



UNIVERSIDAD PERUANA
CAYETANO HEREDIA

IDENTIFICACIÓN DE HUELLAS DE
SELECCIÓN EN *Plasmodium falciparum*
DE LA AMAZONÍA PERUANA

TESIS PARA OPTAR EL GRADO DE
MAESTRO EN BIOQUÍMICA Y BIOLOGÍA
MOLECULAR

ERICK GIANPIERRE FIGUEROA
ILDEFONSO

LIMA – PERÚ

2023

ASESORES:

Dionicia Gamboa, Msc, PhD.

Hugo Valdivia Rodríguez, PhD.

JURADO DE TESIS

DR. DANIEL CLARK LEZA

PRESIDENTE

MG. TERESA VICTORIA BARRETO GAVIRIA

VOCAL

DR. JORGE ENRIQUE RODRIGUEZ BAILON

SECRETARIO

DEDICATORIA

A mi familia, quienes han estado conmigo durante mi desarrollo profesional. En especial a mi madre Elva, gracias a su esfuerzo incondicional pude iniciarme en la carrera de ciencia y me sigue apoyando hasta ahora.

A mis sobrinos Alexandra y Olek, a quienes quisiera compartir mis experiencias y enseñarles que todo esfuerzo conlleva a una recompensa.

A mi compañera de aventuras Lesly. Quien me ha brindado su apoyo, paciencia y honestidad, ayudándome a mantenerme firme en mis convicciones y a luchar por mis objetivos.

A mi tía Esther, quien a fines del 2022 partió, dejando un vacío importante en nuestra familia.

AGRADECIMIENTOS

Mis sinceros agradecimientos a mis asesores Dionicia Gamboa y Hugo Valdivia, por el apoyo, guía y dedicación brindados durante el proceso para lograr satisfactoriamente el grado de maestro. Un agradecimiento especial a la Dra. Gamboa por la confianza al encomendarme el reto de llevar a cabo esta tesis, y por su compromiso al darme el soporte necesario para la ejecución de la misma.

Agradezco también al equipo del laboratorio de malaria, por su amabilidad al acogerme y la ayuda brindada cuando la he necesitado. En especial a Berónica Infante, por todo el apoyo desde que me integré al laboratorio y durante el proceso de la tesis. A Luis y Roberson por la ayuda y sus aportes durante las discusiones del grupo de epidemiología molecular.

Quiero agradecer también al equipo de la unidad de malaria del Instituto de Medicina Tropical de Amberes (ITM), en especial a Anna Rosanas y Eline Kattenberg. El tiempo que pasé con su grupo fue una gran experiencia, además de ser de gran ayuda y de haber tenido un impacto importante en mi trabajo y en mí como profesional.

FUENTES DE FINANCIAMIENTO

Este proyecto contó con las siguientes fuentes de financiamiento:

- Proyecto “Improved infectious diseases control in Peru through sustainable capacity building for bioinformatics and genome sequencing”- VLIR UOS JOINT PE2019JOI018A102, el cual financió los costos en la adquisición de reactivos y material consumible
- “Proyecto colaborativo FA4 DGD-ITM 2017-2021”, el cual financió la adquisición de reactivos y entrenamiento en protocolos de laboratorio para secuenciación de nueