bosques tropicales en el oeste del Amazonas (F Rothhammer et al., 2001). Contradiciendo esta hipótesis, un estudio reciente aseguró que la población Shimaa, que ocupa las yungas amazónicas al Este de los Andes, se habría separado de las poblaciones andinas hace unos 5300 años atrás (Scliar et al., 2014), y que representarían un subset de la diversidad genética existente en las poblaciones Andinas, de las cuales provienen. Se señala, sin embargo, que no es seguro extrapolar estos resultados a las demás poblaciones del Este de los Andes, y por tanto, la procedencia y fecha de origen de los Aymaras sigue debatiéndose.

La ocupación del altiplano, a veces a más de 4000 msnm, implica la adecuación a un entorno con hipoxia (menor disponibilidad de oxígeno); con un alto costo energético de subsistencia (requerimiento calórico duplicado respecto al necesario a nivel del mar); con alta radiación solar, y bajas temperaturas (Rupert & Hochachka, 2001). Además, en la región de Puno (15º50' S), donde habitan los Aymaras que serán evaluados en este estudio (Figura 2), existen también altos niveles de arsénico y otros metales tóxicos en las aguas de consumo, debido a su origen cordillerano (George et al., 2014). Habitar esta zona representa dificultades que podrían haber impulsado adaptaciones de tipo cultural y biológico, que han sido estudiadas desde distintas aproximaciones (A. W. Bigham et al., 2013; Francisco Rothhammer et al., 2015; Valverde et al., 2015).

Figura 2: Localización de las poblaciones en estudio. Los Aymaras provienen de la Región de Puno, Perú, los Pehuenches de la Región del Bío-Bío, Chile, y los Huilliches de la Región de los Lagos, Chile.

Poblamiento del Centro-Sur de Chile

El Sur de Chile, entre el río Maule y el norte de la Isla Grande de Chiloé (35º30' – 42º S), cuenta con un registro arqueológico extenso en cuanto a su cronología. Los primeros sitios con evidencia de intervención humana en la zona están datados hacia fines del pleistoceno (entre 15.000-11.500 AP) y representarían una ocupación intermitente y de

baja densidad poblacional. En este período, el clima de la zona cambiaba hacia temperaturas más templadas con una menor cantidad de lluvias, acompañado por un progresivo retroceso de los glaciares, la extinción de megafauna, y un aumento en la actividad volcánica. En el holoceno temprano y medio (10.000-6.000 AP) se presentan condiciones secas y cálidas, y desde el año 6.000 AP se avanza hacia condiciones climáticas más húmedas, similares a las actuales (Abarzúa et al., 2014).

Las ocupaciones prehistóricas registradas en el territorio están desigualmente distribuidas a nivel cronológico y geográfico entre la costa, los valles, y la cordillera. Durante el arcaico temprano, tanto en los valles como en la cordillera, hay registro de ocupación por grupos altamente móviles. En la zona costera, desde el arcaico intermedio (8.000 a 3.950 cal. AP), se observa una tecnología lítica variada y especializada en recursos marinos, que evidencia una ocupación sólida y continua de la zona. En los valles, por otro lado, hay ocupaciones muy esporádicas entre los 12.000 y los 3.000 AP, aunque esto podría ser reflejo de un efecto arqueodemográfico; es decir, podría ser un artefacto de la información arqueológica disponible (R. Campbell & Quiroz, 2014). En la Cordillera, los primeros signos de ocupación también aparecen hacia los 12.000 AP, seguidos por sitios habitados entre los 9.500 y los 2.200 AP. Luego de esta fecha, que coincide con los primeros hallazgos de cerámica en los tres sectores –costa, valle y cordillera-, hay continuidad en el poblamiento de estas localidades hasta periodos históricos (R. Campbell & Quiroz, 2014).

En el periodo alfarero aparece el complejo cultural Pitrén, asociado a grupos cazadores de animales pequeños, recolectores de plantas salvajes, y con algunas practicas hortícolas, según se puede inferir en sitios cordilleranos. El complejo Pitrén se hace presente con sus innovaciones alfareras hacia el año 300 d.C. en zonas lacustres cordilleranas, y hacia el 430 d.C. en sectores costeros (Adán & Mera, 2011).

Una tradición cultural más tardía del periodo alfarero fue el complejo El Vergel, que aunque pudo recibir influencias desde el norte, la base de su estilo cerámico está en el complejo Pitrén (Aldunate del S., 1993). El complejo El Vergel se presenta en la costa y valles del sector norte del sur de Chile (al norte del Río Toltén) durante el arcaico tardío (1000-1550 d.C), y ha sido asociado a habitantes más sedentarios, con una organización social basada en cacicazgos o grupos igualitarios, que usaban recursos cultivados como la quínoa y el maíz manteniendo también prácticas de caza y recolección. La emergencia de la metalurgia para producir joyas, y la diversificación de las prácticas mortuorias con el uso de urnas de cerámica, troncos ahuecados, y los clásicos montículos Mapuche, son señales de una posterior complejización del modelo social (R. J. Campbell, 2011).

De acuerdo a estimaciones etnohistóricas, hace 500 años la población que habitó el sur de Chile alcanzaba un millón de habitantes, y se estima que 50 años después, la población se redujo a 100.000 individuos (Bengoa, 1996). Los primeros registros sugieren que esta región compartía una lengua, creencias religiosas, y tenían una gran capacidad para articular alianzas locales y regionales (R. J. Campbell, 2011), aunque los distintos etnógrafos y lingüistas proponen agrupaciones y características distintas entre los

conjuntos de habitantes que se encontraban en la zona en los periodos post-hispánicos (Urbina, 2009).

En este trabajo son estudiados los pueblos Pehuenche y Huilliche, dos grupos étnicos que habitan el sur de Chile (Figura 2). Durante el siglo XIX, el Estado chileno y argentino aplicó políticas que promovieron la expulsión de los pueblos indígenas a terrenos reducidos y poco productivos. A pesar de esto, comunidades de Pehuenches y Huilliches se caracterizan por conservar su unidad cultural.

Pehuenches

Actualmente, los grupos reconocidos como Pehuenches ocupan el sector de la Cordillera en la Región del Bío-Bío. Antiguamente eran grupos nómades, que transitaban con sus ganados por los territorios de la pampa argentina y los valles chilenos, hasta que fueron expulsados y reducidos a fines del siglo XIX por el Estado chileno (Bengoa, 1996), Algunos lingüistas y etnógrafos afirman que formaron parte de los Mapuches del centro sur, pero que su especialización al entorno y a la recolección del pehuén habría promovido también una diferenciación cultural, distanciando incluso su lengua del Mapudungún (Torrejón, 2001). La actual comunidad de Trapa-Trapa, en la comuna de Santa Bárbara, ha mantenido un relativo aislamiento de la sociedad chilena hasta los años 80', pues sólo se podía acceder al valle a caballo o en helicóptero (Llop et al., 1993). Estos individuos ocupan tierras aptas para la agricultura, y uno de los principales recursos aún utilizados es el piñón.

Huilliches

El término Huilliche, usado por los indígenas del norte del río Toltén para denominar a aquellos que habitaban al sur de este punto, fue registrado por los españoles desde el siglo XVII (Alcaman, 1994). Los Huilliches actuales habitan en comunidades de la Región de los Lagos y Región de los Ríos, en torno a ciudades como Osorno, Valdivia, y Quellón. Los datos genómicos aquí analizados provienen de una comunidad que habita en San Juan de la Costa, parte de la Provincia de Osorno, en la selva valdiviana costera. Su relieve corresponde a un cordón costero de baja altura, con altas precipitaciones y densidad vegetal. Hasta principios del siglo XIX esta zona estaba casi exclusivamente habitada por grupos indígenas. Se cree que los habitantes actuales de la zona, son descendientes de antiguos Huilliches provenientes de los valles y la cordillera, que se vieron forzados a avanzar hacia los territorios de baja productividad con la llegada de los colonos (Quiroz y Olivares, 1987).

2. Procesos microevolutivos en las poblaciones humanas

La microevolución se define como el cambio en las frecuencias de las variantes genéticas (alelos) dentro de una población a través del tiempo. Este cambio se produce luego de varias generaciones por efecto de distintos factores evolutivos, como la selección natural, el apareamiento selectivo, la acumulación de mutaciones, las migraciones, o la deriva génica. Se diferencia de la macroevolución en que este último concepto refiere al estudio de los procesos evolutivos entre especies (Reznick & Ricklefs, 2009).

Gracias a la acción de uno o varios de los factores evolutivos, cuando hay una subdivisión poblacional es casi inevitable que se produzca algún grado de diferenciación genética entre las subpoblaciones que se originan. El concepto de diferenciación genética se refiere a la adquisición de frecuencias alélicas que difieren entre dos grupos de organismos. Con el paso del tiempo, cada población acumulará variaciones genéticas de manera independiente, aumentando la diferenciación entre ellos (Hartl & Clark, 1997). En este trabajo serán estudiados los efectos poblacionales de la deriva génica, migración y selección natural a través de una aproximación genómica.

Deriva génica

Al producirse una subdivisión poblacional en dos grupos de bajo número de individuos, es posible que ocurra un cambio azaroso en las frecuencias alélicas producto del muestreo de variantes génicas que conformarán la siguiente generación. Luego de algunas generaciones, estos cambios pueden producir la fijación (aumento a una frecuencia alélica de 1) de alelos distintos en cada subpoblación. La magnitud de este cambio es inversamente proporcional al tamaño de la población, y no se produce de forma dirigida. Este proceso se denomina deriva génica, y puede generar divergencia genética entre poblaciones en un tiempo relativamente breve (Hartl & Clark, 1997).

Todas las regiones genómicas tienen la misma probabilidad de ser afectadas por deriva génica, dado que no se relaciona con el efecto que éstas puedan tener en las características de los individuos o en su adaptabilidad biológica. Por lo tanto, se espera que la deriva afecte a todo el genoma de forma homogénea.

Migraciones

La migración produce cambios en las características genéticas de una población en dirección opuesta a la deriva génica. Al producirse movimientos migratorios entre dos grupos con divergencia genética, se produce flujo génico que puede reducir la diferenciación previamente acumulada por cada población de forma independiente. Con este proceso, cambian las frecuencias alélicas de la población que recibe a los inmigrantes, y aparecen individuos mestizos con patrones de variación genética que los asocian tanto con la población local como con la inmigrante.

La magnitud del cambio a nivel genético dependerá del coeficiente de migración entre grupos. Si es alto, es decir, si hay una alta proporción de gametos aportada por los inmigrantes en relación a la población local, se producirá una homogeneización entre los grupos (Fix, 1999).

Selección natural

La selección natural es, según Darwin, el principal mecanismo de transformación y diversificación evolutiva de las especies (Darwin, 1859). Su acción está basada en que, dentro de una población, los organismos difieren en su capacidad de sobrevivir y reproducirse en un determinado ambiente; es decir, tienen distinta adaptabilidad biológica o *fitness*. Aquellos individuos que portan genotipos beneficiosos heredarán estas características ventajosas a un mayor número de descendientes respecto a quienes no cuentan con ellos, cambiando las frecuencias alélicas de la población en las siguientes generaciones. La selección natural tiene un rol importante dentro del proceso evolutivo, pues en muy pocas generaciones puede cambiar notablemente las proporciones alélicas en una población (Hartl & Clark, 1997).

Existen distintas formas de operar de la selección natural: en primer lugar, la selección negativa o purificadora consiste en la eliminación de alelos deletéreos; es decir, que son desventajosos para los individuos que las porten. En segundo lugar, la selección estabilizadora permite que se conserven varios alelos de un locus (posición en el genoma) de manera equilibrada y en frecuencias intermedias. En tercer lugar, la selección positiva favorece una variante genética que resulta ventajosa para la sobrevivencia y reproducción de sus portadores (Wollstein & Stephan, 2015), aumentando rápidamente su frecuencia poblacional. En este caso, se reduce la variación genética no sólo en el locus seleccionado, sino también en las zonas cercanas a este, produciendo un barrido selectivo (*selective sweep*) que aumenta la frecuencia de un haplotipo y genera una región con menos recombinación que bajo neutralidad (Figura 3). Los barridos selectivos conllevan, por tanto, la aparición de segmentos cromosómicos con menores niveles de diversidad, y altamente diferenciados en sus frecuencias alélicas respecto a otras poblaciones.

A diferencia de la deriva, la selección positiva es un proceso dirigido que tiene directa relación con el impacto de los genes en la adaptabilidad biológica de los individuos. Esto produce señales que son reconocibles en la estructura del ADN de una población, y por tanto, permiten distinguir las zonas bajo selección positiva al compararlas con el resto del genoma como referencia de neutralidad selectiva.

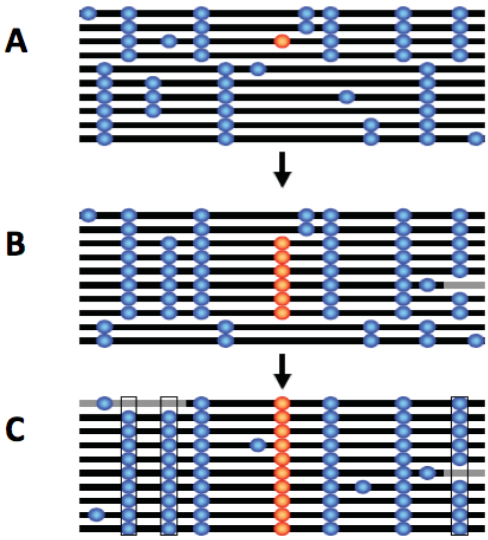

Figura 3: Barridos selectivos. Al aparecer una nueva variante beneficiosa en un individuo de la generación 0 (A), esta es seleccionada positivamente produciendo un barrido selectivo parcial (B) que puede llegar a convertirse en un barrido selectivo completo (C). Imagen modificada de (Hancock & Di Rienzo, 2008).

Selección positiva en humanos

El ser humano moderno ha enfrentado modificaciones importantes en su entorno durante los últimos 50.000 años: la transición climática desde el último máximo glacial a las condiciones modernas, las migraciones desde África y la ocupación de nuevas geografías, sumados a cambios en los modos de subsistencia de cada grupo, impulsaron la emergencia de nuevos agentes selectivos, particulares a cada zona geográfica y estilo de vida (Jeong & Di Rienzo, 2014). Esto supone la aparición de respuestas adaptativas tanto culturales como biológicas a nivel local.

La mayoría de las adaptaciones biológicas estudiadas en profundidad a la fecha han sido identificadas al intentar explicar diferencias fenotípicas evidentes entre poblaciones (pigmentación de la piel, por ejemplo). Ante la hipótesis de que los fenotipos predominantes sólo en algunas poblaciones tendrían un rol adaptativo, los genes involucrados en esta respuesta pueden ser identificados comparando la divergencia genética en los loci candidatos a estar bajo selección. Si la divergencia en estas regiones es mayor a la esperada bajo un modelo que asume neutralidad, se sugiere que la selección positiva podría ser la causal. (“The Neutral Theory and Tests of Neutrality,” 2013). Recientemente, el desarrollo de tecnologías de genotipificación y secuenciación masivos han permitido el desarrollo de aproximaciones genómicas que permiten buscar regiones bajo selección positiva que no dependen de la formulación de hipótesis respecto de genes específicos ni requieren de información fenotípica.

Estudios centrados en encontrar señales de selección asociadas a una presión selectiva específica han analizado adaptación a las bajas temperaturas (Cardona et al., 2014), a la alta radiación UV (Jablonski & Chaplin, 2010), a elementos tóxicos como el arsénico (Schlebusch et al., 2015), al déficit de micronutrientes como el selenio (White, Erlebach, Weihmann, Aida, & Castellano, 2015), entre varios otros factores. Los ejemplos de adaptación con mayor evidencia, como la persistencia de lactasa, la pigmentación de la piel, y la adaptación a la altura, representan adaptaciones desarrolladas paralelamente en distintas poblaciones ante presiones selectivas similares, pero por mecanismos genéticos distintos.

El gen LCT, que codifica la enzima lactasa, es uno de los casos mejor estudiados de selección positiva reciente en humanos. La lactasa es una enzima que permite la digestión de la lactosa –principal azúcar de la leche- en el intestino delgado. En la mayoría de los seres humanos la producción de la lactasa disminuye después del destete, mientras que en algunas poblaciones hay un alto porcentaje de individuos que siguen produciéndola durante la adultez. A esta condición se le domina “persistencia de la lactasa”. Estas poblaciones tienen en común que han sido tradicionalmente pastoralistas y consumidores regulares de leche (Simoons, 1969). La persistencia de lactasa es una característica que otorga ventajas como una ingesta calórica eficiente, asimilación de calcio (muy favorable en zonas de baja radiación solar), y absorción de agua desde la leche en ambientes áridos (Jeong & Di Rienzo, 2014). Este caso es especialmente importante, debido a que evidencia la existencia de adaptaciones muy recientes en humanos -entre 3.000 y 7.000 años atrás (Hancock & Di Rienzo, 2008)- y ejemplifica una relación co-evolutiva entre genes y cultura.

En la literatura científica encontramos otras regiones del genoma y variantes génicas que han sido propuestas como seleccionadas positivamente en ciertas poblaciones. Sin embargo, para la mayoría de estas regiones candidatas no ha sido posible encontrar genes con evidente relación funcional a algún fenotipo, ni comprobar la historia evolutiva del rasgo como en el caso de la persistencia de lactasa (Grossman et al., 2014). Se estima, sin embargo, que la mayoría de las regiones genómicas que presentan huellas de selección en seres humanos, estarían asociadas a la respuesta a patógenos (Fumagalli et al., 2011).

3. Aproximación genómica al estudio de procesos microevolutivos

En los últimos años, el estudio de los distintos factores evolutivos y su efecto en el genoma humano ha crecido velozmente gracias al desarrollo de nuevas tecnologías que permiten secuenciar o genotipificar genomas de gran tamaño a bajo costo. En consecuencia, se han desarrollado numerosos estudios genético-poblacionales y de asociación del genoma completo (GWAS), cuyos datos son depositados en bases de datos públicas con información genética de miles de individuos de distintas poblaciones. La alta disponibilidad de datos ha impulsado el desarrollo de nuevas herramientas

computacionales y estadísticas que permiten hacer estudios microevolutivos a partir de información genotípica de cientos de miles de *SNPs* (por su sigla en inglés *Single Nucleotide Polymorphism*, o polimorfismos de un solo nucleótido) (Wollstein & Stephan, 2015). En la especie humana, se ha investigado la estructuración genética entre grupos continentales (Weir, Cardon, Anderson, Nielsen, & Hill, 2005a), se ha impulsado la reconstrucción de la historia demográfica de distintas poblaciones (Holsinger & Weir, 2009) y también se ha estudiado el impacto de la selección natural en la diversidad genética existente en la especie humana (Barreiro, Laval, Quach, Patin, & Quintana-Murci, 2008).

Estructura poblacional

Al producirse estructura poblacional, se acumula diferenciación genética entre las subpoblaciones que puede ser medida a nivel genómico. El índice de fijación *Fst* es una medida de diferenciación genética que permite evaluar la magnitud de la estructuración genética a partir de las frecuencias alélicas observadas en miles o millones de sitios polimórficos. Este índice representa la probabilidad de que distintas poblaciones con un origen común fijen alelos alternativos luego de haberse separado. Su cálculo se basa en la proporción de la varianza total de las frecuencias alélicas que es atribuible a la estructura poblacional, es decir a la diferencias entre las poblaciones.

La fórmula de Fst de Weir and Cockerham (1984) se encuentra entre los estadísticos más usados en los estudios de genética de poblaciones y evolución y su cálculo no es sesgado en bajos tamaños muestrales (Holsinger & Weir, 2009). Para la interpretación de sus valores, se considera que las frecuencias alélicas en dos poblaciones son semejantes si el *Fst* es pequeño (entre 0 y 0,05). Con valores entre 0,05 y 0,15, se indica una diferenciación moderada entre los grupos; entre 0,15 y 0,25 se señala una marcada diferenciación, mientras que valores sobre 0,25 evidencian una gran diferenciación genética (Hartl & Clark, 1997).

En nuestra especie, al calcular *Fst* entre poblaciones de Nigeria, China, Japón, y del noreste de Europa en cromosomas autosomales (The international HapMap 3 Consortium, 2010) se obtienen valores de *Fst* promedio entre 0,10 a 0,15 dentro de cada grupo, y entre todos, un valor de 0,13 (Weir, Cardon, Anderson, Nielsen, & Hill, 2005b), aunque cuando se incorporan datos de otras poblaciones humanas y una mayor cantidad de individuos, se obtienen valores más bajos, cercanos a 0,05 (Auton et al., 2009). Entre grupos continentales, se han observado valores de Fst de 0,157 entre europeos y africanos, 0,185 entre africanos y asiáticos, y 0,11 entre asiáticos y europeos (The international HapMap 3 Consortium, 2010).

El test Fst permite revelar diferencias entre poblaciones que se deban a procesos evolutivos ocurridos hasta hace 75.000 años atrás, luego de que el ser humano moderno migra desde África hacia distintos continentes, exponiéndose a nuevos ambientes y

disminuyendo o anulando el flujo génico con las poblaciones de las que se separó (P C Sabeti et al., 2006).

Este estadístico también se ha usado, en conjunto con otras pruebas estadísticas, para analizar patrones de diferenciación en genomas completos, proponiendo regiones bajo selección positiva (Barreiro et al., 2008). Si el valor de *Fst* en un locus es notablemente mayor en una región que en el resto del genoma, podemos sugerir que está siendo seleccionado positivamente en una de las poblaciones (Holsinger & Weir, 2009). Con esto, se ha encontrado una alta coincidencia entre regiones de haplotipos largos y valores de Fst altos identificados en humanos.

Inferencias de ancestría en poblaciones indígenas

Cuando se produce migración entre poblaciones reproductivamente aisladas, se producen cambios en la estructura poblacional que pueden ser identificados a nivel genómico. La población que recibe a los migrantes tendrá individuos mestizados, con algunas porciones genómicas que provienen de la población local y otras de los inmigrantes. En tales individuos, se puede estimar el porcentaje de su genoma que es originario de cada población ancestral a partir de sus genotipos a lo largo del genoma. Este tipo de análisis se denomina inferencia de ancestría global, dado que caracteriza ancestría a nivel genómico pero no informa sobre el origen de un cromosoma o locus en particular. El conocimiento de los componentes ancestrales y sus proporciones dentro de una muestra es de alta relevancia al realizar estudios de asociación, en la identificación forense de individuos, y en la evaluación de procesos evolutivos o históricos que influyan en la conformación de una población (Y. Liu et al., 2013).

Usualmente, la definición de poblaciones se realiza a partir de características físicas, lingüísticas, culturales, o por la ubicación geográfica de los individuos muestreados. Todas estas aproximaciones son altamente subjetivas, y pueden no dar cuenta de la estructura poblacional identificada a partir de datos genéticos. Esta última puede ser reconocida utilizando herramientas computacionales que implementan distintas aproximaciones estadísticas (Y. Liu et al., 2013). Los programas más utilizados actualmente se basan en la metodología de agrupamiento basada en modelos propuesta por Pritchard y colaboradores (J K Pritchard, Stephens, & Donnelly, 2000). Esta aproximación tiene la capacidad de inferir la estructura poblacional a partir de datos de genotipos de muchos locus, asumiendo que existe un número K de poblaciones en los datos analizados. El programa ADMIXTURE implementa esta metodología, con algunas modificaciones que permiten que sea más rápida e igualmente precisa que la implementación original (Alexander & Lange, 2011).

Identificación de procesos biológicos seleccionados

Junto al desarrollo de distintos test para identificar selección positiva, hoy existen aproximaciones para evaluar los resultados de estas herramientas que van más allá de la identificación de los genes ubicados en regiones con señales de selección. Últimamente, se ha destacado la relevancia de evaluar si los genes ubicados en las regiones seleccionadas participan o no mayoritariamente en alguna vía metabólica o proceso biológico (Jonathan K Pritchard & Di Rienzo, 2010). Esto estaría relacionado con la naturaleza poligénica de algunas adaptaciones, que si bien representa mayores dificultades para ser identificada en contraste con mecanismos genéticos más sencillos, es como usualmente funciona la selección positiva: afectando fenotipos asociados a muchos genes (Daub, Dupanloup, Robinson-Rechavi, & Excoffier, 2015; Jeong & Di Rienzo, 2014).

Los procesos biológicos en los que está involucrado un grupo de genes pueden ser identificados usando la base de datos de ontología génica (Reference Genome Group of the Gene Ontology Consortium, 2009) y herramientas que gestionan esta información para hacer pruebas estadísticas de sobre-representación o enriquecimiento, como PANTHER (Mi, Muruganujan, Casagrande, & Thomas, 2013). Este proyecto tiene por objetivo describir cada gen y sus productos biológicos usando un estricto vocabulario, y agrupándolos en familias y subfamilias según los procesos biológicos, componentes celulares, o funciones moleculares a las que están asociados.

Problema de investigación

Las poblaciones nativas americanas tienen una extensa historia ocupacional en el Sur del continente de más de 14.000 años AP (Francisco Rothhammer & Dillehay, 2009). A lo largo de este tiempo, han desarrollado adaptaciones locales que han diversificado la forma en que conviven con el entorno. Las herramientas y recursos materiales de los que hicieron uso para sobrevivir en sus respectivos ambientes pueden ser identificados a partir del registro arqueológico (Borrero, 2015). Las adaptaciones biológicas, en cambio, pueden haberse producido sin dejar huellas en los restos óseos humanos. Sin embargo, señales de selección natural podrían ser identificadas mediante el análisis de patrones de variación genética presentes en los genomas de estas poblaciones nativas.

En el caso de los Aymara, la altura y los tóxicos ambientales representan presiones selectivas conocidas en el altiplano andino. En las poblaciones Pehuenches y Huilliches, por otra parte, es más difícil postular factores selectivos fuertes, pues no es evidente la presencia de condiciones ambientales extremas. Sin embargo, estudios recientes han propuesto el desarrollo de adaptaciones a condiciones húmedas (Hancock et al., 2011), al frío en poblaciones de Siberia (Cardona et al., 2014), y al déficit de selenio en la dieta en poblaciones de China (White et al., 2015). Por lo tanto, es razonable pensar que hayan existido presiones selectivas desconocidas en las poblaciones del centro sur chileno, que promovieron la aparición de adaptaciones heredables a lo largo de su historia evolutiva. También es factible que la población ancestral que dio origen a los tres grupos en estudio haya sufrido adaptaciones por presiones selectivas anteriores al momento de su divergencia, produciendo señales comunes en los genomas de los tres grupos.

Una aproximación genómica permitiría identificar características de la diversidad genética en estas poblaciones que son producto de la microevolución, y en particular de la selección, sin necesidad de explicitar hipótesis *a priori* sobre los fenotipos o genes que pudieran estar o haber estado seleccionados. Este tipo de estudio entrega información no sólo son de interés evolutivo, sino que también biomédico e histórico.

Hipótesis y objetivos

Hipótesis

Las poblaciones Aymaras, Pehuenche y Huilliche sufrieron procesos adaptativos recientes que las diferencian a nivel genómico.

Objetivos

- Objetivo general:

Determinar la presencia de señales de selección positiva en el componente amerindio del genoma de individuos Aymara, Pehuenche y Huilliche

- Objetivos específicos:

1. Identificar la estructura genética de poblaciones Aymara, Pehuenche y Huilliche
2. Determinar evidencia de selección positiva ocurrida en el continente Americano
3. Determinar si las poblaciones del altiplano andino y del centro-sur de Chile sufrieron selección positiva posterior a su divergencia.

Material y métodos

Poblaciones en estudio

Se estudió el perfil genético de adultos sanos, de ambos sexos, que reconocieran ascendencia o pertenencia Aymara, Pehuenche o Huilliche. Las muestras consideradas como "Aymara" en este trabajo, corresponden a 106 individuos peruanos con al menos un abuelo de ancestría Aymara, y fueron obtenidas en la región de Puno, ubicada en la zona sureste de Perú, durante noviembre de 2013 por el laboratorio del Dr. Carlos Bustamante de la Universidad de Stanford, quienes autorizaron el uso de estas muestras para esta memoria de título. La extracción de ADN se hizo a partir de una muestra de 5 ml de sangre. El procesamiento y control de calidad de las muestras se realizó en Stanford siguiendo protocolos estándar. Todos los individuos consintieron informadamente su participación en el estudio, el cual fue aprobado por el Comité de Ética de la Universidad de Stanford (protocolo 20839).

Las muestras de los individuos Huilliches fueron obtenidas en la localidad de Misión, en la comuna de San Juan de la Costa, Región de los Lagos, Chile, en el año 1991. Esta comuna es reconocida por el gran número de comunidades Huilliches que habitan en la zona. En el mismo año, se obtuvieron las muestras de Pehuenches, que provienen de la comunidad Pehuenche Trapa-Trapa, en la comuna de Alto Bío Bío, de la Región del Bío Bio, Chile. En este trabajo, usaremos muestras de 20 Huilliches y 15 Pehuenches. De estas, se extrajo ADN desde 5 ml sangre. El análisis de estas muestras se realizó con la aprobación del comité de ética de la Facultad de Medicina de la Universidad de Chile.

Generación de perfiles genético moleculares

Todas las muestras fueron genotipificadas con el microarreglo "*Axiom Human Genome-Wide LAT 1" (World Array 4*) de Affymetrix (Hoffmann et al., 2011) en el *Institute for Human Genetics, University of California*, San Francisco, California, obteniendo información de un total de 817.810 *SNPs*, distribuidos homogéneamente a lo largo del genoma humano.

Control de calidad de los datos

Tal como sugiere Affymetrix, se filtraron las muestras con un valor de "DQC" (Dish quality control) menor a 0,82, y luego se eliminaron los marcadores que presentaban baja calidad con el paquete en R llamado SNPolisher (Nicolazzi, Iamartino, & Williams, 2014) y el programa PLINK versión 1.07 y 1.09 (Purcell et al., 2007). Se conservaron las muestras con más del 95% de los datos y los *SNPs* con información en más del 98% de las muestras. Se eliminaron marcadores duplicados, no autosomales, y aquellos con un valor

de *p* menor a 0,000001 al testear por equilibrio de Hardy-Weinberg, pues estos valores se suelen asociar a errores en la genotipificación (The international HapMap 3 Consortium, 2010). Para evitar el uso de muestras de individuos cercanamente emparentados, sólo se utilizaron los individuos que, al compararlos con todas las muestras del set, tuvieran un valor de IBD <0,3 (identidad por descendencia). Este valor se refiere a la proporción de segmentos del genoma que son idénticos entre individuos, por compartir algún nivel de parentesco.

Estructura genética poblacional

Una primera aproximación a la estructura genética poblacional de la muestra se realizó con el programa ADMIXTURE (Alexander & Lange, 2011). Este permite estimar la proporción de ancestría genética de cada población ancestral (presente en el set de datos utilizado) en cada individuo asumiendo un número hipotético de poblaciones ancestrales que contribuyeron alelos a la muestra genotipificada. Como poblaciones de referencia se incorporaron datos del Proyecto 1000 Genomas (The 1000 Genomes Project Consortium, 2012): 30 individuos Africanos de ancestría Yoruba en Nigeria (YRI), 30 residentes de Utah, EE.UU, con ancestría del Norte y Oeste de Europa (CEU), y 30 individuos de Beijing, China (EAS). Dado que ADMIXTURE asume independencia de individuos y de *SNPs*, para este análisis se eliminaron los marcadores que presentaron una alta correlación por desequilibrio de ligamiento, medido por un $r^2 \geq 0,1$ en ventanas de 50 *SNPs*, dejando solo un marcador por bloque de ligamiento.

Al utilizar el programa ADMIXTURE, es necesario asumir un número de poblaciones ancestrales (K) cuyos componentes genéticos se encuentran presentes en los grupos incorporados en el panel evaluado. En este trabajo se realizaron las inferencias de ancestría global usando valores de K de 3 a 7, y posteriormente, se calculó el error de validación cruzada para conocer qué valor de K tiene mayor capacidad predictiva sobre los datos observados. Este cálculo consiste en enmascarar 1/5 de los genotipos observados por turnos, usando el resto del set de datos para inferir las frecuencias alélicas de cada población y las porciones genómicas contribuidas por cada población a cada individuo. Dados estos estimadores, se realiza una predicción de los genotipos enmascarados, y se elige el valor de K que presente el menor error en la predicción. Tanto la inferencia de ancestría global como el error de validación cruzada fueron calculados con ADMIXTURE y graficados en R (R Core Team, 2013).

Utilizando los resultados de ADMIXTURE, fueron seleccionados los individuos con más del 95% de ancestría amerindia para el resto de los análisis. Se calcularon los valores de *Fst* (Weir & Cockherman) ponderados entre todas las poblaciones utilizadas en el análisis de ancestría global con el programa *Vcftools* (Danecek et al., 2011) a partir de archivos en formato VCF ("*variant call format*"), que contiene la información de los genotipos por individuo. A partir de una matriz con las estimaciones de *Fst* y usando el programa MEGA

(Tamura, Stecher, Peterson, Filipski, & Kumar, 2013), se construyó un dendrograma con el método de agrupamiento UPGMA (Sokal & Michener, 1958).

Ajuste de fase

Este proceso consiste en la reconstrucción de las combinaciones alélicas existentes en cada cromosoma a partir de los datos de genotipado, pues estos últimos sólo nos comunican los genotipos en cada posición y, cuando estos son heterocigotos, no se sabe de qué cromosoma proviene cada alelo. Este procesamiento es necesario para aplicar los test de selección basados en frecuencias de haplotipos. Se usó el programa SHAPEIT2 que es altamente preciso cuando se usan paneles de referencia numerosos (O'Connell et al., 2014). SHAPEIT2 fue utilizado incorporando un mapa genético del ensamble GRCh37 de NCBI, y un archivo con la información de todos los individuos a analizar y del panel de referencia, en el formato de archivos utilizados por el programa PLINK. Debido a que SHAPEIT2 aprovecha la información de parentesco entre individuos con una relación de padre-hijo para identificar segmentos compartidos entre las muestras, se incorporaron como referencia 35 tríos (es decir, muestras de padre, madre, e hijo/a) del proyecto ChileGenomico[1], además de los siguientes individuos del proyecto 1000 Genomas: Mexicanos (de los Ángeles, EEUU), Colombianos (de Medellín, Colombia), Puertorriqueños y Peruanos (Lima, Perú). Este gran panel de referencia (452 individuos) se incorpora en consideración de que con un mayor número de muestras es más precisa la inferencia de haplotipos.

Test de selección positiva

Todos los test de selección fueron aplicados en individuos con evidencia molecular de alta ancestría amerindia (>95%), según los resultados de ancestría global producidos por ADMIXTURE. Se calculó el puntaje de iHS por marcador para las poblaciones nativas americanas Aymaras (AYM) y Pehuenches-Huilliches (PH) de forma separada, y también para Aymaras, Pehuenches y Huilliches considerados como un solo grupo (AMR). Además, se calculó $XPEHH$ entre los grupos AYM y PH, y también entre el grupo AMR y asiáticos (EAS), europeos (EUR) y africanos (AFR). Es importante señalar que ni iHS ni XPEHH confieren una prueba estadística, pues no se conoce la distribución de los datos bajo condiciones neutrales. Los cálculos se realizaron con *selscan* (Szpiech & Hernandez, 2014), que utiliza un archivo PLINK en formato .tped (genotipos transpuestos) con posiciones genéticas (extraídas desde el mapa genético usado en el ajuste de fase) por cada cromosoma. Luego de calcular iHS y $XPEHH$ no normalizados, el programa *norm* (Szpiech & Hernandez, 2014) fue utilizado para estandarizar los estadísticos a partir de los archivos de salida de *selscan*. La normalización se hizo separando los marcadores en 100 grupos de acuerdo a sus frecuencias alélicas, usando la desviación estándar y el promedio por grupo.

[1] http://chilegenomico.uchile.cl

El calculo de iHS y $XPEHH$ se realizó con los valores por defecto de los parámetros para selscan tal como en Voight et al. 2006: no se realizaron cálculos para los marcadores que estuviesen a más de 200 kb de distancia con otro marcador, el cálculo de EHH desde un *SNP* central se detiene al llegar a 0,05, y en caso de que no hubiesen marcadores a más de 20 pb, se escaló la distancia genética multiplicándola por la razón entre 20 y el tamaño del intervalo sin marcadores, para reducir señales espurias inducidas por una baja densidad de datos. Para el test iHS, se eliminaron los marcadores cuyo alelo de menor frecuencia fuera menor a 0,05.

De forma complementaria, se calcularon los valores de *Fst* (Weir & Cockherman) por marcador y por ventana de 200 kb con *Vcftools,* entre todos los pares formados por los grupos de ascendencia Europea (CEU), Asiática (EAS), Africana (YRI), Aymara, Pehuenche, y Huilliche. También se calculó Fst entre los nativos americanos del altiplano andino (Aymaras, AYM), del centro-sur chileno (Pehuenches y Huilliches, PH).

Genes bajo selección positiva

Los resultados obtenidos fueron manipulados en el programa R (R Core Team, 2013), agrupando los puntajes por marcador en ventanas de 200 kb. Dado que las señales de barridos selectivos se expanden unos 0,3-0,5 cM, utilizar ventanas de este tamaño (equivalentes en promedio a 0,2 cM) y con un mínimo de 20 marcadores permitiría captar la señal con iHS manteniendo una densidad de datos adecuada (Voight, Kudaravalli, Wen, & Pritchard, 2006). Ventanas de otros tamaños y el uso de ventanas solapadas producen resultados similares (Pickrell, Coop, & Novembre, 2009). Mediante simulaciones, se ha demostrado que se tiene mayor poder de detección de regiones seleccionadas si se consideran las regiones que agrupen varios marcadores con valores altos de iHS, en vez de los valores por *SNP* individual (Voight et al., 2006). Para el test XPEHH, en cambio, se sugiere que es más conveniente utilizar aquellas regiones donde se encuentre un valor alto, sin considerar los puntajes de los marcadores cercanos (Pardis C Sabeti et al., 2007; Voight et al., 2006). Por esto, en este trabajo se consideran como regiones seleccionadas positivamente a aquellas que se encuentran dentro del 1% de las ventanas de 200 kb con mayor proporción de marcadores con valores altos (mayores a 2, según el valor de corte utilizado en otros estudios) en el caso del test iHS. Para XPEHH y Fst, se usó el marcador con puntaje más alto como el puntaje de la ventana, y el 1% de las ventanas con más alto puntaje fue considerado como seleccionado positivamente. El uso del 1% es un valor de corte conservador en comparación con el utilizado en otras publicaciones (A. W. Bigham et al., 2009; A. Bigham et al., 2010), pero garantiza la identificación de regiones con alta evidencia de selección. En cada región candidata, se identificaron los genes ahí presentes a partir de la base de datos de "*Known Canonical Genes"* (21.680 genes a nivel genómico) publicada en el *Genome Browser* de UCSC (Hsu et al., 2006), la cual fue creada a partir de la agrupación de múltiples entradas para el mismo locus de un gen, desde la base de datos de Genes Conocidos ("*Known Genes*", 44.338 datos transcriptos)

Identificación de procesos biológicos bajo selección

Con la herramienta PANTHER (Mi et al., 2013), se probó si el listado de genes presentes en el 5% de las ventanas con mayores puntajes dentro de cada test estaban sobre-representando o sub-representando algún proceso biológico, en comparación con los genes presentes en todas las ventanas utilizadas. En este análisis consideramos significancia estadística para los procesos biológicos con un valor de $p < 10^{-4}$.

Fenotipos asociados a genes en regiones bajo selección

Fue utilizado el catálogo de GWAS de NHGRI (*National Human Genome Research Institute*) para identificar si existen fenotipos asociados a los genes considerados como seleccionados positivamente (Welter et al., 2013). Este catálogo contiene una colección de datos derivados de estudios de asociación publicados, que analicen al menos 100.000 *SNPs* y que presenten asociaciones entre genes y fenotipos con valores de $p < 10^{-5}$. En este caso, se consideraron los rasgos fenotípicos que presentaron una asociación con los genes seleccionados con un valor de $p < 10^{-8}$ según los datos publicados.

Evaluación de la metodología en una región control

Para comprobar la capacidad de detectar regiones bajo selección positiva reciente con iHS, se calculó este puntaje para individuos europeos en la región del gen LCT (chr2:136.545.410-136.594.750, CRCh37), zona reportada como seleccionada positivamente en poblaciones de ancestría europea (para más detalle, ver Selección positiva en humanos). Estos resultados se compararon con datos públicos de iHS descargados de los repositorios 1000 Genomes Selection Browser[2] (Pybus et al., 2013), Haplotter[3] (Voight et al., 2006) y HGDP Selection Browser[4] (Pickrell et al., 2009).

[2] http://hsb.upf.edu
[3] http://haplotter.uchicago.edu
[4] http://hgdp.uchicago.edu

Resultados

1. Estructura genética en las poblaciones estudiadas

Control de calidad

Luego del control de calidad y los filtros asociados a cada test, el número de *SNPs* utilizado en cada análisis varía entre 120.390 y 728.352. Los detalles de estos resultados son señalados en la Tabla 1. Además de los filtros por *SNP*, resultaron eliminadas dos muestras de Trapa-Trapa por presentar un alto índice de parentesco (IBD>0,3). Para la estimación de ancestría global hubo una gran disminución en el número de marcadores, debido al filtro de marcadores en desequilibrio de ligamiento. Esto, sin embargo, se encuentra dentro de lo esperado. El programa ADMIXTURE indica que 10.000 marcadores son suficientes para inferir ancestrías a nivel continental (Fst>0,05, (Alexander & Lange, 2011) adecuándose al objetivo del presente trabajo.

	Ancestría global	**iHS**	**XPEHH**	**Fst**
AMR	120.390	342.617 / 7.833		
AYM		348.555 / 7.959		
PH		324.638 / 7.427		
AMR.EAS			483.661 / 12.914	728.352 / 13.342
AMR.CEU			483.889 / 12.918	¨
AMR.YRI			484.350 / 12.931	¨
AYM.PH			707.700 / 12.953	¨

Tabla 1: Descripción de los datos utilizados. Por cada test, se señala: Número de marcadores / número de ventanas evaluadas (aquellas con al menos 20 marcadores).

Ancestría global

Utilizando el programa ADMIXTURE, se investigó la estructura poblacional en el grupo de individuos estudiados que presentara mayor soporte según su variación genómica. A partir de un estudio de validación cruzada, se encontró el número de poblaciones ancestrales (K) que mejor representa la variación observada en estas muestras. El mínimo error de predicción de genotipos se obtuvo con K=4 (Figura 5), con un error de 0,27561, muy similar al error con K=5 (0,27590). Estos resultados sugieren que ambos números de poblaciones ancestrales pueden ser soportados por los datos.

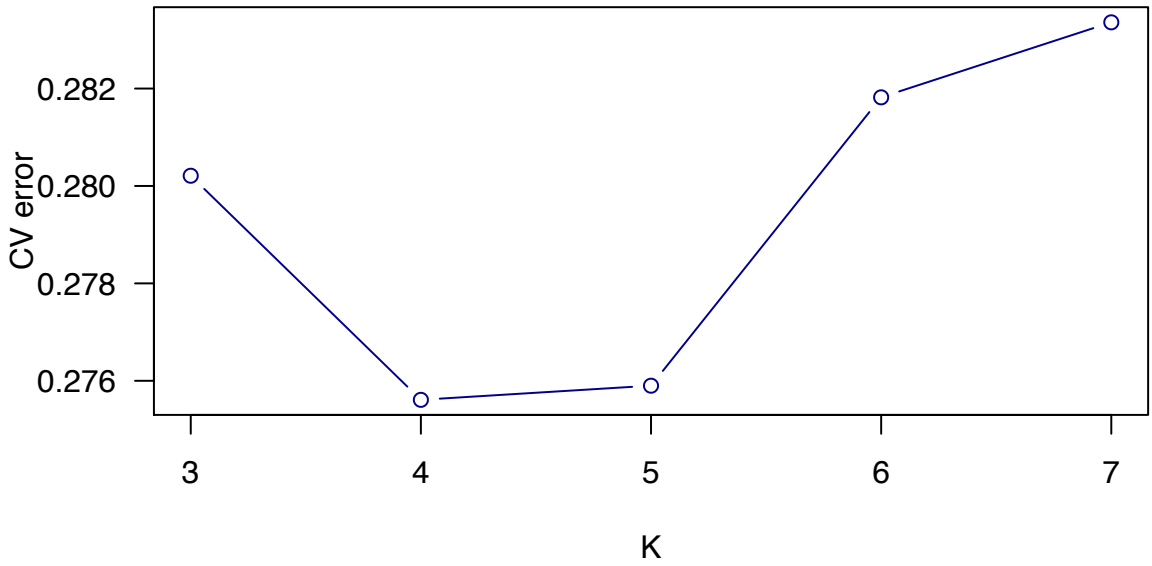

Figura 5: Error de validación cruzada de ADMIXTURE señala el número de poblaciones con menor error predictivo sobre los datos en el análisis de ancestría global.

Con un K=4 el análisis de ancestría global permite separar las poblaciones según su origen continental (Africa, Europa, Asia, América), mientras que con un K=5 el componente nativo americano se diferencia en dos grupos: uno compuesto por los Aymaras de Puno, y otro por los Pehuenches y Huilliches del sur chileno, tal como se observa en el gráfico de barras de la Figura 6. De estos resultados se desprende que existen diferencias genéticas entre las poblaciones del altiplano andino y del centro-sur de Chile, pero no es posible detectar estructura genética entre los Pehuenches y Huilliches a partir de los datos disponibles.

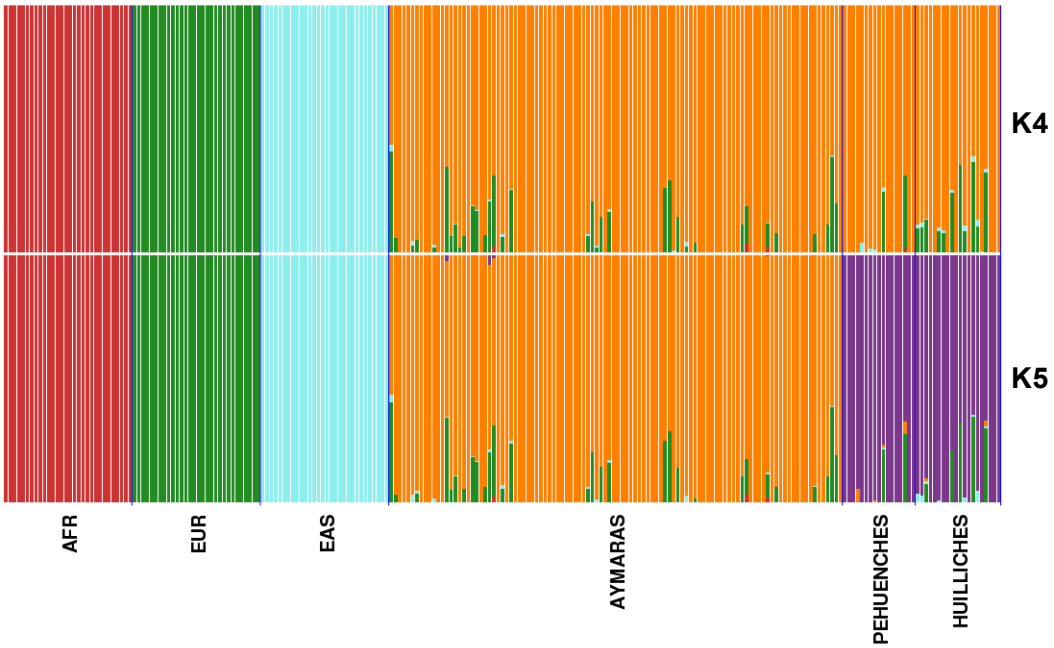

Figura 6: Inferencias de ancestría global de ADMIXTURE con paneles de 1000 Genomas representando poblaciones Africanas, Europeas, y del Este Asiático. Cada barra vertical representa a un individuo, y cada color esta asociado a un K o componente genético ancestral.

En este análisis se puede observar que si bien algunas de las muestras nativas americanas evaluadas presentan huellas de mestizaje con europeos, reflejando el proceso migratorio histórico de estas poblaciones al nuevo mundo, la mayoría de los individuos contiene una alta proporción de ancestría amerindia en sus genomas. Por otro lado, se observa que la porción genómica que puede ser asignada a un origen Asiático o Africano es baja o nula en todos los individuos americanos.

Las inferencias de proporciones de ancestría de cada uno de los grupos analizados coincide en general con los valores esperados, obteniéndose una clara separación entre los componentes genéticos ancestrales de cada continente. Gracias a la migración de individuos Europeos que conllevó la conquista del territorio nativo americano, encontramos individuos pertenecientes a etnias locales con distintos porcentajes de ancestría europea. Dentro de los Amerindios, las muestras de Huilliches son las que presentan un menor porcentaje de ancestría amerindia (Tabla 2), lo cual refleja un mayor porcentaje de mezcla con individuos de origen europeo. Por otro lado, los individuos de origen Pehuenche y Aymara tienen solo un 3-5% de ancestría europea, sin evidencias de mezcla con asiáticos ni africanos. En ninguno de los grupos el porcentaje de ancestría asiática ni africana supera el 0,1%, dando cuenta del bajo número de migrantes de esos continentes en estas poblaciones.

Set de datos	1000 Genomas			Puno, Perú	Trapa-Trapa, Alto Bío Bío	San Juan de la Costa, Osorno
Ascendencia	Europea	Africana	Asiática	Aymara	Pehuenche	Huilliche
N. individuos	30	30	30	106	15	20
Sexo (H / M)	12 / 18	13 / 17	10 / 20	41 / 65	8 / 7	10 / 10
% Ancestría amerindia	0	0	0	95	96,2	89,2
% Ancestría asiática	0	0	99,9	0,1	0,6	0,8
% Ancestría europea	99,9	0	0	4,7	3,2	9,9
% Ancestría africana	0	99,9	0	0,1	0,1	0

Tabla 2: Resultados del análisis de Ancestría Global con el programa ADMIXTURE.

Muestras con alta evidencia de ancestría genética amerindia

Para estudiar la divergencia existente en estas poblaciones de forma anterior a las migraciones recientes de individuos de otros continentes, se seleccionaron las muestras que presentaron más de 95% ancestría amerindia según el resultado de ADMIXTURE para un K=4, quedando en total 52 individuos, 30 de ellos Aymaras, 13 Pehuenches y 9 Huilliches. En la Tabla 3 se describen las muestras seleccionadas. Al analizar los resultados con K=5, se observa que el porcentaje del componente ancestral mayormente asociado al grupo Aymara es prácticamente nulo en los Huilliches, mientras que los Pehuenches presentan un valor un poco más alto pero que se encuentra dentro del margen de error de estas estimaciones.

	Aymara	Pehuenche	Huilliche	Total
N. individuos	30	13	9	52
Sexo (H / M)	16 / 14	6 / 7	5 / 4	27 / 25
% Ancestría amerindia promedio (K=4)	99	99	99	99
% Ancestría asociada a AYM promedio (K=5)	100	0,5	<0,1	57,8
% Ancestría asociada a PH promedio (K=5)	<0,1	99	100	42,1

Tabla 3: Muestras utilizadas para cálculo de Fst, iHS y XPEHH. Se conservaron en el set de datos sólo a los individuos con más del 95% de ancestría genética amerindia.

Diferenciación genética entre poblaciones

Tal como es esperado, se obtuvo que en promedio la mayor diferenciación genética se produce entre los nativos americanos (AMR) y los individuos de ancestría africana (Fst promedio: 0,234), seguida de la diferenciación respecto a los europeos (CEU, Fst: 0,152) y por último, respecto a los individuos del este asiático (EAS, Fst promedio: 0,118). Entre los indígenas del altiplano andino y del centro-sur de Chile, se obtuvo un valor de Fst de 0,046, y entre Pehuenches y Huilliches hay un Fst de 0,025.

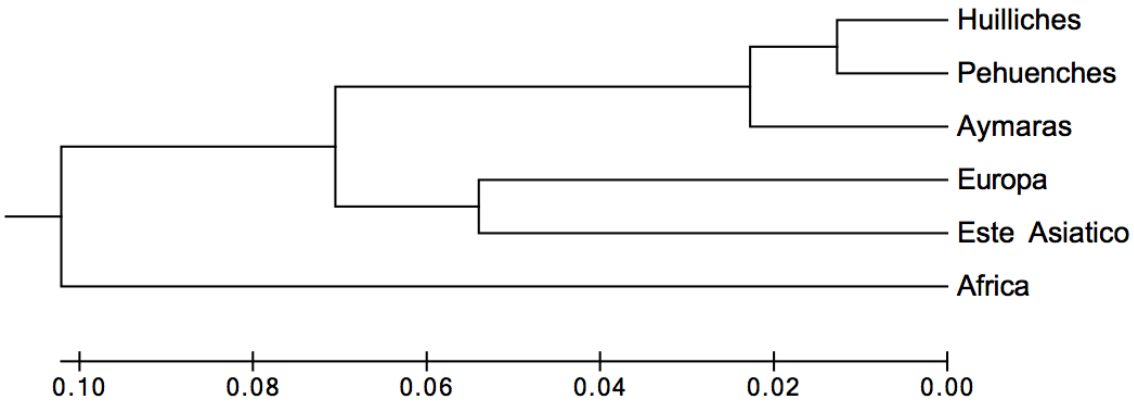

Figura 7: Dendrograma construido a partir de matriz de Fst entre las poblaciones estudiadas, usando el método UPGMA.

Si bien es posible que exista diferenciación genética entre las poblaciones Pehuenches y Huilliches, los resultados del análisis de ancestría global no permiten evaluarlas como

grupos separados. Por esto, en adelante, los test de selección fueron aplicados sólo considerando a todos los indígenas del sur de Chile como una población.

2. Test de selección positiva en Amerindios

Regiones con alta diferenciación genética

Al visualizar los valores de Fst por cada marcador a lo largo de todo el genoma, se encuentran regiones que presentan una alta diferenciación entre nativos americanos y las otras tres poblaciones con las que se comparó (Figura 8). Esto evidencia que no hay una distribución homogénea de los valores de Fst, y que hay segmentos altamente diferenciados solamente en amerindios que podrían ser explicados por procesos asociados a selección en esta población y no por deriva génica. La distribución de valores de Fst en el genoma completo se muestra en el Anexo (Figura S1).

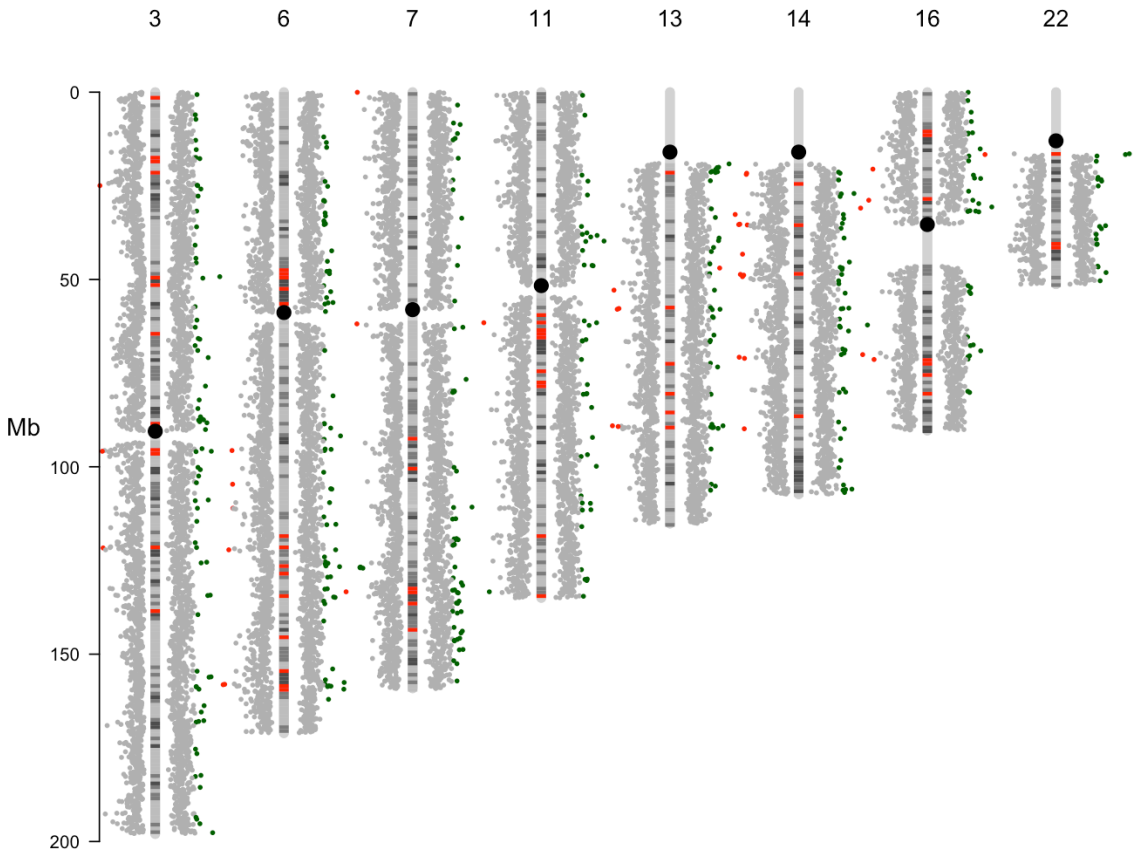

Figura 8: Valores de Fst a nivel genómico entre Americanos y Asiáticos (a la izquierda del cromosoma), Europeos (cuerpo cromosómico), y Africanos (lado derecho del cromosoma). En rojo y verde aparecen destacados los valores de Fst por región mayores a 0,25.

Si definimos las regiones con alta divergencia como las ventanas de 200 kb con puntajes máximos mayores al 1% más alto, se confirma lo señalado anteriormente respecto a la diferenciación entre los americanos y las poblaciones provenientes de otros continentes

(Anexo, Figura S2), pues los valores del 1% más alto son más altos al calcular Fst entre amerindios y africanos, y los más bajos se obtienen al comparar con asiáticos, aunque siguen siendo puntajes que señalan una alta diferenciación (Fst=0,76). De esta forma, todas las ventanas consideradas de interés (alrededor de 140 por población) presentan niveles de diferenciación muy alto.

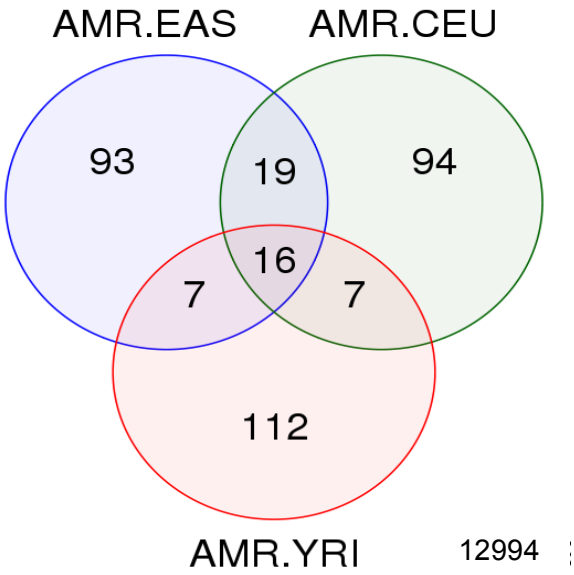

Figura 9: Diagrama de Venn con las ventanas de mayor puntaje de Fst (1% más alto), al comparar la población amerindia (AMR) con asiáticos (EAS), europeos (CEU) y africanos (YRI). El número al margen señala el número total de ventanas evaluadas con este test.

Al comparar las regiones de interés obtenidas de los tres cálculos (Figura 9), se identifican 16 ventanas comunes que comprenden a 66 genes (detalle en Anexo, Tabla S3). El listado de genes candidatos según Fst entre Asiáticos y Amerindios indica al transporte de lípidos como significativamente sobre-representado (p=1,29 x 10^{-7}), junto al proceso de transporte de carbohidratos.

iHS

Según la literatura, puntajes de iHS mayores a |2,5| o |2,0| señalan regiones candidatas a estar seleccionadas positivamente, lo que suele coincidir con el 1% o 5% (respectivamente) de los marcadores con puntaje más alto al estudiar genomas completos. De manera similar, en este estudio el 1% de los puntajes más altos equivale a valores de iHS sobre 2,86, y el 5% a valores sobre 1,98 (Distribución de los puntajes en Anexo, Figura S3). Tal como se ha visto en otros trabajos, se encuentran marcadores iHS > |2,5| o en el 1% más alto en todos los cromosomas (Figura 10).

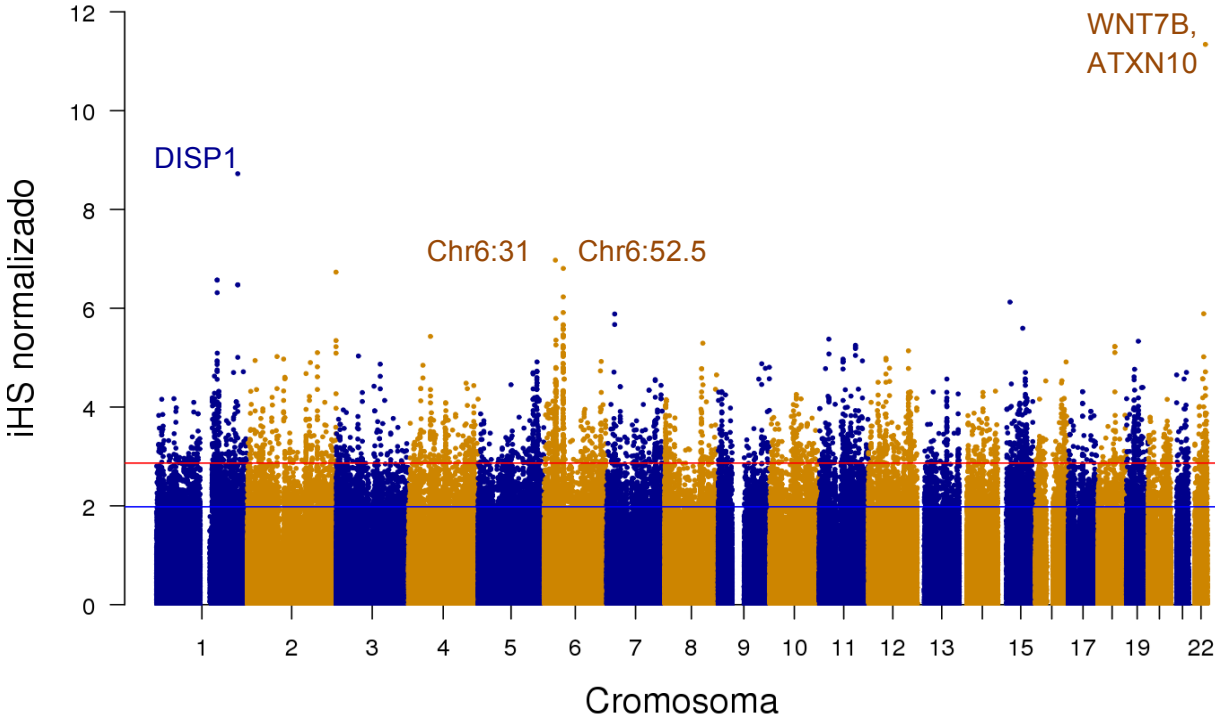

Figura 10: Manhattan plot de iHS por marcador en la población de Amerindios (AMR), y algunos genes cercanos a los marcadores con puntaje más alto a nivel genómico. Las líneas roja y azul señalan el 1% y 5% de los puntaes más altos, respectivamente.

El valor máximo de iHS obtenido es de 11,34, en el marcador rs9330783 del cromosoma 22, cercano a los genes WNT7B (desarrollo embrionario) y ATXN10 (neurogénesis). Estos genes han sido señalados previamente como candidatos en poblaciones asiáticas y africanas mediante el test XP-CLR de selección positiva (Chen, Patterson, & Reich, 2010). El segundo valor más alto (iHS=8,72) se encuentra en la región chr1:223,0-223,1 Mb, donde está el gen DISP1, involucrado en el desarrollo embrionario. Aquí se concentran varios de los marcadores con puntajes más altos (rs6658697, rs124098339) y el gen ha sido señalado como candidato en poblaciones Europeas y del Este Asiático por presentar altos valores en el test XP-EHH (Higasa et al., 2009). El tercer y el cuarto marcador, cercanos a los genes MUC22 (chr6:31 Mb) y TRAM2 (chr6:52,5 Mb), se encuentran en dos regiones del cromosoma 6 que presentan una gran concentración de puntajes altos (Figura 12).

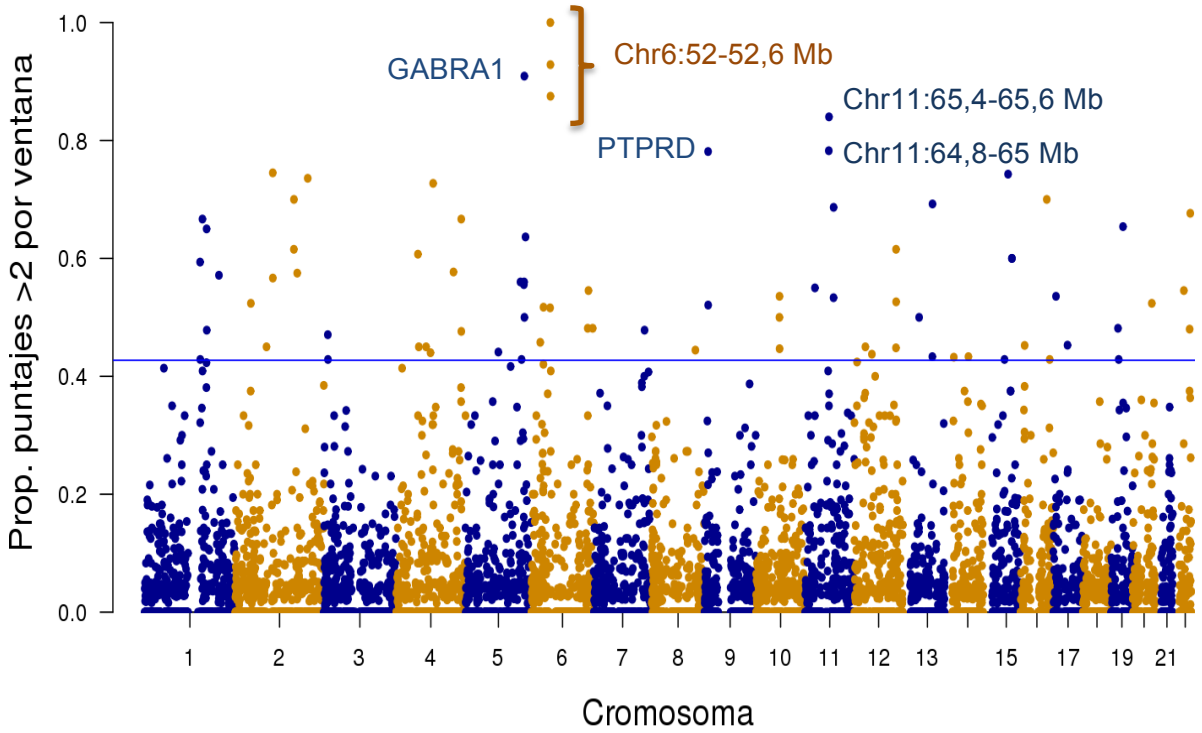

Figura 11: Proporción de marcadores con valores de |iHS| >2 en ventanas de 200kb. Se señalan las 5 regiones con puntajes más altos en todo el genoma. La linea azul indica el puntaje de corte del 1% más alto (0,427).

Al evaluar las ventanas de 200 kb a partir de las proporción de valores altos (|iHS|>2) en ellas, encontramos que el 50% de ellas no tiene marcadores con puntajes de iHS sobre 2, y que el 1% de las ventanas tiene más del 42% de sus *SNPs* con puntajes altos (Distribución de los puntajes por ventana en Anexo, Figura S4). De las 7833 ventanas evaluadas, que incluyen 15.977 genes anotados, 79 fueron consideradas como regiones de interés, y contienen 257 genes candidatos. El listado completo de las ventanas dentro del 1% más alto, junto con los genes en esas regiones y sus fenotipos asociados, se encuentra en el Anexo, Tabla S1.

Al visualizar la proporción de puntajes altos de iHS por ventana se pueden identificar algunas regiones de interés: la principal, se localiza en el cromosoma 6 (Figura 11). Al analizar con mayor detalle este cromosoma y comparar estos resultados con valores de iHS en otras poblaciones, se observan dos grandes agrupaciones de valores altos ubicadas entre los 31,0-33,1 Mb y. 51,9-52,3 Mb (Figura 12). Esta última región contiene a la única ventana de 200 kb del genoma con el 100% de sus marcadores con valores de |iHS| sobre 2 (25 marcadores). Además, presenta puntajes altos sólo en el grupo de Nativos americanos y no en las otras poblaciones evaluadas. Esta región contiene a 10 genes, entre los cuales se encuentran MCM3 e IL17F, que han sido señalados como seleccionados positivamente en individuos Americanos de la base de datos HGDP (López Herráez et al., 2009) y en un estudio realizado en Collas (Eichstaedt et al., 2014) TRAM2, por otro lado, aparece como región candidata a ser seleccionada en individuos europeos (Lappalainen et al., 2010).

La región 31,0-33,1 Mb presenta puntajes altos de iHS en todas las poblaciones. En ella se encuentran 97 genes, que en su mayoría están asociados al HLA (sistema de antígenos leucocitarios humanos).

A nivel genómico, los genes en el 5% de las ventanas con mayor proporción de puntajes altos de iHS en amerindios se encuentra sobre-representada en dos procesos biológicos según PANTHER: metabolismo de carbohidratos (valor de p = 2,27 x 10^{-5}) y la función "activación de células asesinas naturales" (valor de p = 2,69 x 10^{-5}), de importancia en la respuesta inmune.

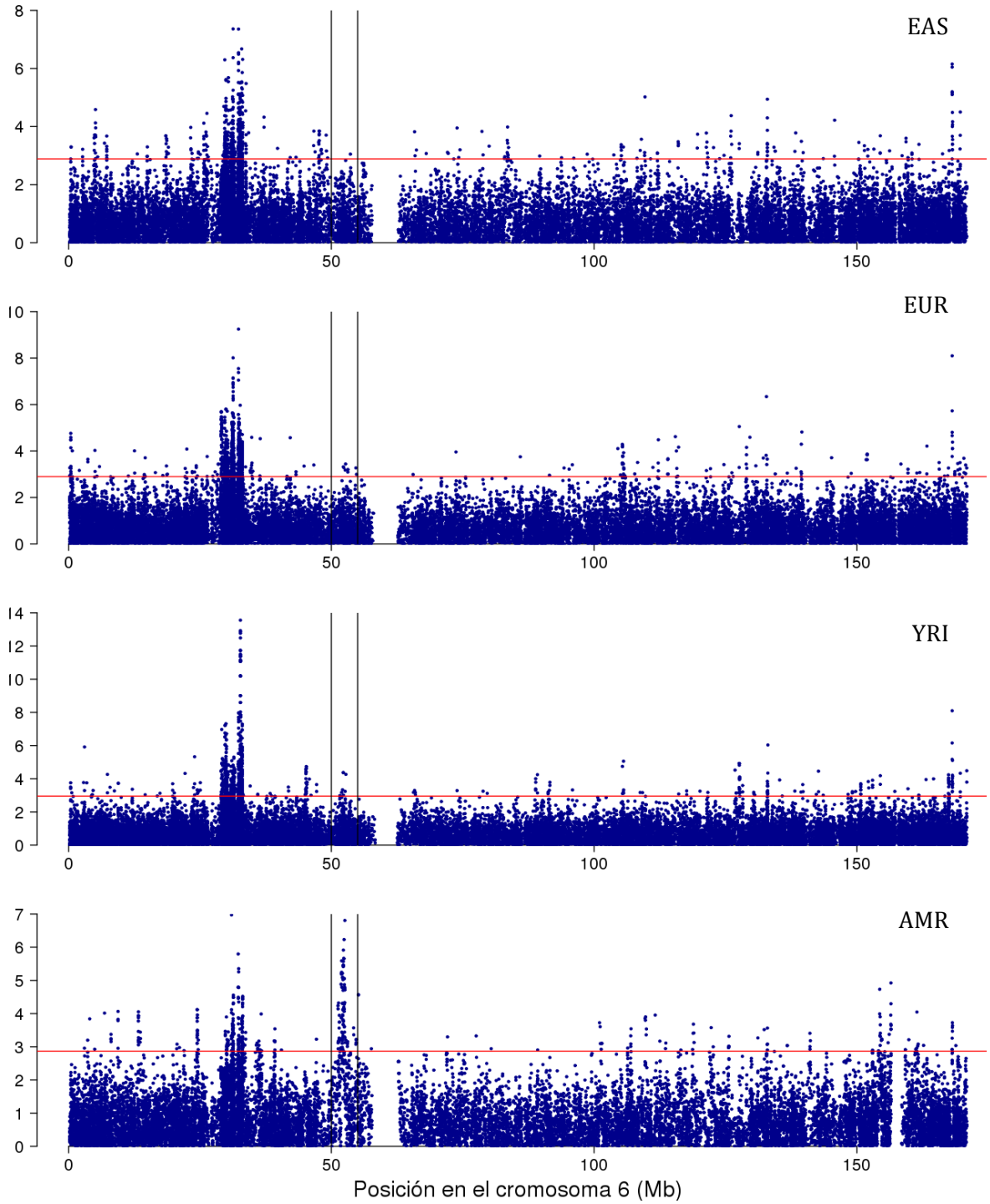

Figura 12: Manhattan plot de iHS en cromosoma 6 en poblaciones del Este Asiático, Europa y África, por marcador. La línea horizontal roja en cada gráfico señala el valor de corte del 1% de los puntajes más altos dentro de cada población. Las líneas verticales negras indican la región entre los 50 y los 55 Mb.

XPEHH

Los puntajes altos en este test representan la existencia de haplotipos notablemente más extensos en una población que en otra. En el histograma de los valores por marcador (S, Figura 5) se observa que la distribución de los puntajes es muy semejante al comparar a los Americanos con Asiáticos, Europeos o Africanos. Al comparar el 1% de las ventanas con mayores puntajes de las tres comparaciones, encontramos coincidencia en 20 de 12.596 ventanas (Figura 13). Estas son de especial interés, pues representan la existencia de haplotipos de extensión notoriamente distinta para las poblaciones nativas americanas respecto a todos los otros grupos continentales estudiados (el detalle de las regiones candidatas se encuentra en el Anexo, Tabla S2)

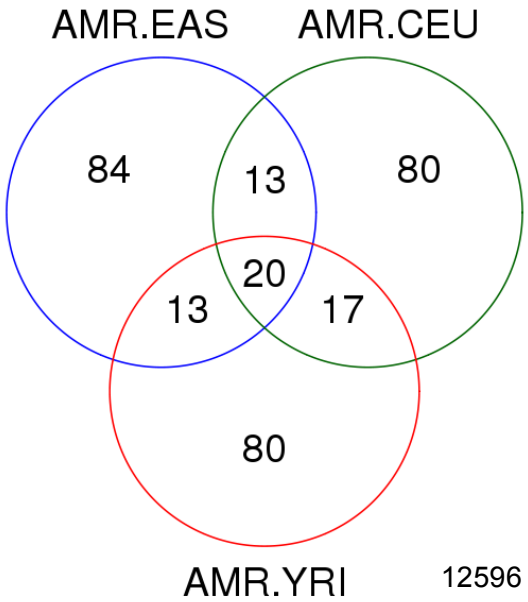

Figura 13: Diagrama de Venn con las ventanas de 200 kb con puntaje de XPEHH más alto (1%), al comparar la población amerindia (AMR) con asiáticos (EAS), europeos (CEU) y africanos (YRI). El número al margen señala el número total de ventanas evaluadas.

Dentro de estas 20 ventanas comunes para los tres cálculos de XPEHH, 3 corresponden a la región con mayor proporción de puntajes altos de iHS ya señalada (chr6:51,8-52,6 Mb), lo cual aumenta la evidencia para considerarla una región de relevancia. Además, esta región presenta los puntajes más altos del test por marcador, al calcular XPEHH entre amerindios y asiáticos, como se observa en la Figura 14.

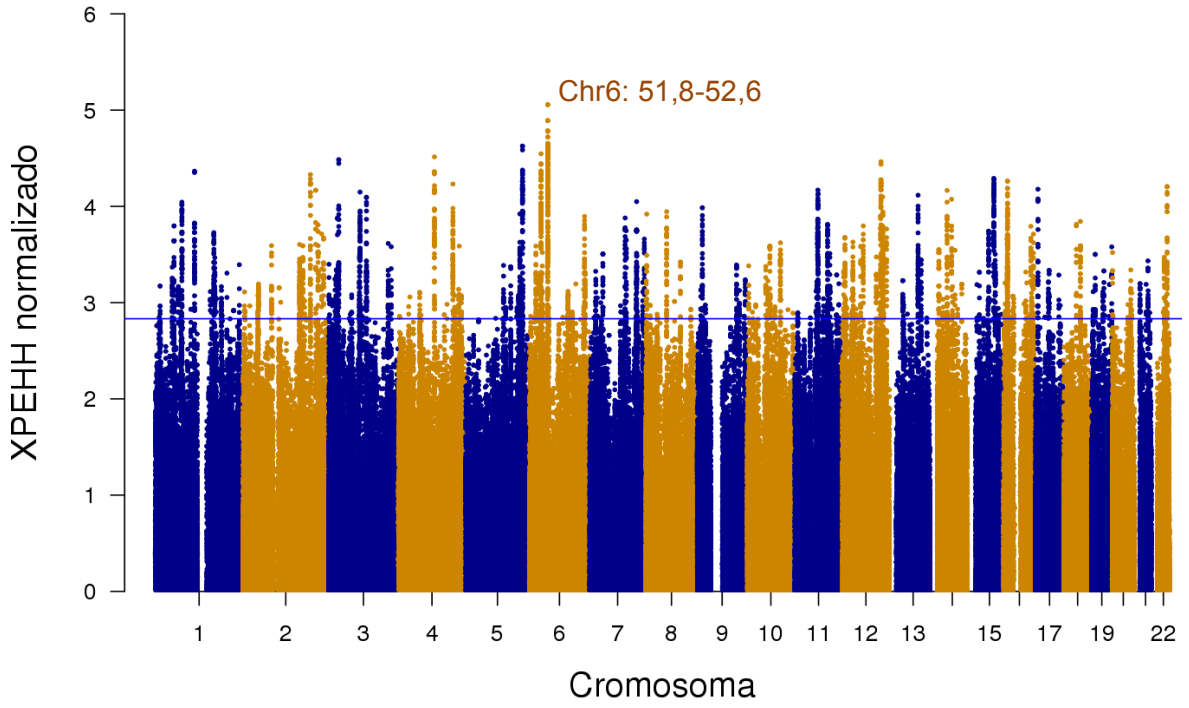

Figura 14: Manhattan plot de XPEHH por *SNP* entre nativos amerindios e individuos del este asiático (EAS). La línea horizontal azul señala el valor de corte del 1% más alto.

También resulta interesante una región que se extiende por 3 ventanas (600 kb) de las 20 comunes en las tres comparaciones. Se ubica en el cromosoma 16 (11-11,6 Mb), ha sido señalada como región seleccionada en mexicanos (X. Liu et al., 2013), y presenta asociación con varias patologías, como la esclerosis múltiple, la diabetes tipo 1, enfermedad de Crohn, cirrosis y enfermedad celiaca (gen CLC16A). Tres ventanas consecutivas del cromosoma 7 también se encuentran entre las 20 ventanas comunes, todas asociadas al gen EXOC4, relacionado a fenotipos como peso corporal, presión arterial y niveles de triglicéridos.

Al evaluar de forma global los genes candidatos según XPEHH entre Amerindios y Asiáticos aparecen varios procesos biológicos sobre-representados. Algunos de los que presentan valores significativos son el transporte de lípidos, con un p = 6,11 x 10^{-7}, que se suma al metabolismo de lípidos, transporte de cationes, señalización célula-célula, transmisión sináptica y transporte de iones. También está sobre-representado el proceso de transporte (movimiento de sustancias entre, dentro y fuera de células) y localización (proceso en el que una célula o sustancia se transporta hacia una ubicación o se mantiene en ella).

3. Test de selección positiva entre Aymaras y Pehuenches-Huilliches

Fueron evaluados los resultados de iHS por marcador para los Aymaras (AYM), separados de los Pehuenches y Huilliches (PH). La razón de este agrupamiento de individuos radica en la ausencia de estructura genética evidente entre muestras Pehuenche y Huillliche, lo cual sugiere que no ha transcurrido suficiente tiempo de asilamiento entre estas poblaciones para evidenciar procesos microevolutivos distintos. Los resultados de iHS por separado en AYM y PH mostraron nuevas zonas candidatas de haber sufrido selección positiva (Figura 15). En Aymaras, el puntaje más alto está en el cromosoma 11 (iHS=6,93), en el gen LUZP2, que también presenta altos puntajes de iHS en Japoneses y Europeos (Higasa et al., 2009).

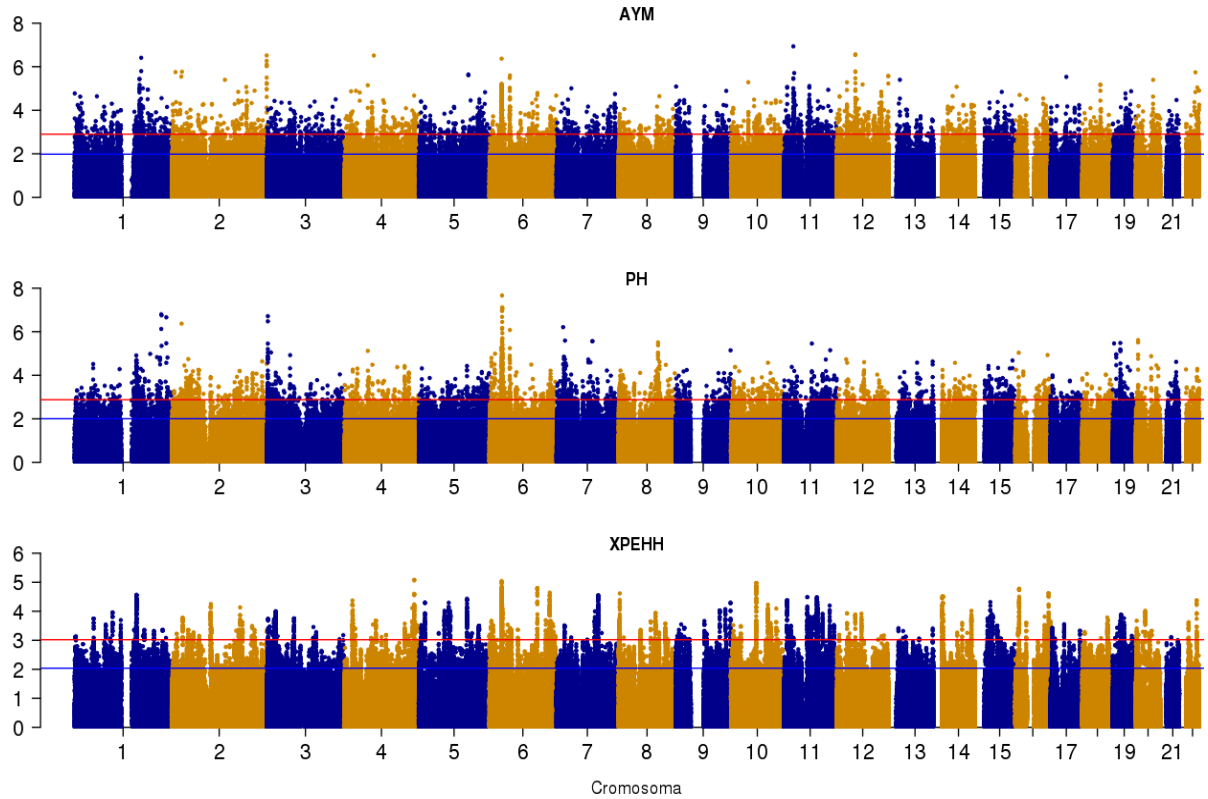

Figura 15: Manhattan plot de iHS en Aymaras (arriba) y en los Pehuenches-Huilliches (al centro), por marcador, junto a los valores de XPEHH que resultan de la comparación entre estos dos grupos. Las líneas rojas señalan el valor del 1% más alto y las azules el 5% más alto en cada set de datos.

En las regiones candidatas según iHS en Aymaras, aparecen sobre-representados los procesos de transporte de iones, y sub-representado el transporte celular mediado por vesículas y la percepción sensorial.

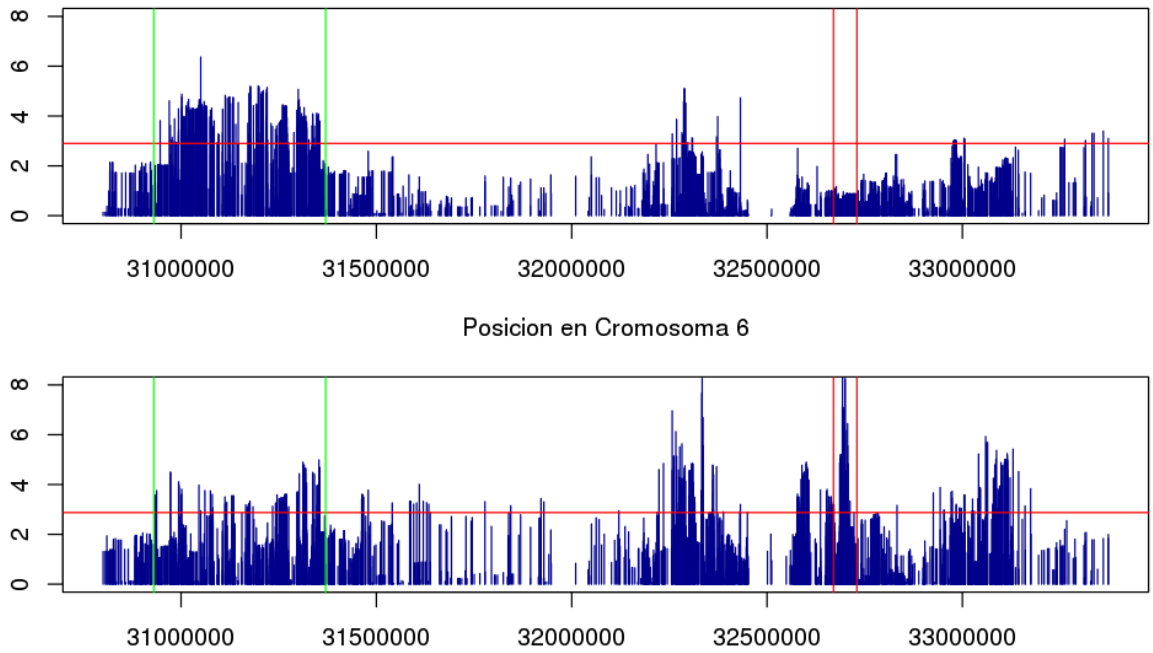

Figura 16: Puntajes de iHS en la región chr6: 30.800.000-33.400.000, en Aymaras (arriba) y Pehuenches y Huilliches (abajo). Entre líneas verdes, se indica la región 30,93-31,37 Mb. Entre líneas rojas verticales, se señala la región 32,67-32,72 Mb. La línea horizontal roja señala el valor del 1% más alto a nivel genómico en cada set de datos.

En los Pehuenche-Huilliches los puntajes más altos se ubican en la región 31,8-33,2 Mb del cromosoma 6, que también fue indicada como una zona de relevancia en los test presentados anteriormente. Analizando la región en detalle, descubrimos que en la población de Pehuenche-Huilliches, existe una acumulación de puntajes altos en una región de 50 kb entre las posiciones 32,67 y 32,72 Mb que no aparece como seleccionada en individuos Aymaras (Figura 16). En ella se encuentran genes asociados al complejo HLA (antígenos leucocitarios humanos), involucrado en el sistema inmune, como se señaló previamente. De forma inversa, la población Aymara presenta mayores puntajes entre las posiciones 30,93 y 31,37 Mb, aunque también se observa cierta señal en esa zona en los Pehuenche-Huilliches (Figura 15). En los resultados de iHS en Pehuenches y Huilliches, está sobre-representados procesos referidos a la adhesión entre células (p=6,65 x 10^{-7}), o entre estas y un sustrato (p=2,26 x 10^{-5}). En el caso de la población Aymara, está sobre-representado el proceso de transporte de iones (p=1,31 x 10^{-5}) y el transporte mediado por vesículas (p=1,7 x 10^{-5}).

Al comparar los resultados de los test XPEHH y Fst entre las poblaciones AYM y PH, se identifican 10 ventanas en común (Figura 17; descritas en detalle en el Anexo, Tabla S6). Entre estas, encontramos nuevamente la región del cromosoma 6 (31-31,6 Mb) asociada al Complejo de histocompatibilidad (o complejo HLA), como altamente diferenciada entre las dos poblaciones.

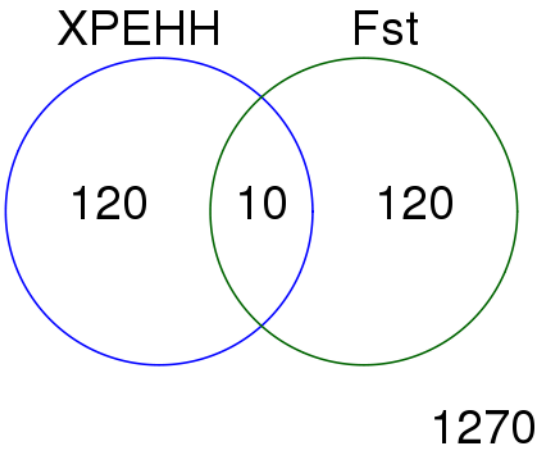

Figura 17: Diagrama de Venn que señala coincidencia entre regiones candidatas (1% más alto) según test Fst y XPEHH entre población Aymara y Pehuenches y Huilliches.

Este acotado listado de 10 ventanas comunes, comprende también a la región chr2:99,2-99,6 Mb, con el gen MGAT4A asociado a neoplasmas (Murabito et al., 2007), y a la región chr15:45-45,4 Mb, con el gen B2M relacionado al complejo de histocompatibilidad mayor (D'Urso et al., 1991)

Del análisis global de las regiones con altos valores de XPEHH entre Aymaras y Pehuenches-Huilliches, observamos sobre-representación de genes relacionados al procesamiento y presentación de antígenos (p=1,24 x 10^{-7}), de potencial relevancia para las reacciones inmunológicas. Con Fst, obtenemos los procesos de muerte celular (p=9,95 x 10^{-9}), y transporte de lípidos como sobre-representados.

Discusión

A través de los análisis de ancestría global (con ADMIXTURE), detectamos que los individuos de ancestría amerindia presentaron un bajo o nulo aporte genético de poblaciones provenientes de otros continentes. Esto complementa la información existente respecto al alto componente nativo de los individuos Aymaras puneños en Perú (Barbieri, 2011), y al aislamiento y bajo mestizaje de las poblaciones Pehuenches y Huilliches hasta tiempos históricos recientes (Llop et al., 1993).

En los resultados del análisis de ancestría global, sumado al de diferenciación genética (Fst), encontramos diferenciación entre las poblaciones del altiplano andino y las del centro-sur de Chile (Fst=0,046) comparable a la encontrada entre indígenas mexicanos Huichol (norte) y Maya (sudeste), que tienen un Fst de 0,045 (Moreno-Estrada et al., 2014). Dentro de las poblaciones europeas, se observan diferencias menores a las encontradas entre Aymaras y Pehuenches-Huilliches. De hecho, entre Orcadianos (nativos del norte del Reino Unido) y Beduinos (Arabia Saudita) hay un Fst de 0,021 (Tian et al., 2009), menor al encontrado entre Pehuenches y Huilliches (0,025). Esta diferenciación entre los nativos del centro-sur es equivalente a la encontrada entre indígenas mexicanos Nahua (del Oeste central de México) y Maya (del sudeste de México), que tienen un Fst de 0,026 (Moreno-Estrada et al., 2014). Sin embargo, con los datos genómicos utilizados y el análisis de ADMIXTURE, no fue posible identificar estructura genética entre Pehuenches y Huilliches. Estos resultados sugieren que el impacto de la deriva génica en estos dos grupos ha sido menor, posiblemente por el breve de tiempo de divergencia entre ambas o por flujo génico producto de migraciones entre ellas, lo cual es esperable dada la proximidad geográfica y las actividades de intercambio que sostuvieron (Gissi, 1997). Si bien estos resultados nos llevaron a agrupar a Pehuenches y Huilliches en la búsqueda de señales de selección positiva, no es posible extrapolar esta agrupación a dimensiones socio-culturales, pues estas últimas pueden variar de forma distinta a los factores genéticos. Adicionalmente, las estimaciones de estructura genética pueden estar sesgadas por la selección de *SNPs* en el microarreglo empleado para la genotipificación, cuyo diseño no consideró datos de ninguna de estas dos poblaciones y por lo tanto pueden tender a agruparlas. La generación de datos de secuenciación en Pehuenches y Huilliches permitiría realizar evaluaciones más exactas del nivel de estructura genética presente entre ellas.

En este trabajo proponemos 44 regiones (52 ventanas de 200 kb. con 181 genes en total) con señales de selección según el test iHS que se presentan sólo en las poblaciones amerindias y no en asiáticos, europeos ni africanos. Utilizando múltiples estimadores de selección positiva, la región del chr6 comprendida entre las posiciones 52-52,6 Mb presentó una fuerte señal de selección en las muestras amerindias. Esta región muestra un alto puntaje en los test de iHS en Aymaras y en Pehuenches-Huilliches. También tiene altos puntajes al calcular XPEHH contra las poblaciones de otros tres continentes (asiáticos, europeos y africanos). Adicionalmente, no presenta señales de selección en las poblaciones de otros continentes. En esta región se encuentran genes como IL17F, asociados al funcionamiento renal; IL17A, a los niveles de colesterol, hematocrito y

hemoglobina (Yang, Kathiresan, Lin, Tofler, & O'Donnell, 2007), TRAM2 asociado a la presión sanguínea (Levy et al., 2007) y EFHC1, a la función respiratoria (Wilk, Walter, Laramie, Gottlieb, & O'Connor, 2007).

Otra de las regiones con fuerte señal de selección halladas en este trabajo, se encuentra también en el cromosoma 6, en 30,8-31,2 Mb, y aparece seleccionada según el test de iHS de Aymaras, y en los test XPEHH y Fst entre Aymaras y Pehuenches-Huilliches. En esta región se encuentran genes como MUC22 y PSORS1C2, cuyas variantes nativas americanas han sido asociadas a un menor riesgo de asma (Galanter et al., 2014). Cercana a esta región, entre 32,6-33,2, existe una alta señal se selección en Pehuenches-Huilliches según iHS. Aquí se ubica una gran cantidad de genes asociados al sistema HLA o en animales no-humanos, complejo de histocompatibilidad mayor (MHC). Los genes de este complejo son responsables de determinar histocompatibilidad entre individuos y son esenciales en la activación y regulación de la respuesta inmune ante agentes patógenos. Esto es muy relevante en la capacidad reproductiva de los individuos de una población y podría representar una adaptación a la carga patógena local. Identificar señales de selección en genes relacionados con respuestas inmunes es común en este tipo de trabajos (Pardis C Sabeti et al., 2007; Tang, Thornton, & Stoneking, 2007; Voight et al., 2006) e incluso se ha propuesto que, en comparación con el clima, la dieta y otros factores ecológicos, el principal motor de la selección positiva es la adaptación local al ambiente patogénico (Fumagalli et al., 2011).

El número de estudios que proponen genes candidatos a ser seleccionados en poblaciones amerindias es bajo, pero entre los existentes, se encuentra el gen SIK3. Este se encuentra en la región del cromosoma 11, 116,6-118,8 Mb, que en este trabajo contiene 33 marcadores con valores de iHS altos en los Aymaras. En la literatura se ha propuesto que el componente nativo americano otorga una mayor susceptibilidad genética a diabetes tipo 2, obesidad y dislipidemia (Aguilar-Salinas, 2009), y el gen SIK3 (asociado con la mantención de altos niveles de triglicéridos en el plasma) podría ser uno de los responsables (Ko et al., 2014). La hipótesis que se ha planteado al respecto, es que en momentos tempranos de la ocupación americana, la mantención de altos niveles de lípidos en la sangre podría ser una característica favorable en tiempos de escasez.

La existencia de señales de selección positiva diferentes entre las poblaciones del altiplano andino y del centro-sur de Chile se puede producir en alguno de los siguientes escenarios: (1) selección positiva fuerte ocurrida luego de la separación de los grupos, (2) suficiente tiempo de separación entre poblaciones para que el "relajo" de las presiones selectivas en alguna de ellas (al emigrar, por ejemplo) permitiera visualizar diferencias en las regiones genómicas involucradas en la adaptación (Hancock & Di Rienzo, 2008). En el presente estudio, el metabolismo de lípidos aparece como sobre-representado en el listado de genes dentro del 5% de las ventanas con mayor puntaje en los test XPEHH y Fst entre amerindios y asiáticos; lo cual señala que las regiones con diferencias de extensión de haplotipos entre las dos poblaciones están asociadas significativamente a este proceso, de alta relevancia para la salud pública local.

Es necesario mencionar que la aproximación aquí utilizada para buscar señales de selección positiva tiene limitaciones importantes de considerar. Los tests estadísticos basados en extensión de homocigosidad de haplotipos dependen de que se tenga el suficiente número de marcadores en la región de interés. En nuestra experiencia, los filtros que se utilizan para calcular iHS y XP-EHH basado en frecuencias alélicas mínimas pueden eliminar muchos *SNPs* en una región bajo selección positiva, dejando un número insuficiente de marcadores para realizar el cálculo de estos estadísticos. Esto está condicionado el diseño de la plataforma de genotipificación usada. Por ejemplo, el locus LCT no presentó un número suficiente de marcadores para detectar selección en individuos CEU cuando se utilizó el mismo set de marcadores disponibles para individuos de ancestría Amerindia estudiados en este trabajo. Tampoco hay marcadores en esta región en los resultados de 1000G ni HGDP. Sin embargo, al repetir el test con todos los marcadores en HapMap (fase 3) para CEU, sí pudimos detectar la señal de selección reportada en trabajos anteriores (detalles en Anexo, Apéndice 1).

El uso de microarreglos en este tipo de estudios tiene al menos dos limitantes: en primer lugar, suelen tener una baja densidad de datos en regiones con alto desequilibrio de ligamiento, y en segundo lugar, tienden a utilizar alelos comunes en la población continental haciendo que -probablemente- los sitios variantes asociados a adaptaciones locales estén sub-representados. De hecho, el diseño del microarreglo usado en este trabajo no consideró frecuencias alélicas de ninguna población indígena del sur de chile (Hoffmann et al., 2011). Por lo tanto, es posible que el diseño de la plataforma usada produzca un sesgo en las frecuencias alélicas promedio del genoma lo cual podría reducir artificialmente los valores estimados de divergencia genética en entre Pehuenches y Huilliches. Los resultados aquí presentados para estas poblaciones deben ser confirmados mediante el uso de microarreglos alternativos o por datos de secuenciación masiva que evitan este sesgo.

El poder de identificación de señales de selección positiva a través de los test utilizados varía según múltiples factores, como el escenario demográfico de la población estudiada, las características del barrido selectivo, la densidad de datos presentes en esa región, el tamaño muestral, entre otros (Wilson 2014). Por tanto, la ausencia de señales de selección en un loci no debe ser interpretada como una ausencia de selección, tal como ocurre con el caso del gen LCT en varios estudios (Lappalainen, 2010).

A pesar de lo anterior, las metodologías usadas permiten identificar haplotipos seleccionados de manera reciente, que se encuentren en las poblaciones en frecuencias intermedias a altas en el caso del test iHS, y de alta frecuencia o fijación en el test XPEHH. El test Fst es sensible a eventos selectivos de mayor antigüedad y tiene poca capacidad para identificar selección por si solo, pero se complementa de buena forma si se considera en conjunto con el test XPEHH, según lo demuestran simulaciones realizadas en otros trabajos (Barreiro et al. 2008; Lappalainen et al. 2010).

La interpretación de los resultados obtenidos a través de aproximaciones genómicas se dificulta por la gran cantidad de genes que se encuentran en las regiones consideradas

como seleccionadas. Sumado a esto, muchos de estos genes no tienen función conocida o, si la tienen, estas pueden no tener una relación evidente con la adaptabilidad biológica de las poblaciones, complejizando la postulación de hipótesis evolutivas. Sin embargo, a partir de la identificación de regiones genómicas con señales de selección en dos grupos de ancestría amerindia, pudimos encontrar procesos biológicos sobre-representados en los genes de estas regiones. Es posible que la selección actúe a nivel de redes genéticas complejas, desarrollando adaptaciones poligénicas asociadas con ciertos procesos biológicos. En este escenario, la búsqueda de genes específicos bajo selección por mapeo posicional puede no entregar resultados satisfactorios, aun con grandes tamaños poblacionales o disponibilidad de secuencias completas. Es posible, entonces, que la utilización de pruebas estadísticas de selección positiva que evalúen redes génicas asociadas a algún proceso biológico pueda tener una mayor sensibilidad que las pruebas actualmente disponibles.

Conclusiones

Los resultados obtenidos nos permiten concluir que sí existen regiones genómicas con alta evidencia de selección positiva en las poblaciones nativas americanas, y que algunas de estas son distintas entre las poblaciones evaluadas del altiplano andino y el centro-sur de Chile.

Estas señales de selección están asociadas al metabolismo de carbohidratos, de lípidos, y a las respuestas inmunes: tres procesos biológicos que representan adaptaciones de alta relevancia biomédica y bioantropológica. El estudio en detalle de estos procesos biológicos ayudará a la comprensión de las patologías asociadas y su origen, a la identificación de variantes de relevancia funcional en las poblaciones americanas, y la generación de conocimiento biomédico sensible a las particularidades genéticas de las poblaciones locales.

Los resultados obtenidos nos revelan parte de la historia microevolutiva de las poblaciones analizadas. Si bien las regiones propuestas como candidatas pueden haber sido seleccionadas antes de que se iniciaran las primeras migraciones al continente americano o de forma posterior a este evento, el uso de muestras de alta ancestría genética amerindia y la aproximación metodológica utilizada nos permitió identificar regiones que fueron seleccionadas sólo en los genomas amerindios, revelando características importantes sobre la diversidad biológica existente en las poblaciones originarias.

Bibliografía

Abarzúa, A. M., Pinchicura, A. G., Jarpa, L., Sterken, M., Vega, R., & Q, M. P. (2014). Environmental Responses to Climatic and Cultural Changes. In *The Teleoscopic polity. Andean patriarchy and materiality.* (pp. 123–141). New York: Springer.

Adán, L., & Mera, R. (2011). Variabilidad interna en el alfarero temprano del centro-sur de Chile: el complejo Pitrén en el valle central del Cautín y el sector lacustre andino. *Chungará*, *43*(1), 3–23.

Alcaman, E. (1994). La sociedad mapuche- huilliche del futahuillimapu septentrional. 1750-1792. *Boletín Del Museo Histórico Municipal de Osorno*, *Nº1*, 64–90.

Aldunate del S., C. (1993). Estadio alfarero en el Sur de Chile (500 a ca. 1800 d. C.). In *Culturas de Chile. Prehistoria, desde sus origenes hasta los albores de la conquista* (2da ed., pp. 329–348). Santiago: Andrés Bello.

Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, *12*(1), 246. http://doi.org/10.1186/1471-2105-12-246

Apata, M. A., & Moraga, M. (2015). *Señales de adaptación humana al hidroarsenicismo en los valles andinos de la Región de Arica y Parinacota: evidencias de variantes protectoras del gen arsénico [+3] metiltransferasa*. Universidad de Chile.

Auton, A., Bryc, K., Boyko, A. R., Lohmueller, K. E., Novembre, J., Reynolds, A., … Bustamante, C. D. (2009). Global distribution of genomic diversity underscores rich complex history of continental human populations. *Genome Research*, 795–803. http://doi.org/10.1101/gr.088898.108.51

Barreiro, L. B., Laval, G., Quach, H., Patin, E., & Quintana-Murci, L. (2008). Natural selection has driven population differentiation in modern humans. *Nature Genetics*, *40*(3), 340–345. http://doi.org/10.1038/ng.78

Bengoa, J. (1996). *Historia del pueblo Mapuche* (5a. Edicio). Santiago, Chile: Ediciones Sur.

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., Mao, X., Mei, R., Brutsaert, T., Wilson, M. J., Julian, C. G., … Shriver, M. D. (2009). Identifying positive selection candidate loci for high-altitude adaptation in

Andean populations. *Human Genomics*, *4*(2), 79–90. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2857381&tool=pmcentrez&rendertype=abstract

Bigham, A. W., Wilson, M. J., Julian, C. G., Kiyamu, M., Vargas, E., Leon-Velarde, F., … Shriver, M. D. (2013). Andean and Tibetan patterns of adaptation to high altitude. *American Journal of Human Biology*, *25*(2), 190–197. http://doi.org/10.1002/ajhb.22358

Bodner, M., Perego, U. a, Huber, G., Fendt, L., Ro, A. W., Lancioni, H., … Torroni, A. (2012). Rapid coastal spread of First Americans: Novel insights from South America's Southern Cone mitochondrial genomes. *Genome Research*, 811–820. http://doi.org/10.1101/gr.131722.111

Borrero, L. A. (2015). Moving: Hunter-gatherers and the cultural geography of South America. *Quaternary International*, *363*, 1–8.

Campbell, R. J. (2011). *Socioeconomic differentiation, leadership, and residential patterning at an araucanian chiefly center (Isla Mocha, AD 1000-1700)*. University of Pittsburgh.

Campbell, R., & Quiroz, D. (2014). Chronological database for Southern Chile (35°30'-42° S), ~33000 BP to present: Human implications and archaeological biases. *Quaternary International*, *356*, 39–53. http://doi.org/10.1016/j.quaint.2014.07.026

Capriles, J. M., & Albarracin-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

Cardona, A., Pagani, L., Antao, T., Lawson, D. J., Eichstaedt, C. a, Yngvadottir, B., … Kivisild, T. (2014). Genome-wide analysis of cold adaptation in indigenous Siberian populations. *PloS One*, *9*(5), e98076. http://doi.org/10.1371/journal.pone.0098076

Chen, H., Patterson, N., & Reich, D. (2010). Population differentiation as a test for selective sweeps. *Genome Research*, *20*(3), 393–402. http://doi.org/10.1101/gr.100545.109

D'Urso, C. M., Wang, Z., Cao, Y., Tatake, R., Zeff, R. a., & Ferrone, S. (1991). Lack of HLA class I antigen expression by cultured melanoma cells FO-1 due to a defect in B2m gene expression. *Journal of Clinical Investigation*, *87*(1), 284–292. http://doi.org/10.1172/JCI114984

Danecek, P., Auton, A., Abecasis, G., Albers, C. a., Banks, E., DePristo, M. a., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. http://doi.org/10.1093/bioinformatics/btr330

Darwin, C. (1859). *On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life.* London ; Henry Frowde,. http://doi.org/10.5962/bhl.title.2109

Daub, J. T., Dupanloup, I., Robinson-Rechavi, M., & Excoffier, L. (2015). Inference of evolutionary forces acting on human biological pathways. *Genome Biology and Evolution*, *7*(6), 1546–1558. http://doi.org/10.1093/gbe/evv083

De Saint Pierre, M., Gandini, F., Perego, U. a, Bodner, M., Gómez-Carballa, A., Corach, D., … Olivieri, A. (2012). Arrival of Paleo-Indians to the southern cone of South America: new clues from mitogenomes. *PloS One*, *7*(12), e51311. http://doi.org/10.1371/journal.pone.0051311

Dillehay, T. D., Ramírez, C., Pino, M., Collins, M. B., Rossen, J., & Pino-Navarro, J. D. (2008). Monte Verde: seaweed, food, medicine, and the peopling of South America. *Science (New York, N.Y.)*, *320*(5877), 784–6. http://doi.org/10.1126/science.1156533

Eichstaedt, C. a., Antao, T., Cardona, A., Pagani, L., Kivisild, T., & Mormina, M. (2015). Positive selection of AS3MT to arsenic water in Andean populations. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. http://doi.org/10.1016/j.mrfmmm.2015.07.007

Eichstaedt, C. a., Antão, T., Pagani, L., Cardona, A., Kivisild, T., & Mormina, M. (2014). The Andean adaptive toolkit to counteract high altitude maladaptation: Genome-wide and phenotypic analysis of the Collas. *PLoS ONE*, *9*(3). http://doi.org/10.1371/journal.pone.0093314

Fix, A. G. (1999). *Migration and Colonization in Human Microevolution*. Cambridge: Cambridge University Press.

Fumagalli, M., Sironi, M., Pozzoli, U., Ferrer-Admettla, A., Pattini, L., & Nielsen, R. (2011). Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genetics*, *7*(11). http://doi.org/10.1371/journal.pgen.1002355

Galanter, J. M., Gignoux, C. R., Torgerson, D. G., Roth, L. a., Eng, C., Oh, S. S., … Burchard, E. G. (2014). Genome-wide association study and admixture mapping identify different asthma-associated loci in Latinos: The Genes-environments & Admixture in Latino Americans study. *Journal of Allergy and Clinical Immunology*, *134*(2), 295–305. http://doi.org/10.1016/j.jaci.2013.08.055

Gayo, E. M., Latorre, C., & Santoro, C. M. (2015). Timing of occupation and regional settlement patterns revealed by time-series analyses of an archaeological radiocarbon database for the South-Central Andes (16°–25°S). *Quaternary International*, *356*, 4–14. http://doi.org/10.1016/j.quaint.2014.09.076

George, C. M., Sima, L., Helena, M., Arias, J., Mihalic, J., Cabrera, L. Z., … Marie, C. (2014). Arsenic exposure in drinking water: an unrecognized health threat in Peru. *Bull World Health Organ*, *92*, 565–572.

Gissi, N. (1997). *Aproximación al conocimiento de la memoria Mapuche-Huilliche en San Juan de la Costa*. Universidad de Chile.

Grossman, S. R., Andersen, K. G., Shlyakhter, I., Winnicki, S., Yen, A., Park, D. J., … Rinn, J. L. (2014). Identifying Recent Adaptations in Large-scale Genomic Data. *Cell*, *152*(4), 703–713. http://doi.org/10.1016/j.cell.2013.01.035.Identifying

Hancock, A. M., & Di Rienzo, A. (2008). Detecting the Genetic Signature of Natural Selection in Human Populations: Models, Methods, and Data. *Annual Review of Anthropology*, *37*, 197–217. http://doi.org/10.1146/annurev.anthro.37.081407.085141

Hancock, A. M., Witonsky, D. B., Alkorta-Aranburu, G., Beall, C. M., Gebremedhin, A., Sukernik, R., … Di Rienzo, A. (2011). Adaptations to climate-mediated selective pressures in humans. *PLoS Genetics*, *7*(4), e1001375. http://doi.org/10.1371/journal.pgen.1001375

Hartl, D. L., & Clark, A. G. (1997). *Principles of population genetics* (Third). Massachusetts.

Higasa, K., Kukita, Y., Kato, K., Wake, N., Tahira, T., & Hayashi, K. (2009). Evaluation of Haplotype Inference Using Definitive Haplotype Data Obtained from Complete Hydatidiform Moles, and Its Significance for the Analyses of Positively Selected Regions. *PLoS Genetics*, *5*(5). http://doi.org/10.1371/journal.pgen.1000468

Hoffmann, T. J., Zhan, Y., Kvale, M. N., Hesselson, S. E., Gollub, J., Iribarren, C., … Risch, N. (2011). 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

Holsinger, K. E., & Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting Fst. *Nature Reviews. Genetics*, *10*(9), 639–650. http://doi.org/10.1038/nrg2611

Hsu, F., Kent, J. W., Clawson, H., Kuhn, R. M., Diekhans, M., & Haussler, D. (2006). The UCSC known genes. *Bioinformatics*, *22*(9), 1036–1046. http://doi.org/10.1093/bioinformatics/btl048

Jablonski, N. G., & Chaplin, G. (2010). Human skin pigmentation as an adaptation to UV radiation. *Proceedings of the National Academy of Sciences of the United States of America*, *107 Suppl*, 8962–8968. http://doi.org/10.1073/pnas.0914628107

Jeong, C., & Di Rienzo, A. (2014). Adaptations to local environments in modern human populations. *Current Opinion in Genetics and Development*, *29*, 1–8. http://doi.org/10.1016/j.gde.2014.06.011

Ko, A., Cantor, R. M., Weissglas-Volkov, D., Nikkola, E., Reddy, P. M. V. L., Sinsheimer, J. S., … Pajukanta, P. (2014). Amerindian-specific regions under positive selection harbour new lipid variants in Latinos. *Nature Communications*, *5*, 3983. http://doi.org/10.1038/ncomms4983

Laland, K. N., Odling-Smee, J., & Myles, S. (2010). How culture shaped the human genome: bringing genetics and the human sciences together. *Nature Reviews. Genetics*, *11*(2), 137–148. http://doi.org/10.1038/nrg2734

Lappalainen, T., Salmela, E., Andersen, P. M., Dahlman-Wright, K., Sistonen, P., Savontaus, M.-L., … Kere, J. (2010). Genomic landscape of positive natural selection in Northern European populations. *European Journal of Human Genetics*, *18*(4), 471–478. http://doi.org/10.1038/ejhg.2009.184

Levy, D., Larson, M. G., Benjamin, E. J., Newton-Cheh, C., Wang, T. J., Hwang, S.-J., … Mitchell, G. F. (2007). Framingham Heart Study 100K Project: genome-wide associations for blood pressure and arterial stiffness. *BMC Medical Genetics*, *8 Suppl 1*, S3. http://doi.org/10.1186/1471-2350-8-S1-S3

Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., Ramachandran, S., … Myers, R. M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science (New York, N.Y.)*, *319*(5866), 1100–4. http://doi.org/10.1126/science.1153717

Liu, X., Ong, R. T.-H., Pillai, E. N., Elzein, A. M., Small, K. S., Clark, T. G., … Teo, Y.-Y. (2013). Detecting and characterizing genomic signatures of positive selection in global populations. *American Journal of Human Genetics*, *92*(6), 866–81. http://doi.org/10.1016/j.ajhg.2013.04.021

Liu, Y., Nyunoya, T., Leng, S., Belinsky, S. a, Tesfaigzi, Y., & Bruse, S. (2013). Softwares and methods for estimating genetic ancestry in human populations. *Human Genomics*, *7*(1), 1. http://doi.org/10.1186/1479-7364-7-1

Llop, E., Harb D., Z., Acuña, M., Moreno, R., Barton, S., Aspillaga, E., & Rothhammer, F. (1993). Composición genética de la población chilena: los Pehuenches de Trapa-Trapa. *Rev. Med. Chile*, *121*, 494–498.

López Herráez, D., Bauchet, M., Tang, K., Theunert, C., Pugach, I., Li, J., … Stoneking, M. (2009). Genetic variation and recent positive selection in worldwide human populations: Evidence from nearly 1 million SNPs. *PLoS ONE*, *4*(11). http://doi.org/10.1371/journal.pone.0007888

Mi, H., Muruganujan, A., Casagrande, J. T., & Thomas, P. D. (2013). Large-scale gene function analysis with the PANTHER classification system. *Nature Protocols*, *8*(8), 1551–66. http://doi.org/10.1038/nprot.2013.092

Moran, E. F. (2008). *Human adaptability. An introduction to ecological anthropology* (Third). Westview Press.

Moreno-Estrada, A., Gignoux, C. R., Fernández-López, J. C., Zakharia, F., Sikora, M., Contreras, A. V, … Bustamante, C. D. (2014). The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science (New York, N.Y.)*, *344*(6189), 1280–5. http://doi.org/10.1126/science.1251688

Murabito, J. M., Rosenberg, C. L., Finger, D., Kreger, B. E., Levy, D., Splansky, G. L., … Hwang, S.-J. (2007). A genome-wide association study of breast and prostate cancer in the NHLBI's Framingham Heart Study. *BMC Medical Genetics*, *8 Suppl 1*, S6. http://doi.org/10.1186/1471-2350-8-S1-S6

O'Connell, J., Gurdasani, D., Delaneau, O., Pirastu, N., Ulivi, S., Cocca, M., … Marchini, J. (2014). A General Approach for Haplotype Phasing across the Full Spectrum of Relatedness. *PLoS Genetics*, *10*(4). http://doi.org/10.1371/journal.pgen.1004234

Perego, U. a, Angerhofer, N., Pala, M., Olivieri, A., Lancioni, H., Hooshiar Kashani, B., … Torroni, A. (2010). The initial peopling of the Americas: a growing number of founding mitochondrial genomes from Beringia. *Genome Research*, *20*(9), 1174–9. http://doi.org/10.1101/gr.109231.110

Pickrell, J. K., Coop, G., & Novembre, J. (2009). Signals of recent positive selection in a worldwide sample of human populations, 826–837. http://doi.org/10.1101/gr.087577.108

Pritchard, J. K., & Di Rienzo, A. (2010). Adaptation - not by sweeps alone. *Nature Reviews. Genetics*, *11*(10), 665–667. http://doi.org/10.1038/nrg2880

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–59. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1461096&tool=pmcentrez&rendertype=abstract

Pybus, M., Dall'Olio, G. M., Luisi, P., Uzkudun, M., Carreño-Torres, A., Pavlidis, P., … Engelken, J. (2013). 1000 Genomes Selection Browser 1.0: a genome browser dedicated to signatures of natural selection in modern humans. *Nucleic Acids Research*, *42*(Database issue), D903–9. http://doi.org/10.1093/nar/gkt1188

R Core Team, T. (2013). R: A Language and Environment for Statistical Computing.

Raghavan, M., Steinrücken, M., Harris, K., Schiffels, S., Degiorgio, M., Albrechtsen, A., … Fuller, B. T. (2015). Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science*, (July), 1–20.

Reference Genome Group of the Gene Ontology Consortium, T. (2009). The gene ontology's reference genome project: A unified framework for functional annotation across species. *PLoS Computational Biology*, *5*(7). http://doi.org/10.1371/journal.pcbi.1000431

Reznick, D. N., & Ricklefs, R. E. (2009). Darwin's bridge between microevolution and macroevolution. *Nature*, *457*(February), 837–842. http://doi.org/10.1038/nature07894

Rothhammer, F., & Dillehay, T. D. (2009). The late pleistocene colonization of South America: An interdisciplinary perspective. *Annals of Human Genetics*, *73*(5), 540–549. http://doi.org/10.1111/j.1469-1809.2009.00537.x

Rothhammer, F., Fuentes Guajardo, M., Chakraborty, R., Lorenzo Bermejo, J., & Dittmar, M. (2015). Neonatal Variables, Altitude of Residence and Aymara Ancestry in Northern Chile. *Plos One*, *10*(4), e0121834. http://doi.org/10.1371/journal.pone.0121834

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., & Hochachka, P. W. (2001). The evidence for hereditary factors contributing to high altitude adaptation in Andean natives: a review. *High Altitude Medicine & Biology*, *2*(2), 235–256. http://doi.org/10.1089/152702901750265332

Sabeti, P. C., Schaffner, S. F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., … Lander, E. S. (2006). Positive natural selection in the human lineage. *Science (New York, N.Y.)*, *312*(5780), 1614–20. http://doi.org/10.1126/science.1124309

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., … Lander, E. S. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, *449*(October), 913–919. http://doi.org/10.1038/nature06250

Scally, A., & Durbin, R. (2012). Revising the human mutation rate: implications for understanding human evolution. *Nature Reviews Genetics*, *13*(11), 824–824. http://doi.org/10.1038/nrg3353

Schlebusch, C. M., Gattepaille, L. M., Engström, K., Vahter, M., & Broberg, K. (2015). Human Adaptation to Arsenic-Rich Environments. *Molecular Biology and Evolution*, *32*(6), 1544–1555.

Schurr, T. G. (2004). The Peopling of the New World: Perspectives from Molecular Anthropology. *Annual Review of Anthropology*, *33*(1), 551–583. http://doi.org/10.1146/annurev.anthro.33.070203.143932

Schutkowski, H. (2006). *Human Ecology. Biocultural adaptations in human communities* (Vol. 182).

Scliar, M. O., Gouveia, M. H., Benazzo, A., Ghirotto, S., Fagundes, N., Leal, T. P., … Tarazona-Santos, E. (2014). Bayesian inferences suggest that Amazon Yunga Natives diverged from Andeans less than 5000 ybp: implications for South American prehistory. *BMC Evolutionary Biology*, *14*(1), 174. http://doi.org/10.1186/s12862-014-0174-3

Simoons, F. J. (1969). Primary adult lactose intolerance and the milking habit: a problem in biological and cultural interrelations. *The American Journal of Digestive Diseases*, *14*(12), 819–836.

Sokal, R., & Michener, C. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Scientific Bulletin*, *28*, 1409–1438.

Szpiech, Z. a, & Hernandez, R. D. (2014). selscan: An Efficient Multithreaded Program to Perform EHH-Based Scans for Positive Selection. *Molecular Biology and Evolution*, (3), 1–4. http://doi.org/10.1093/molbev/msu211

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, *30*(12), 2725–2729. http://doi.org/10.1093/molbev/mst197

Tang, K., Thornton, K. R., & Stoneking, M. (2007). A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biology*, *5*(7), e171. http://doi.org/10.1371/journal.pbio.0050171

The international HapMap 3 Consortium. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature*, *467*(September). http://doi.org/10.1038/nature09298

The Neutral Theory and Tests of Neutrality. (2013). In *An Introduction to Population Genetics* (First Ed., pp. 179–193). Sunderlan, MA 01375, USA: Sinauer Associates, Inc.

Tian, C., Kosoy, R., Nassir, R., Lee, A., Villoslada, P., Klareskog, L., … Seldin, M. F. (2009). European population genetic substructure: further definition of ancestry informative markers for distinguishing among diverse European ethnic groups. *Molecular Medicine (Cambridge, Mass.)*, *15*(11-12), 371–383. http://doi.org/10.2119/molmed.2009.00094

Torrejón, F. (2001). Variables geohistóricas en la evolución del sistema económico pehuenche durante el periodo colonial. *Revista Universum*, (16), 12–28.

Urbina, M. X. (2009). *La frontera de arriba en Chile colonial. Interacción hispano-indígena en el territorio entre Valdivia y Chiloé e imaginario de sus bordes geográficos, 1600-1800*. Chile: Ediciones Universitarias de Valparaiso.

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

Voight, B. F., Kudaravalli, S., Wen, X., & Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biology*, *4*(3), e72. http://doi.org/10.1371/journal.pbio.0040072

Wang, S., Lewis, C. M., Jakobsson, M., Ramachandran, S., Ray, N., Bedoya, G., … Ruiz-Linares, A. (2007). Genetic variation and population structure in Native Americans. *PLoS Genetics*, *3*(11), 2049–2067. http://doi.org/10.1371/journal.pgen.0030185

Weir, B. S., Cardon, L. R., Anderson, A. D., Nielsen, D. M., & Hill, W. G. (2005a). Measures of human population structure show heterogeneity among genomic regions. *Genome Research*, *15*(11), 1468–1476. http://doi.org/10.1101/gr.4398405

Weir, B. S., Cardon, L. R., Anderson, A. D., Nielsen, D. M., & Hill, W. G. (2005b). Measures of human population structure show heterogeneity among genomic regions. *Genome Research*, *15*(11), 1468–76. http://doi.org/10.1101/gr.4398405

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., … Parkinson, H. (2013). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research*, *42*(D1), 1001–1006. http://doi.org/10.1093/nar/gkt1229

White, L., Erlebach, E., Weihmann, A., Aida, M., & Castellano, S. (2015). Genetic adaptation to levels of dietary selenium in recent human history. *Molecular Biology and Evolution*, 1–28.

Wilk, J. B., Walter, R. E., Laramie, J. M., Gottlieb, D. J., & O'Connor, G. T. (2007). Framingham Heart Study genome-wide association: results for pulmonary function measures. *BMC Medical Genetics*, *8 Suppl 1*, S8. http://doi.org/10.1186/1471-2350-8-S1-S8

Wollstein, A., & Stephan, W. (2015). Inferring positive selection in humans from genomic data. *Investigative Genetics*, *6*(1), 1–8. http://doi.org/10.1186/s13323-015-0023-1

Yang, Q., Kathiresan, S., Lin, J.-P., Tofler, G. H., & O'Donnell, C. J. (2007). Genome-wide association and linkage analyses of hemostatic factors and hematological phenotypes in the Framingham Heart Study. *BMC Medical Genetics*, *8 Suppl 1*, S12. http://doi.org/10.1186/1471-2350-8-S1-S12

Anexos

Apéndice 1: Evaluación del programa selscan en una región control

A forma de control positivo, evaluamos nuestra capacidad de detectar señales de selección positiva usando datos de 49 *SNPs* del locus LCT en los individuos CEU de Hapmap fase 3 (The international HapMap 3 Consortium, 2010). Este locus es ampliamente reconocido como seleccionado positivamente en europeos. Como resultado, encontramos 37 *SNPs* con altos valores de iHS (≥2). A modo de comparación, inspeccionamos los resultados de iHS para este locus reportados en la literatura.

Los puntajes de iHS en la región del gen LCT y el número de marcadores evaluados resultaron ser distintos entre las bases de datos inspeccionadas (Tabla 4). Para nuestra sorpresa, el Proyecto 1000 Genomas no cuenta con ningún SNP en esta región a pesar de contar con una gran cantidad de marcadores distribuidos a lo largo del genoma. HapMap y HGDP sí muestran señales de selección en la región del gen LCT, aunque con una gran diferencia en el número de marcadores con valores altos: HapMap muestra una fuerte señal con 61 *SNPs*, mientras que solo 2 marcadores fueron seleccionados en HGDP.

Base de datos	Nº indiv.	*SNPs* en región LCT	Puntaje promedio región LCT	Puntaje máximo región LCT	*SNPs* con iHS >2
Resultados de iHS en este estudio con selscan					
HapMap fase 3 *Este estudio*	112	49	2,605	4,00	37
Resultados de iHS publicados					
HapMap fase 2 *Voight et al. 2006*	60	82	2,581	3,884	61
1000G *Pybus et al. 2013*	97	0	0	0	0
HGDP *Pickrell et al. 2009*	161	27	1,049	2,768	2

Tabla 2: Puntajes de iHS en europeos publicados y calculados por selscan, para la región del gen LCT (cromosoma 2: 136.545.410-136.594.750)

Otros trabajos han discutido problemas para detectar señales de selección en LCT. La ausencia de *SNPs* en la región del gen LCT en la base de datos de 1000 Genomas se explica en su publicación (Pybus et al., 2013) como resultado de las características inherentes del test, porque: (1) el haplotipo seleccionado está demasiado cercano a la

fijación o (2) el valor de EHH dentro de una ventana de 200kb no decae hasta 0.15. Este valor de corte de EHH es una modificación de lo que se usa en Voight 2006 y en el resto de las publicaciones de las bases de datos, en las que se utiliza un valor de 0.05.

Por otro lado, en la publicación de los resultados para LCT del panel HGDP, explican que el microarreglo utilizado fue diseñado con una baja densidad de marcadores en las regiones con alto desequilibrio de ligamiento, lo que disminuiría la capacidad de identificar esta característica propia de los barridos selectivos. Esto se evidencia en que, en las regiones de 200 kb con señales de selección más potente según HapMap, el 25% tiene menos de 20 *SNPs* en el microarreglo Illumina usado en HGDP (es decir, son regiones no evaluadas por el test) (Pickrell et al., 2009). Esto es notoriamente menor respecto al promedio de 40 *SNPs* por ventana en HGDP y de 180 *SNPs* en HapMap. Pickrell y colaboradores (2009) concluyen en este trabajo que el uso de estos datos tiene un bajo poder para confirmar los barridos selectivos identificados por HapMap, pero que no afectaría la capacidad de detectar barridos selectivos nuevos en otras poblaciones.

Por último, es posible que los valores de iHS para LCT en HGDP sean más bajos que en HapMap debido a que la muestra utilizada comprende individuos de diversos sectores de Europa, donde no se ha identificado selección positiva asociada al gen LCT, mientras que en HapMap se usan muestras con ancestría del Norte y Oeste de Europa, mostrando una señal de selección positiva fuerte.

Apéndice 2: Estadísticos de selección positiva utilizados

Aquí describimos en detalle los métodos implementados por el programa selscan para identificar señales de selección positiva, según Szpiech & Hernandez (2014).

iHS (Integrated Haplotype score)

En una muestra de *n* cromosomas, *C* denota el conjunto de todos los haplotipos distintos posibles en un locus de interés (x_0), y $C(x_i)$ denota el set de todos los haplotipos distintos posibles entre el locus x_0 y el *i* marcador río arriba o río abajo desde el x_0 o SNP central. Siendo x_0 un SNP bialélico donde 0 representa al alelo ancestral y 1 al alelo derivado, $C := \{0,1\}$. Si x_1 es un marcador inmediatamente adyacente al locus x_0, entonces el conjunto de todos los haplotipos posibles será $C(x_1) := \{11, 10, 00, 01\}$. El EHH para toda la muestra, desde el locus x_0 al marcador x_1 se calcula:

$$EHH(x_i) = \sum_{h \epsilon C(x_i)} \frac{\binom{n_h}{2}}{\binom{n}{2}}$$

donde n_h es el número de haplotipos observado de tipo $h \epsilon C(x_i)$.

Si consideramos a $\mathcal{H}_c(x_i)$ un subgrupo de $C(x_1)$ que contiene los distintos haplotipos que portan el alelo central, entonces:

$$C(x_i) = \bigcup_{c \in C} \mathcal{H}_c(x_i)$$

Si el alelo derivado (1) es considerado el alelo central, entonces $\mathcal{H}_1(x_1) = \{11, 10\}$. De forma semejante, si el alelo ancestral (0) es considerado alelo central, entonces $\mathcal{H}_0(x_1) = \{00, 01\}$. El valor de EHH de los cromosomas que cargan el haplotipo central c en el marcador x_1 se calcula:

$$EHH_c(x_i) = \sum_{h \epsilon \mathcal{H}(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$

donde n_h es el número de haplotipos observado de tipo $h \epsilon \mathcal{H}(x_i)$ y n_c es el número de haplotipos observados que portan el haplotipo central ($c \epsilon C$).

De esta forma, iHS se calcula a partir de esta última ecuación, para registrar el decaimiento de la homocigosidad haplotípica en los haplotipos que porten el alelo ancestral y el derivado en el SNP central. En primer lugar, se calcula iHH (homocigosidad haplotípica integrada) para alelo ancestral y derivado ($C := \{0,1\}$), de la siguiente forma:

$$iHH_c = \sum_{i=1}^{|D|} \frac{1}{2}\big(EHH_c(x_{i-1}) + EHH_c(x_i)\big)g(x_{i-1}, x_i) + \sum_{i=1}^{|U|} \frac{1}{2}\big(EHH_c(x_{i-1}) + EHH_c(x_i)\big)g(x_{i-1}, x_i)$$

donde D es el set de marcadores río abajo desde el locus central, y $x_i \in D$ denota al marcador número *i* más cercano al marcador de interés x_0. De forma semejante a D se define U, pero en la dirección opuesta (río arriba). $g(x_{i-1}, x_i)$ representa la distancia genética entre los dos marcadores. La sumatoria se detiene con un valor de $EHH_c(x_i) < 0.05$. El iHS no estandarizado, se calcula como

$$\ln\left(\frac{iHH_1}{iHH_0}\right)$$

Si el valor del decaimiento de EHH es similar entre los alelos evaluados, el valor de la razón entre iHH_1 e iHH_0 será cercano a 1, y el logaritmo natural de la razón (el iHS no estandarizado) será cercano a 0. Valores grandes y negativos indicarán haplotipos largos con el alelo 0, y valores grandes positivos indicarán haplotipos largos con el alelo 1.

En modelos neutrales, los alelos con baja frecuencia son generalmente los más jóvenes, y se asocian con haplotipos más largos que los alelos de alta frecuencia y por tanto, más antiguos. Por esto, la ecuación presentada es ajustada para obtener un estadístico final con una media igual a 0 y varianza de 1, sin importar la frecuencia alélica del SNP central. Los puntajes no estandarizados se normalizan en grupos de marcadores creados a partir de sus frecuencias alélicas

$$iHS = \frac{\ln\left(\frac{iHH_1}{iHH_0}\right) - E_p\left[\ln\left(\frac{iHH_1}{iHH_0}\right)\right]}{SD_p\left[ln\left(\frac{iHH_1}{iHH_0}\right)\right]}$$

donde la esperanza y la desviación estándar son las equivalentes a cada agrupación de marcadores. Cuando hay baja densidad de *SNPs*, con una distancia física de b mayor a 20 kb entre dos marcadores, entonces $g(x_{i-1}, x_i)$ se multiplica por un factor de $20/b$, para reducir posibles señales espurias inducidas por ausencia de datos en una región. Si al calcular iHH se alcanza el límite de un cromosoma antes de que $EHH_c(x_i) < 0.05$, o si se encuentra con una zona mayor a 200 kb sin datos, no se realiza el calculo para ese locus.

XP-EHH (Cross Population Extended Haplotype Homozygosity)

Para calcular XP-EHH entre las poblaciones A y B para un marcador x_0, primero se calcula iHH para cada población de forma separada, integrando el EHH de todas las muestras en esa población:

$$iHH = \sum_{i=1}^{|D|} \frac{1}{2}\big(EHH(x_{i-1}) + EHH(x_i)\big)g(x_{i-1}, x_i) + \sum_{i=1}^{|U|} \frac{1}{2}\big(EHH(x_{i-1}) + EHH(x_i)\big)g(x_{i-1}, x_i)$$

Las sumatorias en el calculo de iHH para cada población se detienen en el marcador donde el EHH de todos los haplotipos de ambas poblaciones es menor a 0,05 ($EHH(x_i) < 0.05$). De la misma forma en que se calculó iHS, se escala el valor de $g(x_{i-1}, x_i)$ multiplicandolo por $20/b$, y no se calculan puntajes en marcadores cuyo $EHH(x_i)$ no haya disminuido hasta menos de 0,05 al llegar al límite de un cromosoma o a regiones sin datos mayores a 200 kb.

Si $\mathrm{iHH_A}$ e $\mathrm{iHH_B}$ corresponden a los valores integrados de EHH para las poblaciones A y B, entonces el XP-EHH no estandarizado es:

$$\ln\left(\frac{iHH_A}{iHH_B}\right)$$

y la normalización del genoma completo:

$$XP - EHH = \frac{\ln\left(\frac{iHH_A}{iHH_B}\right) - E\left[\ln\left(\frac{iHH_A}{iHH_B}\right)\right]}{SD\left[ln\left(\frac{iHH_A}{iHH_B}\right)\right]}$$

El estadístico $\ln\left(iHH_A/iHH_B\right)$, resultará en un valor inusualmente positivo cuando sugiere selección en la población A, y un valor negativo cuando propone selección en población B.

Tablas suplementarias

Tabla S1: iHS Amerindios

Ranking	Chr	Inicio	Termino	NºSNPs	Prop. iHS	Genes	Fenotipos (GWAS p<10-8)	Exclusiva amerindia
1	chr6	52000001	52200001	25	1	MIR206, MIR133B, IL17F, IL17A, MCM3	IL17F: Response to gemcitabine in pancreatic cancer	*
2	chr6	52200001	52400001	28	0.928571	PAQR8, EFHC1, TRAM2		*
3	chr5	161200001	161400001	22	0.909091	Gabra1		*
4	chr6	52400001	52600001	32	0.875	LOC730101, TRAM2-AS1,TMEM14A, TRAM2		*
5	chr11	65400001	65600001	25	0.84	MIR4690, SIPA1, MIR4489, DQ587981, AP5B1, OVOL1, PCNXL3, RELA, KAT5, RNASEH2C	OVOL1: Atopic dermatitis	*
6	chr11	64800001	65000001	23	0.782609	SAC3D1, AB231703, ZFPL1, LOC100130348, ZNHIT2, BC104003, FAU, MRPL49, SPDYC, SLC22A20, ARL2-SNX15, NAALADL1, CDCA5, VPS51, SYVN1, CAPN1, SLC22A20, TM7SF		
7	chr9	10000001	10200001	32	0.78125	PTPRD	Type 2 diabetes	*
8	chr2	102800001	103000001	51	0.745098	IL1RL2, IL1RL1, IL18R1	IL1RL1: Celiac disease, eosinophil counts	
9	chr15	68000001	68200001	35	0.742857	SKOR1, MAP2K5	Body mass index	
10	chr2	198600001	198800001	53	0.735849	BOLL, BC021693, PLCL1	Intracranial aneurysm	*
11	chr4	102000001	102200001	22	0.727273	PPP3CA	HP,HPR,DHX38: LDL, cholesterol, aptoglobin levels	
12	chr16	72000001	72200001	20	0.7	PKD1L3, DHODH, HPR, TXNL4B, DHX38, PMFBP1, HP		*
13	chr2	160800001	161000001	20	0.7	ITGB6, PLA2R1	Nephropathy (idiopathic membranous)	*
14	chr13	81000001	81200001	26	0.692308	-		*
15	chr11	78000001	78200001	67	0.686567	GAB2, NARS2	Alzheimer, Menarche (age at onset)	*
16	chr22	48000001	48200001	34	0.676471	LINC00898, AK093107		
17	chr1	159200001	159400001	48	0.666667	OR10J3, FCER1A, BC038194	IgE levels	

18	chr4	179000001	179200001	21	0.666667	-		*
19	chr19	32600001	32800001	26	0.653846	-		*
20	chr1	170600001	170800001	20	0.65	PRRX1		*
21	chr5	164400001	164600001	22	0.636364	-		*
22	chr12	114400001	114600001	26	0.615385	RBM19		*
23	chr2	160600001	160800001	26	0.615385	MARCH7, LY75-CD302, PLA2R1		
24	chr4	60400001	60600001	28	0.607143	-		*
25	chr15	78400001	78600001	25	0.6	CIB2,IDH3A,DNAJA4, WDR61, ACSBG1		*
26	chr15	78600001	78800001	30	0.6	Mir_544, CRABP1, AGPHD1, IREB2	Chronic obstructive pulmonary disease	
27	chr1	152800001	153000001	32	0.59375	LCE1A, LCE6A, SMCP, IVL, SPRR1A, SPRR3, SPRR4, PDGFC, GLRB		*
28	chr4	157800001	158000001	26	0.576923	ATG7, PDGFC, GLRB		
29	chr2	169800001	170000001	40	0.575	DHRS9, LRP2, ABCB11	Liver enzyme levels (alkaline phosphatase)	
30	chr1	204400001	204600001	21	0.571429	AK097184, PIK3C2B, MDM4, LRRN2		*
31	chr2	103000001	103200001	60	0.566667	MIR4772, IL18R1, IL18RAP, SLC9A4	Celiac disease, Chron	
32	chr5	150600001	150800001	50	0.56	GM2A, SLC36A3, SLC36A2, CCDC69		*
33	chr5	160600001	160800001	25	0.56	GABRB2		*
34	chr5	160000001	160200001	27	0.555556	Mir_544, ATP10B		*
35	chr11	26800001	27000001	20	0.55	-		
36	chr22	30400001	30600001	22	0.545455	HORMAD2, MTMR3	Crohn, Inflammatory bowel disease (early onset), Nephropathy	
37	chr6	156400001	156600001	22	0.545455	-		*
38	chr10	65000001	65200001	28	0.535714	AX747628, MIR1296, JMJD1C, NRBF2, REEP3,	Sex hormone-binding globulin levels, Liver enzyme levels (alkaline phosphatase), Mean platelet volume, Triglycerides	*
39	chr17	7400001	7600001	28	0.535714	SENP3-EIF4A1, SNORA48, SNORD10, SNORA67, CD68, SOX15, TRNA_Pseudo, SAT2, ATP1B2, HV941431, HV941433, HV941428, HV941434, HV941486, HV941429, HV941440,	SHBG: Sex hormone-binding globulin levels, Androgen levels, Testosterone levels MPDU1: nephropathy	*

						HV941478, HV941442, HV941444, HV941430, POLR2A, TNFSF12-TNFSF13, MPDU1, FXR2, SHBG, TP53, SENP3, WRAP53		
40	chr11	77800001	78000001	60	0.533333	ALG8, KCTD21, USP35, LOC100289388, GAB2		
41	chr12	114600001	114800001	38	0.526316	TBX5	Electrocardiographic traits, Ventricular conduction	*
42	chr2	42800001	43000001	21	0.52381	OXER1, HAAO, MTA3		*
43	chr20	52600001	52800001	63	0.52381	MIR4756, BCAS1, CYP24A1	CYP24A1 Multiple sclerosis,	*
44	chr9	10200001	10400001	32	0.520833	PTPRD	Diabetes tipo 2	*
45	chr6	33200001	33400001	29	0.517241	HCG25, B3GALT4, WDR46, PFDN6, ZBTB22, DAXX, BC146941, PHF1, CUTA, VPS52, RPS18, RGL2, TAPBP, KIFC1, SYNGAP1	LDL cholesterol	
46	chr6	51400001	51600001	31	0.516129	PKHD1		*
47	chr10	64800001	65000001	22	0.5	NRBF2, JMJD1C		*
48	chr13	44400001	44600001	28	0.5	LACC1, LINC00284, CCDC122		
49	chr5	161400001	161600001	22	0.5	GABRG2		*
50	chr19	18800001	19000001	27	0.481481	COMP, UPF1, CERS1, CRTC1	CRTC1: menarche (age at oncet)	*
51	chr6	154400001	154600001	27	0.481481	OPRM1, IPCEF1		*
52	chr6	168000001	168200001	27	0.481481	AK127120, LOC441178, C6orf123		
53	chr22	46200001	46400001	25	0.48	WNT7B, ATXN10		
54	chr1	171000001	171200001	46	0.478261	MIR1295A, MIR1295B, FMO3, FMO6P, FMO2, MROH9	FMO3, FMO6: Lentiform nucleus volume	
55	chr7	139400001	139600001	23	0.478261	TBXAS1, HIPK2		
56	chr4	179200001	179400001	21	0.47619	-		*
57	chr3	10800001	11000001	34	0.470588	LINC00606, SLC6A11		
58	chr6	24400001	24600001	59	0.457627	MRS2, GPLD1, ALDH5A1, KIAA0319	ALDH5A1, GPLD1: Liver enzyme levels (alkaline phosphatase)	*
59	chr17	39000001	39200001	53	0.45283	KRT12, KRT20, KRT39, KRTAP3-3, KRTAP3-2, KRTAP3-1, KRTAP1-5, KRTAP1-4, KRTAP1-1, KRT23, KRT40, KRTAP1-3,		

60	chr16	11400001	11600001	42	0.452381	RMI2, AK126539		*
61	chr12	30400001	30600001	20	0.45	-		
62	chr2	85600001	85800001	20	0.45	Y_RNA, SH2D6, LOC100630918, CAPG, MAT2A, GGCX, ELMOD3,	GGCX, VAMP8, VAMP5, RNF181: Prostate cancer	*
63	chr4	62800001	63000001	20	0.45	BC039452, LPHN3		*
64	chr4	82800001	83000001	20	0.45	-		*
65	chr12	114200001	114400001	29	0.448276	TRNA_Pseudo, AK096932, RBM19,		*
66	chr10	65200001	65400001	47	0.446809	JMJD1C-AS1, REEP3, JMJD1C		*
67	chr8	119800001	120000001	45	0.444444	TNFRSF11B	Osteoporosis	
68	chr5	90000001	90200001	34	0.441176	GPR98,		*
69	chr4	94800001	95000001	25	0.44	-		
70	chr12	48200001	48400001	48	0.4375	TMEM106C, HDAC7, VDR, COL2A1		*
71	chr13	80600001	80800001	30	0.433333	-		*
72	chr14	64600001	64800001	30	0.433333	SYNE2, ESR2	Atrial fibrillation	
73	chr14	24600001	24800001	37	0.432432	FITM1, PSME1, PSME2, IRF9, TSSK4, TINF2, DHRS1, CIDEB, LTB4R2, LTB4R, EMC9, RNF31, IPO4, TM9SF1, NEDD8-MDP1, GMPR2, TGM1, RABGGTA, HP08474, ADCY4, REC8, NOP9		*
74	chr1	153200001	153400001	35	0.428571	LOR, S100A9, S100A12, S100A8, S100A7A, PGLYRP3, PGLYRP4,		*
75	chr5	153600001	153800001	21	0.428571	GALNT10, SAP30L-AS1		*
76	chr16	80000001	80200001	28	0.428571	-		*
77	chr19	20800001	21000001	21	0.428571	ZNF626, ZNF66		*
78	chr3	11200001	11400001	21	0.428571	HRH1, ATG7		
79	chr15	58200001	58400001	35	0.428571	ALDH1A2		
Regiones candidatas (dentro del 1% del puntaje más alto) para el test iHS en Amerindios								

Tabla S2: XPEHH

Chr	inicio	termino	NºSNPs	XPEHH			Genes	Fenotipos (GWAS p<10-8)
				EAS	CEU	YRI		
chr1	53600001	53800001	46	3.469	3.711	3.736	C1orf123, MAGOH, AK097571, SLC1A7, CPT2, LOC100507564, LRP8	
chr1	166400001	166600001	44	3.506	3.679	3.349	FMO9P	
chr1	166800001	167000001	20	3.692	3.645	3.359	POGK, TADA1, ILDR2, MAEL	
chr11	93000001	93200001	26	3.813	3.473	3.379	CCDC67	
chr11	93200001	93400001	36	3.466	3.492	3.737	SMCO4, KIAA1731	
chr12	96200001	96400001	47	3.475	3.838	3.537	SNRPF, CCDC38, HAL, LTA4H, AMDHD1	Pulmonary function
chr12	109600001	109800001	43	4.464	3.362	3.706	ACACB, FOXN4	
chr14	48000001	48200001	22	4.044	4.021	3.587	MDGA2	
chr14	48200001	48400001	30	4.166	4.020	3.406	MIR548Y, LINC00648	
chr16	11000001	11200001	48	3.472	4.269	4.566	AX748339, DEXI, CIITA, CLEC16A	Multiple sclerosis, Type 1 diabetes, Primary biliary cirrosis, Crohn disease, psoriasis, Celiac disease
chr16	11200001	11400001	85	4.262	5.239	4.797	SOCS1, TNP2, PRM3, PRM2, PRM1, CLEC16A	
chr16	11400001	11600001	79	3.978	4.842	4.125	RMI2, AK126539	
chr5	161000001	161200001	19	4.626	3.391	3.473	GABRA6	
chr6	51800001	52000001	31	3.657	4.794	3.882	PKHD1	
chr6	52000001	52200001	72	4.891	4.202	4.227	MIR206, MIR133B, IL17F, IL17A, MCM3	Response to gemcitabine in pancreatic cancer
chr6	52200001	52400001	62	5.056	4.312	4.169	PAQR8, EFHC1, TRAM2	
chr6	52400001	52600001	51	4.719	4.067	4.193	LOC730101, TRAM2-AS1, TMEM14A, TRAM2	
chr7	133000001	133200001	11	3.513	3.862	4.093	EXOC4	
chr7	133200001	133400001	30	4.049	4.484	3.728	EXOC4	
chr7	133600001	133800001	24	3.482	3.428	3.436	EXOC4	
Regiones candidatas (dentro del 1% del puntaje más alto) para el test XPEHH al calcular la extension de haplotipos entre nativos amerindios y asiáticos, europeos y africanos.								

Tabla S3: Fst

chr	inicio	termino	NºSNPs	Fst EAS	Fst CEU	Fst YRI	Genes	Fenotipos (GWAS p<10-8)
chr1	100400001	100600000	25	0.910913	1	0.987599	BC112312, SLC35A3, HIAT1, SASS6, TRMT13	
chr1	171200001	171400000	76	1	1	0.962477	FMO1, Y_RNA, FMO4, TOP1P1	
chr11	56400001	56600000	69	0.987686	0.987686	0.987686	OR8U8, OR5AP2, OR5AR1, OR9G9, OR9G4, AB231741	Hemostatic factors and hematological phenotypes
chr11	63000001	63200000	75	1	1	0.987599	SLC22A10, SLC22A9	
chr13	52600001	52800000	25	0.987686	0.987686	0.987686	UTP14C, UTP14C, NEK5, NEK3, MRPS31P5	
chr13	111800001	112000000	133	0.910913	0.962477	0.962477	ARHGEF7, AX748212, TEX29	
chr16	16200001	16400000	101	1	1	1	ABCC1, ABCC6, NOMO3, Mir_548, MIR3179-2	
chr18	3600001	3800000	71	0.987686	0.974936	0.987686	DLGAP1, BC094703	
chr19	22600001	22800000	57	1	1	1	ZNF98, LOC440518, BC030765	
chr19	45400001	45600000	97	1	1	0.962477	TOMM40, APOE, APOC1, APOC1P1, APOC4-APOC2, CLPTM1, RELB, CLASRP, ZNF296, GEMIN7, PPP1R37	APOE: C-reactive protein, Lipid metabolism phenotypes, Triglycerides, Cardiovascular disease risk factors. APOC1: Carotid intima media thickness, Longevity, C-reactive protein. Alzheimer. LDL, HDL colesterol,
chr2	1400001	1600000	91	1	1	1	TPO	
chr2	204600001	204800000	100	0.897741	0.923972	0.949752	CD28, CTLA4	Rheumatoid arthritis, Alopecia areata, Type 1 diabetes autoantibodies, Rheumatoid arthritis,
chr20	50400001	50600000	71	0.975092	0.830137	1	SALL4	
chr6	33200001	33400000	83	0.975092	0.923622	1	HCG25, VPS52, RPS18, B3GALT4, WDR46, PFDN6, RGL2, TAPBP, ZBTB22, DAXX, KIFC1, BC146941, PHF1, CUTA, SYNGAP1	
chr6	154400001	154600000	142	1	1	1	OPRM1, IPCEF1	
chr1	100400001	100600000	25	0.910913	1	0.987599	BC112312, SLC35A3, HIAT1, SASS6, TRMT13	
chr1	171200001	171400000	76	1	1	0.962477	FMO1, Y_RNA, FMO4, TOP1P1	
chr14	86600000	86800000	47	1	1	0.962477		
Regiones candidatas para el test Fst: 1% de las ventanas con máximo puntaje al comparar población americana con asiáticos, europeos y africanos.								

Tabla S4: ihs AYM

Ranking	chr	Inicio	Termino	NºSNPs	Prop.iHS	Genes	Fenotipos (GWAS p<10-8)	Exclusivo Amerindio
1	chr7	133200001	133400001	25	0.96	EXOC4		
2	chr9	10000001	10200001	34	0.794118	PTPRD	Type 2 diabetes, Restless legs syndrome	
3	chr12	114400001	114600001	24	0.75	RBM19		
4	chr6	31000001	31200001	255	0.745098	MUC22, HCG22, C6orf15, PSORS1C1, CDSN, PSORS1C2, CCHCR1, TCF19, POU5F1, POU5F1, POU5F1, PSORS1C3, HCG27	Hypothyroidism, HIV-1 control, Systemic sclerosis, Nevirapine-induced rash, Prostate cancer, Stevens-Johnson syndrome and toxic epidermal necrolysis (SJS-TEN), Coronary heart disease	
5	chr2	198600001	198800001	53	0.735849	BC021693, BOLL, PLCL1	Intracranial aneurysm, Intracranial aneurysm, Crohn disease,	
6	chr22	48000001	48200001	36	0.722222	LINC00898, AK093107		
7	chr10	65200001	65400001	30	0.7	JMJD1C, JMJD1C-AS1, REEP3	Triglycerides, Liver enzyme levels (alkaline phosphatase), Platelet aggregation, Sex hormone-binding globulin levels	
8	chr6	52200001	52400001	33	0.69697	PAQR8, EFHC1, TRAM2		
9	chr11	64800001	65000001	23	0.695652	ARL2-SNX15, SAC3D1, NAALADL1, AB231703, CDCA5, ZFPL1, LOC100130348, VPS51, TM7SF2, ZNHIT2, BC104003, FAU, MRPL49, SYVN1, SPDYC, CAPN1, SLC22A20		
10	chr11	78000001	78200001	68	0.676471	GAB2, NARS2	Alzheimer disease	
11	chr12	114600001	114800001	36	0.666667	TBX5	Electrocardiographic traits, Diastolic blood pressure	
12	chr17	7400001	7600001	29	0.655172	POLR2A, TNFSF12-TNFSF13, SENP3, SENP3-EIF4A1, SNORA48, SNORD10, SNORA67, CD68, MPDU1, SOX15, FXR2, SAT2, SHBG, ATP1B2, TP53, HV941431, HV941433, HV941428, HV941434, HV941486, HV941429, HV941440, HV941478, HV941442, HV941444, HV941430, WRAP53	IgA nephropathy, Testosterone levels, Sex hormone-binding globulin levels, Androgen levels	

13	chr16	28400001	28600001	20	0.65	EIF3C, NPIPL1, CLN3, APOBR, IL27, NUPR1, CCDC101	Crohn disease, type 1 diabetes, Inflammatory bowel disease (early onset)	
14	chr2	204000001	204200001	25	0.64	NBEAL1, CYP20A1, ABI2		
15	chr11	8400001	8600001	22	0.636364	STK33, SCARNA20		
16	chr7	150200001	150400001	27	0.62963	GIMAP7, GIMAP4, GIMAP6, GIMAP2		
17	chr12	48200001	48400001	50	0.62	HDAC7, VDR, TMEM106C, COL2A1		
18	chr6	52400001	52600001	38	0.605263	TRAM2, TRAM2-AS1, LOC730101, TMEM14A		
19	chr16	11200001	11400001	25	0.6	CLEC16A, SOCS1, TNP2, PRM3, PRM2, PRM1	Multiple sclerosis, type 1 diabetes, primary biliary cirrhosis, crohn disease, psoriasis, celiac disease	
20	chr11	65400001	65600001	32	0.59375	PCNXL3, MIR4690, SIPA1, MIR4489, DQ587981, RELA, KAT5, RNASEH2C, AP5B1, OVOL1	Atopic dermatitis	
21	chr11	60200001	60400001	22	0.590909	MS4A5, MS4A1, MS4A12, MS4A13, LINC00301		
22	chr16	10000001	10200001	31	0.580645	GRIN2A	Hepatitis B	
23	chr1	204400001	204600001	21	0.571429	PIK3C2B, MDM4, LRRN2, AK097184		
24	chr4	71200001	71400001	30	0.566667	CABS1, SMR3A, PROL1, MUC7, AMTN		
25	chr11	77800001	78000001	60	0.55	ALG8, LOC100289388, KCTD21, USP35, GAB2	Alzheimer disease	
26	chr1	72000001	72200001	26	0.538462	NEGR1	Body mass index	
27	chr19	41400001	41600001	30	0.533333	CYP2G1P, CYP2B7P1, CYP2B6, CYP2A13		
28	chr11	24400001	24600001	47	0.531915	LUZP2		
29	chr5	15600001	15800001	23	0.521739	FBXL7		
30	chr6	168000001	168200001	27	0.518519	AK127120, LOC441178, C6orf123		
31	chr18	46400001	46600001	51	0.509804	SMAD7, DYM, MIR4744	Height, Colorectal cancer	
32	chr11	26600001	26800001	53	0.509434	ANO3, SLC5A12		
33	chr1	205000001	205200001	20	0.5	CNTN2, TMEM81, RBBP5, DSTYK, TMCC2	Mean platelet volume	
34	chr3	68400001	68600001	32	0.5	FAM19A1, FRMD4B		
35	chr7	29800001	30000001	26	0.5	WIPF3, SCRN1		
36	chr9	10200001	10400001	45	0.488889	PTPRD	Type 2 diabetes, Restless legs syndrome	
37	chr3	10800001	11000001	33	0.484848	LINC00606, SLC6A11		
38	chr10	135000001	135200001	29	0.482759	KNDC1, UTF1, VENTX, MIR202, ADAM8, TUBGCP2, ZNF511, BC047942, CALY, FUOM, ECHS1, MIR3944, PAOX		
39	chr3	58200001	58400001	54	0.481481	ABHD6, RPP14, RPP14, PXK	Systemic lupus erythematosus, Rheumatoid arthritis	

40	chr2	242600001	242800001	25	0.48	ATG4B, DTYMK, AK126180, ING5, D2HGDH, GAL3ST2, PABL, NEU4, PDCD1		
41	chr10	44800001	45000001	48	0.479167	CXCL12	Coronary heart disease	
42	chr11	116600001	116800001	71	0.478873	BUD13, ZNF259, APOA5, APOA4, APOC3, APOA1, SIK3	Triglycerides, LDL cholesterol, Metabolic sundrome, waist circumference, HDL cholesterol. Coronary heart disease, Vitamin E levels, Phospholipid levels (plasma), Hypertriglyceridemia	
43	chr2	191600001	191800001	21	0.47619	GLS		
44	chr4	15400001	15600001	21	0.47619	C1QTNF7, CC2D2A, AX746699	Conduct disorder (symptom count),	
45	chr12	114200001	114400001	36	0.472222	AK096932, RBM19		
47	chr14	23400001	23600001	36	0.472222	HAUS4, AJUBA, C14orf93, PSMB5, PSMB11, CDH24, ACIN1, C14orf119, BC153822, CEBPE, SLC7A8		
48	chr14	96000001	96200001	51	0.470588	SNHG10, GLRX5, BC038791, TCL6, TCL1B, TCL1A, BX247990	Atrial fibrillation	
49	chr1	170600001	170800001	20	0.45	PRRX1	Renal cell carcinoma	
50	chr12	26600001	26800001	40	0.45	ITPR2		
51	chr14	75400001	75600001	20	0.45	PGF, EIF2B2, MLH3, ACYP1, ZC2HC1C, NEK9, TMED10	LDL cholesterol, Haptoglobin levels, Liver enzyme levels (alkaline phosphatase)	
52	chr16	72000001	72200001	20	0.45	PKD1L3, DHODH, HP, HPR, TXNL4B, DHX38, PMFBP1	Breast cancer, Crohn disease, Mammographic density	
53	chr10	64400001	64600001	27	0.444444	ZNF365, ADO, EGR2		
54	chr5	150600001	150800001	52	0.442308	CCDC69, GM2A, SLC36A3, SLC36A2	hair color	
55	chr11	68800001	69000001	43	0.44186	TPCN2, BC064339	Red blood cell traits, Hematology traits, Mean corpuscular hemoglobin, White blood cell types, Other erythrocyte phenotypes	
56	chr6	135200001	135400001	34	0.441176	ALDH8A1, HBS1L, MIR3662		
57	chr12	55000001	55200001	25	0.44	GLYCAM1, LACRT, DCD	HDL cholesterol, Ulcerative colitis, Primary biliary cirrhosis	
58	chr17	37800001	38000001	25	0.44	STARD3, TCAP, PNMT, PGAP3, ERBB2, MIR4728, MIEN1, GRB7, IKZF3	Nephropathy (idiopathic membranous)	
59	chr2	160600001	160800001	25	0.44	MARCH7, LY75-CD302, PLA2R1		
60	chr19	32800001	33000001	32	0.4375	ZNF507, LOC400684, DPY19L3		
61	chr4	178600001	178800001	39	0.435897	LOC285501		
62	chr5	113800001	114000001	23	0.434783	KCNN2, AK097686	Lentiform nucleus volume	

63	chr1	171000001	171200001	49	0.428571	MROH9, FMO3, MIR1295A, MIR1295B, FMO6P, FMO2		
64	chr12	109600001	109800001	21	0.428571	ACACB, FOXN4	Type 1 diabetes autoantibodies	
65	chr1	157600001	157800001	26	0.423077	FCRL3, FCRL2, FCRL1		
66	chr16	17400001	17600001	26	0.423077	XYLT1, AX747757	Body mass index	
67	chr15	68000001	68200001	45	0.422222	MAP2K5, SKOR1	Menarche (age at onset), Body mass index	
68	chr11	8600001	8800001	24	0.416667	STK33, TRIM66, BC068088, RPL27A, SNORA3, SNORA45, ST5		
69	chr22	46200001	46400001	22	0.409091	ATXN10, WNT7B	Esophageal cancer	
70	chr4	100000001	100200001	22	0.409091	ADH5, LOC100507053, LOC100507053, ADH4, PCNAP1, ADH6, BC038532, ADH1A		

Tabla S5: iHS PH

Ranking	chr	Inicio	Termino	Nº SNPs	Prop.iHS	Genes	Fenotipos (GWAS p<10-8)
1	1	159200001	159400001	42	0.904762	FCER1A, OR10J3, BC038194	IgE levels
2	1	152800001	153000001	30	0.766667	LCE1A, LCE6A, SMCP, IVL, SPRR4, SPRR1A, SPRR3	
3	6	52200001	52400001	27	0.740741	PAQR8, EFHC1, TRAM2	
4	2	42800001	43000001	22	0.727273	MTA3, OXER1, HAAO	
5	5	74600001	74800001	28	0.714286	HMGCR, COL4A3BP,	LDL cholesterol, Metabolite levels, Body mass index
6	2	48600001	48800001	23	0.695652	FOXN2, PPP1R21	
7	2	227800001	228000001	58	0.637931	RHBDD1, SNORA48, COL4A4	Immune reponse to smallpox (secreted IFN-alpha)
8	3	15200001	15400001	37	0.594595	COL6A4P1, CAPN7, SH3BP5	
9	4	102000001	102200001	21	0.571429	PPP3CA	Blood pressure
10	6	52400001	52600001	21	0.571429	TRAM2, LOC730101, TMEM14A	
11	2	69800001	70000001	20	0.5	AAK1, ANXA4	
12	16	89600001	89800001	31	0.483871	SPG7, DQ596025, SPG7, RPL13, SNORD68, CPNE7, DPEP1, CHMP1A, C16orf55, BC031657, CDK10, SPATA2L, VPS9D1, LOC100128881, ZNF276	
13	5	81000001	81200001	29	0.482759	SSBP2	
14	1	159400001	159600001	44	0.477273	BC038194, OR10J1, OR10J5, APCS	
15	19	20800001	21000001	21	0.47619	ZNF626, ZNF66	
16	5	90000001	90200001	34	0.470588	GPR98	Response to antipsychotic treatment
17	1	159600001	159800001	47	0.468085	CRP, DUSP23, FCRL6, SLAMF8	C-reactive protein, Metabolic traits, Inflammatory biomarkers
18	4	21400001	21600001	28	0.464286	KCNIP4	
19	6	32200001	32400001	385	0.45974	AK123889, C6orf10, HCG23, BTNL2	Coronary heart disease, Vitiligo, Asthma, Lung Adenocarcinoma, Osteoarthritis, Vitiligo, Ulcerative colitis
20	10	70000001	70200001	24	0.458333	PBLD, HNRNPH3, RUFY2, DNA2	Vertical cup-disc ratio, Optic disc parameters
21	8	95800001	96000001	24	0.458333	DPY19L4, INTS8, CCNE2, TP53INP1	Type 2 diabetes
22	20	5800001	6000001	68	0.455882	C20orf196, CHGB, TRMT6, MCM8, CRLS1	Menarche and menopause (age at onset)

23	2	36800001	37000001	66	0.454545	FEZ2, VIT	
24	12	108800001	109000001	20	0.45	FICD, SART3, ISCU, TMEM119	
25	13	36800001	37000001	20	0.45	CCDC169, SOHLH2, U6, SPG20, SPG20OS	
26	4	62800001	63000001	20	0.45	LPHN3, BC039452	
27	8	11000001	11200001	38	0.447368	XKR6, MTMR9, CR749668, CR749668, SLC35G5, TDH	Triglycerides
28	10	6800001	7000001	32	0.4375	LINC00707	
29	11	64800001	65000001	23	0.434783	ARL2-SNX15, SAC3D1, NAALADL1, AB231703, CDCA5, ZFPL1, LOC100130348, VPS51, TM7SF2, ZNHIT2, BC104003, FAU, MRPL49, SYVN1, SPDYC, CAPN1, SLC22A20, SLC22A20	
30	19	19200001	19400001	33	0.424242	SLC25A42, TMEM161A, MEF2B, MEF2BNB, RFXANK, NR2C2AP, NCAN, NCAN, HAPLN4, TM6SF2, SUGP1	Bipolar disorder, LDL cholesterol, triglycerides
31	1	153000001	153200001	26	0.423077	SPRR1B, SPRR2D, SPRR2A, SPRR2B, SPRR2E, SPRR2F, SPRR2C, SPRR2G, LELP1, PRR9, Mir_584	
32	1	170200001	170400001	26	0.423077	LOC284688	
33	4	109000001	109200001	24	0.416667	LEF1, LEF1-AS1	
34	6	46600001	46800001	24	0.416667	CYP39A1, SLC25A27, TDRD6, PLA2G7, ANKRD66, MEP1A	Lipoprotein-associated phospholipase A2 activity and mass
35	11	99400001	99600001	29	0.413793	CNTN5	Immune reponse to smallpox (secreted IL-2)
36	12	27000001	27200001	22	0.409091	ASUN, FGFR1OP2, TM7SF3, MED21	
37	13	44400001	44600001	22	0.409091	CCDC122, LACC1, LINC00284	Vertical cup-disc ratio
38	5	161400001	161600001	22	0.409091	GABRG2	
39	1	153200001	153400001	25	0.4	LOR, PGLYRP3, PGLYRP4, S100A9, S100A12, S100A8, S100A7A	
40	5	161200001	161400001	20	0.4	GABRA1	
41	20	52600001	52800001	56	0.392857	BCAS1, MIR4756, CYP24A1	Multiple sclerosis
42	6	31600001	31800001	46	0.391304	PRRC2A, BAG6, APOM, C6orf47, GPANK1, CSNK2B, LY6G5C, ABHD16A, ABHD16A, MIR4646, LY6G6F, LY6G6E, LY6G6C, C6orf25, DDAH2, CLIC1, MSH5-SAPCD1, MSH5-SAPCD1, VWA7, VARS, LSM2, HSPA1L, HSPA1A, HSPA1B	Menopause (age at onset), Lung adenocarcinoma, Rheumatoid arthritis, Multiple sclerosis, asthma
43	6	13200001	13400001	54	0.388889	PHACTR1, LOC100130357, TBC1D7, AK127555, GFOD1	Coronary heart disease, Migraine, Myocardial infarction (early onset)
44	6	163400001	163600001	31	0.387097	PACRG, AK058177, AK296276	

45	3	183800001	184000001	26	0.384615	HTR3E, EIF2B5, DVL3, AP2M1, ABCF3, VWA5B2, MIR1224, ALG3, ECE2, CAMK2N2	
46	13	36600001	36800001	47	0.382979	DCLK1, CCDC169, SOHLH2	
47	7	132000001	132200001	34	0.382353	PLXNA4	
48	1	190400001	190600001	21	0.380952	FAM5C, CR936711, LOC440704	
49	5	22800001	23000001	21	0.380952	CDH12	
50	8	6200001	6400001	71	0.380282	LOC100287015, MCPH1, ANGPT2	Epirubicin-induced leukopenia
51	11	107400001	107600001	37	0.378378	ALKBH8, ELMOD1, LOC643923, SLN	
52	15	53800001	54000001	40	0.375	WDR72, U2	Cognitive function, Renal function-related traits (BUN), Chronic kidney disease
53	2	166600001	166800001	24	0.375	GALNT3, LOC100506124, TTC21B, LOC100506134	Bone mineral density
54	6	29000001	29200001	137	0.372263	LOC100129636, OR2W1, OR2B3, OR2J3, OR2J2	Height
55	11	126000001	126200001	27	0.37037	RPUSD4, FAM118B, U4, SRPR, FOXRED1, TIRAP, DCPS	
56	5	160000001	160200001	27	0.37037	ATP10B, Mir_544	
57	14	64600001	64800001	30	0.366667	SYNE2, ESR2	Atrial fibrillation
58	10	119200001	119400001	41	0.365854	EMX2OS, EMX2	
59	10	25600001	25800001	33	0.363636	GPR158	Immune response to smallpox vaccine (IL-6)
60	12	2200001	2400001	33	0.363636	CACNA1C, CACNA1C-AS4, CACNA1C-IT3	Bipolar disorder and major depressive disorder (combined)
61	15	42000001	42200001	22	0.363636	MGA, MAPKBP1, BC045779, JMJD7-PLA2G4B, SPTBN5, 5S_rRNA, MIR4310, EHD4	

Tabla S6. XPEHH y Fst entre población Aymara y población Pehuenche y Huilliche

chr	Inicio	Termino	NºSNPs	Fst	XPEHH	Genes	Fenotipos (GWAS p<10-8)
15	45000001	45200001	41	0.661961	3.84903	PATL2, B2M, TRIM69, U1	
15	45200001	45400001	22	0.559531	3.78465	C15orf43, SORD, DUOX2	
2	99200001	99400001	40	0.445383	3.89569	INPP4A, COA5, UNC50, MGAT4A	
2	99400001	99600001	28	0.448685	3.97074	KIAA1211L	
4	21200001	21400001	58	0.484912	4.3686	KCNIP4	
6	144000001	144200001	58	0.450026	3.68579	PHACTR2, LTV1, AX746989, ZC2HC1B	
6	31000001	31200001	373	0.438183	3.92351	MUC22, HCG22, C6orf15, PSORS1C1, CDSN, PSORS1C2, CCHCR1, TCF19, POU5F1, POU5F1, POU5F1, PSORS1C3, HCG27	Hipotiroidismo, VIH 1, sclerosis, psoriasis, enfermedades coronarias,
6	31200001	31400001	910	0.45604	5.02709	HLA-C, MICA	Carcinoma nasofaringeo, VIH-1, psoriasis, progresion de SIDA, carcinoma hepatocellular, artritis reumatoide, peso.
8	3400001	3600001	180	0.536517	4.61354	CSMD1	Esquizofrenia.
9	112600001	112800001	61	0.483872	4.01931	PALM2-AKAP2	

Tabla S7: Resultados de PANTHER, iHS Nativos americanos

PANTHER GO- proceso biológico	Genes en referencia	Lista genes candidatos	Nº genes esperado	Sobre o sub-representación	Proporción sobre o sub-representación	Valor de p	Valores de p significativos
cytokine production	3	2	0.25	+	> 5	2.59E-02	
protein lipidation	17	5	1.4	+	3.57	1.42E-02	
complement activation	31	8	2.55	+	3.13	4.75E-03	
natural killer cell activation	**62**	**15**	**5.11**	**+**	**2.94**	**2.69E-04**	*
antigen processing and presentation	42	9	3.46	+	2.6	9.09E-03	
antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	28	6	2.31	+	2.6	3.02E-02	
amino acid transport	36	7	2.97	+	2.36	3.16E-02	
carbohydrate transport	58	11	4.78	+	2.3	9.92E-03	
cell proliferation	57	10	4.7	+	2.13	2.18E-02	
macrophage activation	110	19	9.06	+	2.1	2.48E-03	
DNA replication	70	12	5.77	+	2.08	1.51E-02	
anion transport	118	19	9.72	+	1.95	5.18E-03	
hemopoiesis	69	11	5.69	+	1.93	3.05E-02	
carbohydrate metabolic process	**322**	**50**	**26.54**	**+**	**1.88**	**2.27E-05**	**
fatty acid metabolic process	124	18	10.22	+	1.76	1.67E-02	
steroid metabolic process	121	17	9.97	+	1.7	2.57E-02	
generation of precursor metabolites and energy	139	19	11.45	+	1.66	2.45E-02	
blood circulation	111	15	9.15	+	1.64	4.56E-02	
blood coagulation	120	16	9.89	+	1.62	4.41E-02	
immune response	341	42	28.1	+	1.49	7.55E-03	
cation transport	393	44	32.39	+	1.36	2.75E-02	
ion transport	488	54	40.22	+	1.34	1.93E-02	
biosynthetic process	417	46	34.36	+	1.34	3.05E-02	
immune system process	879	89	72.44	+	1.23	2.72E-02	
primary metabolic process	3795	346	312.74	+	1.11	1.13E-02	

regulation of catalytic activity	648	41	53.4	-	0.77	4.25E-02	
protein transport	617	39	50.85	-	0.77	4.66E-02	
intracellular protein transport	605	38	49.86	-	0.76	4.48E-02	
mesoderm development	464	28	38.24	-	0.73	4.88E-02	
vesicle-mediated transport	513	26	42.28	-	0.62	4.19E-03	
mitosis	192	9	15.82	-	0.57	4.59E-02	
muscle organ development	208	9	17.14	-	0.53	2.33E-02	
visual perception	150	6	12.36	-	0.49	3.65E-02	
skeletal system development	161	6	13.27	-	0.45	2.14E-02	
sensory perception	**292**	**8**	**24.06**	**-**	**0.33**	**1.24E-04**	**
pattern specification process	127	3	10.47	-	0.29	7.10E-03	
female gamete generation	66	1	5.44	-	< 0.2	2.77E-02	
segment specification	96	1	7.91	-	< 0.2	3.18E-03	

*valor de $p < 0.05$. **valores de p significativos luego de aplicar corrección de Bonferroni para testeos múltiples

Tabla S8: Resultados de PANTHER, iHS Aymaras

PANTHER GO- proceso biológico	Genes en referencia	Lista genes candidatos	Nº genes esperado	Sobre o sub-representación	Proporción sobre o sub-representación	Valor de p	Valores de p significativos
cytokine production	3	2	0.24	+	> 5	2.48E-02	
response to pheromone	4	2	0.32	+	> 5	4.19E-02	
protein lipidation	16	4	1.29	+	3.11	4.17E-02	
natural killer cell activation	65	13	5.23	+	2.49	2.85E-03	
complement activation	32	6	2.57	+	2.33	4.69E-02	
fatty acid biosynthetic process	32	6	2.57	+	2.33	4.69E-02	
anion transport	**123**	**23**	**9.9**	**+**	**2.32**	**2.33E-04**	*
cell proliferation	60	11	4.83	+	2.28	1.06E-02	
fatty acid metabolic process	130	19	10.46	+	1.82	1.07E-02	
respiratory electron transport chain	112	16	9.01	+	1.78	2.17E-02	
ion transport	**499**	**65**	**40.15**	**+**	**1.62**	**1.31E-04**	**
generation of precursor metabolites and energy	142	18	11.43	+	1.58	4.24E-02	
cation transport	401	49	32.26	+	1.52	3.06E-03	
response to stress	368	40	29.61	+	1.35	3.68E-02	
biosynthetic process	424	46	34.11	+	1.35	2.74E-02	
nitrogen compound metabolic process	607	61	48.84	+	1.25	4.66E-02	
neurological system process	678	41	54.55	-	0.75	3.02E-02	
protein transport	629	33	50.61	-	0.65	4.56E-03	
intracellular protein transport	617	32	49.64	-	0.64	4.13E-03	
endocytosis	237	11	19.07	-	0.58	3.22E-02	
cell-cell adhesion	256	11	20.6	-	0.53	1.50E-02	
exocytosis	147	6	11.83	-	0.51	4.93E-02	
visual perception	150	6	12.07	-	0.5	4.31E-02	
vesicle-mediated transport	**526**	**21**	**42.32**	**-**	**0.5**	**1.70E-04**	**
heart development	126	4	10.14	-	0.39	2.61E-02	
sensory perception	**288**	**9**	**23.17**	**-**	**0.39**	**6.41E-04**	*
*valor de p < 0.05. **valores de p significativos luego de aplicar corrección de Bonferroni para testeos múltiples							

Tabla S9: Resultados de PANTHER, iHS Pehuenches y Huilliches

PANTHER GO- proceso biológico	Genes en referencia	Lista genes candidatos	Nº genes esperado	Sobre o sub-representación	Proporción sobre o sub-representación	Valor de p	Valores de p significativos
phagocytosis	13	4	0.99	+	4.03	1.85E-02	
antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	27	7	2.06	+	3.39	5.26E-03	
response to toxic substance	31	7	2.37	+	2.96	1.07E-02	
antigen processing and presentation	37	8	2.83	+	2.83	8.44E-03	
RNA catabolic process	35	7	2.67	+	2.62	1.94E-02	
DNA replication	65	12	4.96	+	2.42	5.02E-03	
cell-cell adhesion	**250**	**41**	**19.1**	**+**	**2.15**	**6.65E-06**	**
respiratory electron transport chain	100	16	7.64	+	2.09	5.22E-03	
generation of precursor metabolites and energy	126	18	9.62	+	1.87	9.64E-03	
cellular defense response	152	21	11.61	+	1.81	7.83E-03	
biological adhesion	**383**	**50**	**29.25**	**+**	**1.71**	**2.26E-04**	**
cell adhesion	**361**	**47**	**27.57**	**+**	**1.7**	**3.66E-04**	
translation	206	25	15.73	+	1.59	1.77E-02	
ectoderm development	414	44	31.62	+	1.39	1.92E-02	
immune system process	842	82	64.31	+	1.28	1.49E-02	
biological regulation	2160	139	164.98	-	0.84	1.17E-02	
regulation of nucleobase-containing compound metabolic process	901	55	68.82	-	0.8	4.31E-02	
regulation of biological process	1540	93	117.63	-	0.79	6.68E-03	
regulation of transcription from RNA polymerase II promoter	711	42	54.31	-	0.77	4.45E-02	
cellular component organization or biogenesis	699	37	53.39	-	0.69	9.40E-03	
cellular component organization	646	33	49.34	-	0.67	7.31E-03	
organelle organization	260	12	19.86	-	0.6	3.97E-02	
catabolic process	223	10	17.03	-	0.59	4.67E-02	
chromatin organization	105	3	8.02	-	0.37	4.11E-02	

regulation of phosphate metabolic process	109	3	8.33	-	0.36	3.33E-02	
segment specification	90	2	6.87	-	0.29	3.21E-02	
*valor de p < 0.05. **valores de p significativos luego de aplicar corrección de Bonferroni para testeos múltiples							

Tabla S10: Resultados de PANTHER, XPEHH entre Amerindios y Asiáticos

PANTHER GO- proceso biológico	Genes en referencia	Lista genes candidatos	Nº genes esperado	Sobre o sub-representación	Proporción sobre o sub-representación	Valor de p	Valores de p significativos
neuromuscular synaptic transmission	6	3	0.41	+	> 5	8.61E-03	
fatty acid biosynthetic process	40	7	2.75	+	2.54	2.23E-02	
vitamin transport	52	9	3.58	+	2.52	1.11E-02	
amino acid transport	53	9	3.65	+	2.47	1.24E-02	
carbohydrate transport	73	12	5.02	+	2.39	5.52E-03	
lipid transport	**295**	**43**	**20.29**	**+**	**2.12**	**6.11E-06**	*
cholesterol metabolic process	76	11	5.23	+	2.1	1.81E-02	
cytoskeleton organization	73	10	5.02	+	1.99	3.23E-02	
phosphate ion transport	95	13	6.53	+	1.99	1.64E-02	
anion transport	161	22	11.07	+	1.99	2.32E-03	
cation transport	**534**	**71**	**36.73**	**+**	**1.93**	**2.06E-07**	**
steroid metabolic process	166	22	11.42	+	1.93	3.31E-03	
extracellular transport	109	14	7.5	+	1.87	2.11E-02	
cell-cell signaling	**565**	**71**	**38.86**	**+**	**1.83**	**1.51E-06**	**
fatty acid metabolic process	184	23	12.65	+	1.82	5.39E-03	
synaptic transmission	**294**	**36**	**20.22**	**+**	**1.78**	**8.69E-04**	
ion transport	**666**	**80**	**45.8**	**+**	**1.75**	**1.79E-06**	**
response to external stimulus	342	40	23.52	+	1.7	1.08E-03	
blood coagulation	154	18	10.59	+	1.7	2.30E-02	
angiogenesis	182	20	12.52	+	1.6	3.02E-02	
lipid metabolic process	**804**	**86**	**55.29**	**+**	**1.56**	**5.16E-05**	*
ectoderm development	582	55	40.03	+	1.37	1.28E-02	
transport	**2248**	**210**	**154.6**	**+**	**1.36**	**2.86E-06**	**
localization	**2367**	**220**	**162.79**	**+**	**1.35**	**2.21E-06**	**
proteolysis	654	60	44.98	+	1.33	1.66E-02	
nervous system development	724	66	49.79	+	1.33	1.41E-02	
cell adhesion	502	45	34.52	+	1.3	4.68E-02	

response to stress	573	51	39.41	+	1.29	4.02E-02	
regulation of molecular function	1000	87	68.77	+	1.27	1.62E-02	
regulation of catalytic activity	980	84	67.4	+	1.25	2.45E-02	
cellular protein modification process	1189	101	81.77	+	1.24	1.82E-02	
protein metabolic process	2433	203	167.33	+	1.21	2.09E-03	
immune system process	1234	101	84.87	+	1.19	4.18E-02	
response to stimulus	1906	155	131.08	+	1.18	1.67E-02	
cell communication	2665	213	183.28	+	1.16	1.05E-02	
developmental process	2185	171	150.27	+	1.14	4.06E-02	
primary metabolic process	6086	468	418.56	+	1.12	1.59E-03	
cellular process	5957	444	409.69	+	1.08	2.00E-02	
metabolic process	7344	543	505.08	+	1.08	1.43E-02	
Unclassified	**6557**	**395**	**450.95**	**-**	**0.88**	**4.18E-04**	*****
macrophage activation	160	5	11	-	0.45	3.68E-02	
female gamete generation	92	2	6.33	-	0.32	4.85E-02	
tRNA metabolic process	76	1	5.23	-	< 0.2	3.32E-02	
*valor de p < 0.05. **valores de p significativos luego de aplicar corrección de Bonferroni para testeos múltiples							

Tabla S11: Resultados de PANTHER, Fst entre Amerindios y Asiáticos

PANTHER GO- proceso biológico	Genes en referencia	Lista genes candidatos	Nº genes esperado	Sobre o sub-representación	Proporción sobre o sub-representación	Valor de p	Valores de p significativos
regulation of sequence-specific DNA binding transcription factor activity	13	2	0.24	+	> 5	2.49E-02	
carbohydrate transport	**73**	**8**	**1.36**	**+**	**> 5**	**8.24E-05**	*
vitamin transport	55	5	1.02	+	4.88	3.97E-03	
extracellular transport	**110**	**10**	**2.05**	**+**	**4.88**	**5.20E-05**	*
phosphate ion transport	95	7	1.77	+	3.96	2.26E-03	
lipid transport	**299**	**20**	**5.57**	**+**	**3.59**	**1.29E-06**	**
cellular defense response	214	11	3.99	+	2.76	2.58E-03	
carbohydrate metabolic process	521	21	9.7	+	2.16	9.16E-04	
response to external stimulus	349	14	6.5	+	2.15	6.54E-03	
synaptic transmission	307	12	5.72	+	2.1	1.37E-02	
cell-cell signaling	581	21	10.82	+	1.94	3.33E-03	
proteolysis	674	24	12.55	+	1.91	2.12E-03	
immune response	472	16	8.79	+	1.82	1.69E-02	
immune system process	1279	43	23.82	+	1.81	1.40E-04	
regulation of molecular function	1020	34	19	+	1.79	8.39E-04	
regulation of catalytic activity	1000	33	18.62	+	1.77	1.18E-03	
lipid metabolic process	821	27	15.29	+	1.77	3.42E-03	
phosphate-containing compound metabolic process	853	28	15.89	+	1.76	2.97E-03	
response to stimulus	1991	49	37.08	+	1.32	2.67E-02	
primary metabolic process	6273	137	116.83	+	1.17	1.27E-02	
metabolic process	7571	163	141.01	+	1.16	8.88E-03	
*valor de p < 0.05. **valores de p significativos luego de aplicar corrección de Bonferroni para testeos múltiples							

Tabla S12: Resultados de PANTHER, XPEHH entre Aymaras y Pehuenches-Huilliches

PANTHER GO- proceso biológico	Genes en referencia	Lista genes candidatos	Nº genes esperado	Sobre o sub-representación	Proporción sobre o sub-representación	Valor de p	Valores de p significativos
antigen processing and presentation	**49**	**16**	**3.6**	**+**	**4.44**	**1.24E-06**	**
antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	33	10	2.42	+	4.13	2.14E-04	**
response to interferon-gamma	38	8	2.79	+	2.87	7.92E-03	
RNA catabolic process	53	10	3.89	+	2.57	6.72E-03	
induction of apoptosis	119	18	8.74	+	2.06	3.86E-03	
nucleobase-containing compound transport	107	15	7.86	+	1.91	1.47E-02	
hemopoiesis	99	13	7.27	+	1.79	3.45E-02	
cellular defense response	210	26	15.43	+	1.69	8.26E-03	
respiratory electron transport chain	183	22	13.44	+	1.64	1.91E-02	
angiogenesis	183	22	13.44	+	1.64	1.91E-02	
generation of precursor metabolites and energy	236	26	17.34	+	1.5	2.99E-02	
immune system process	1238	112	90.94	+	1.23	1.46E-02	
cellular component organization	1081	60	79.41	-	0.76	1.18E-02	
cellular component organization or biogenesis	1181	65	86.76	-	0.75	7.20E-03	
organelle organization	497	26	36.51	-	0.71	4.12E-02	
cellular component morphogenesis	438	22	32.18	-	0.68	3.64E-02	
pattern specification process	175	6	12.86	-	0.47	2.76E-02	

*valor de $p < 0.05$. **valores de p significativos luego de aplicar corrección de Bonferroni para testeos múltiples

Figuras suplementarias

Figura S1: Representación genómica de valores de Fst en todos los cromosomas

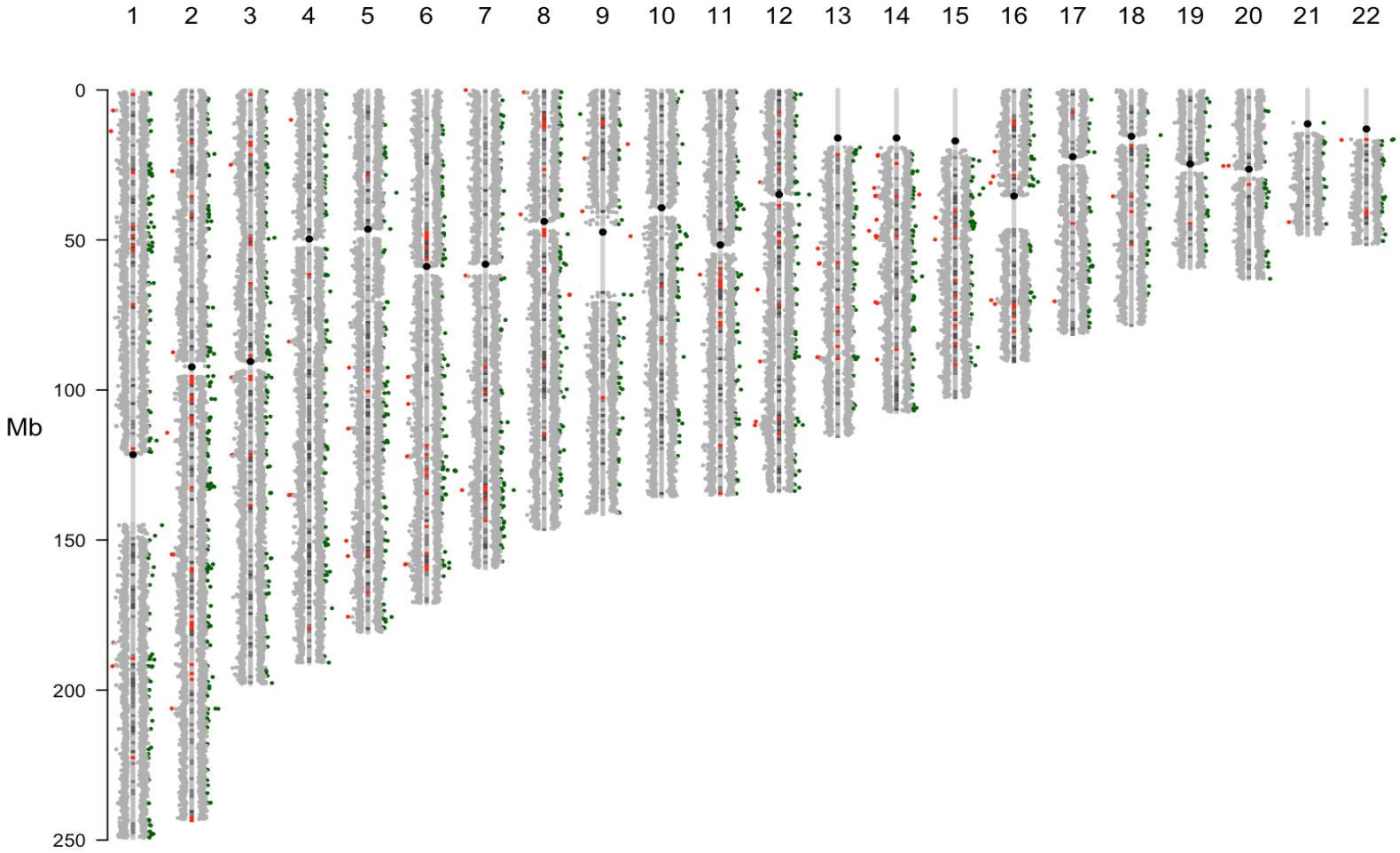

Figura S2: Distribución de valores máximos por ventana en test de Fst

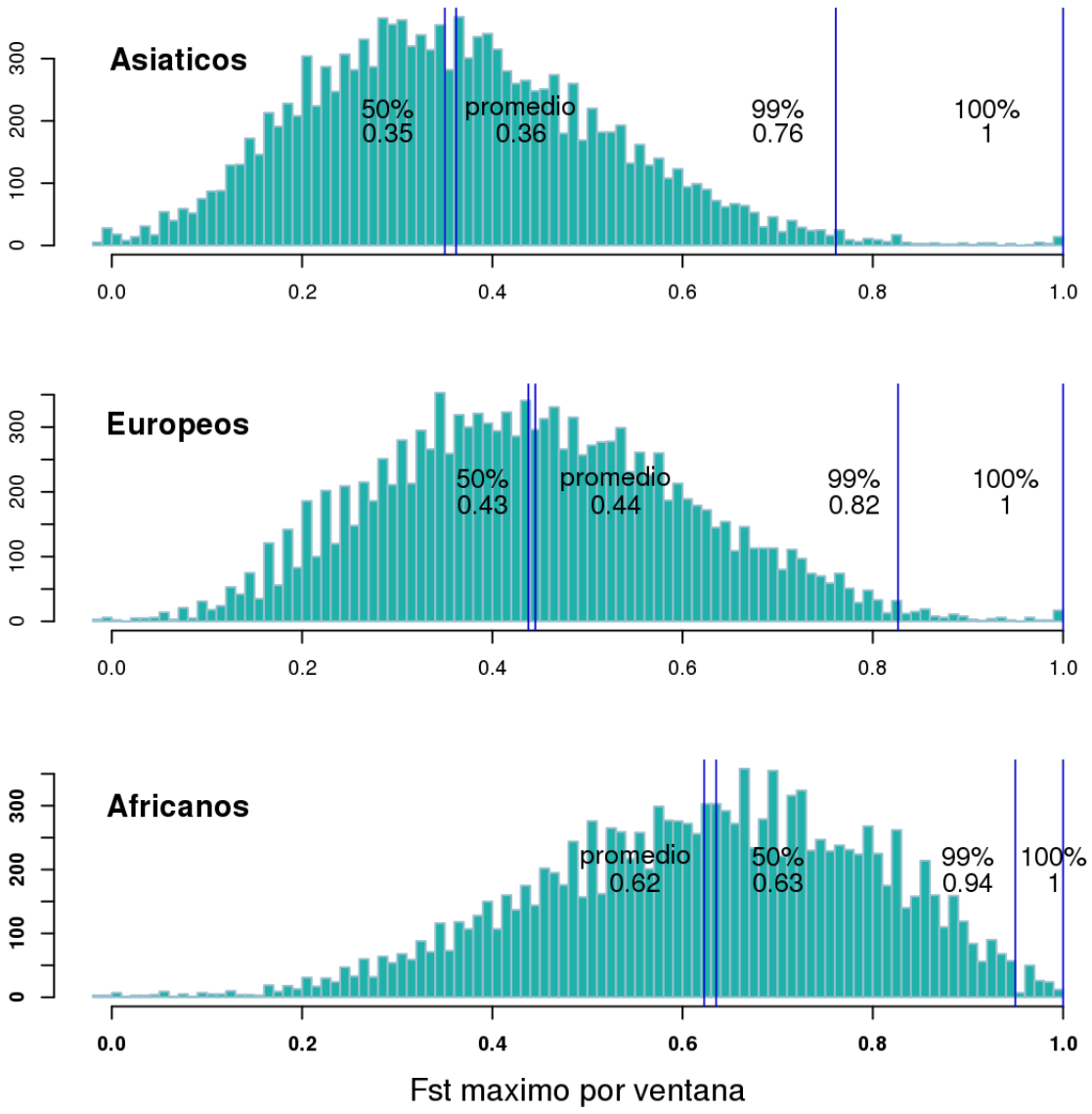

Figura S3: Distribución de valores de iHS por marcador en AMR, AYM y PH

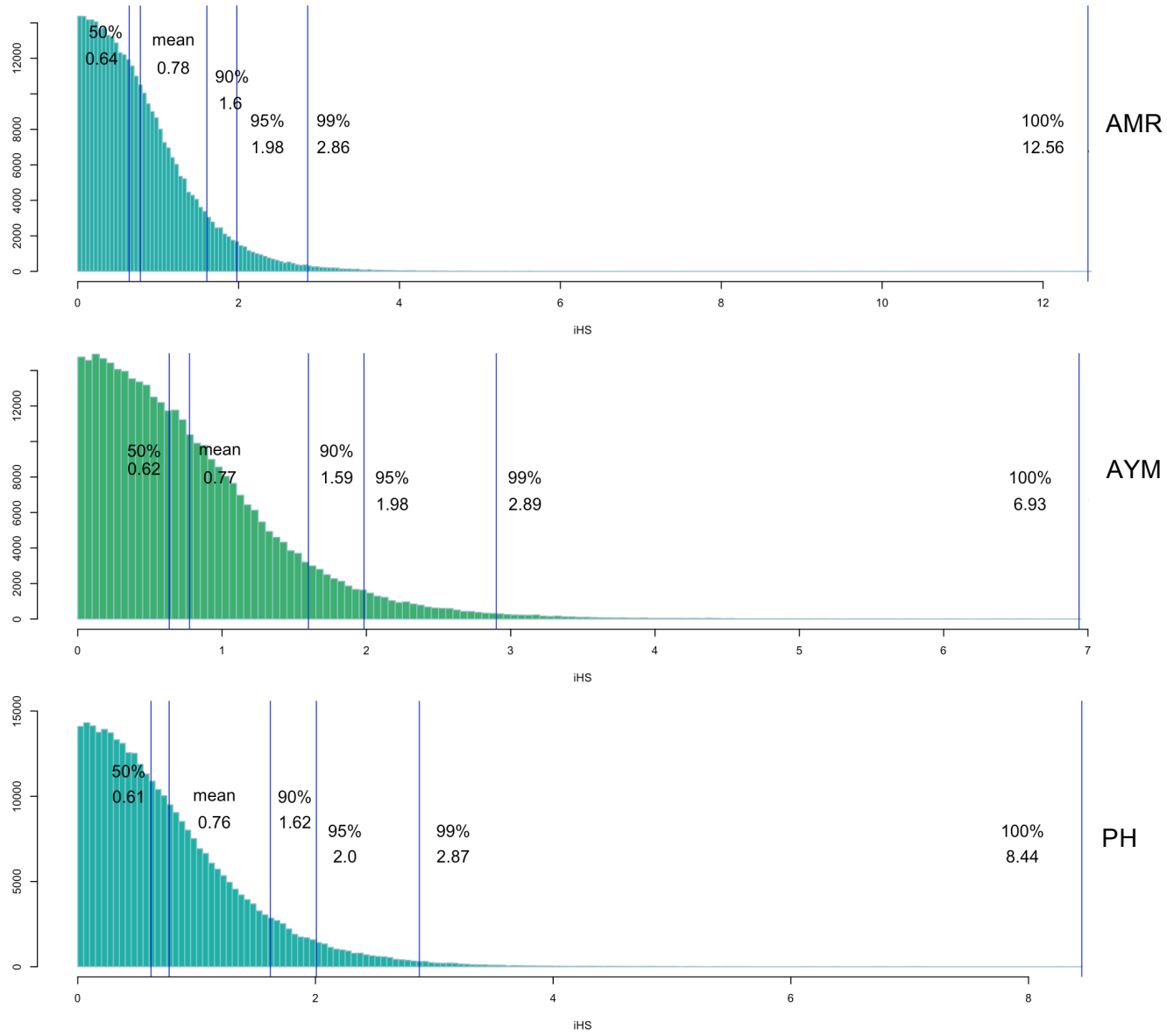

Figura S4: Distribución de porcentaje de valores altos de iHS por ventana de 200kb

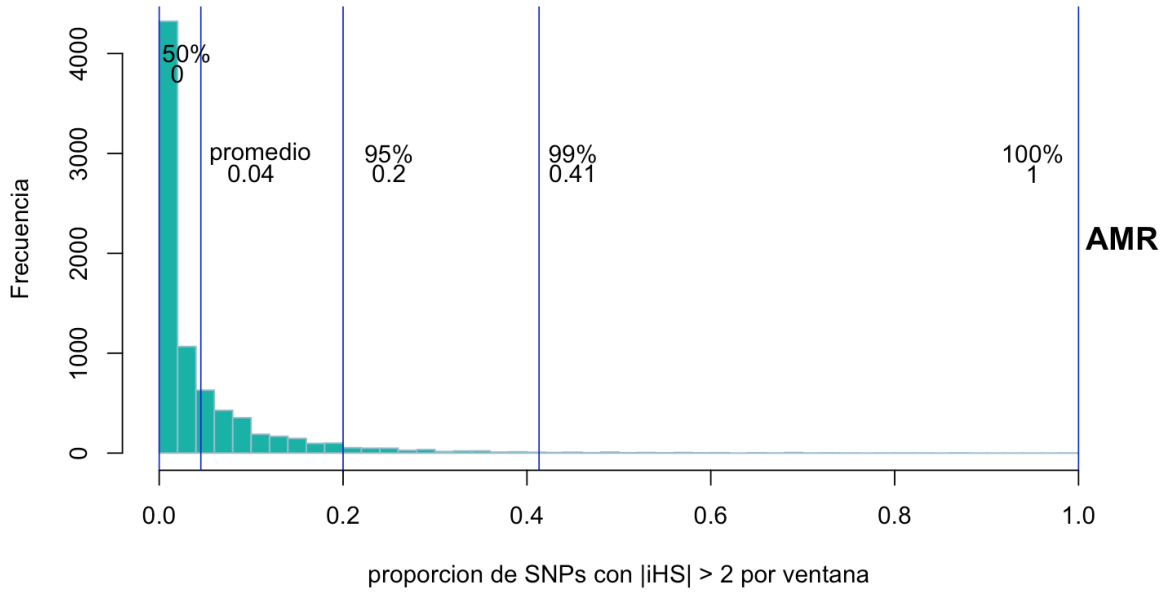

Figura S5: Distribución de valores por marcador en test XPEHH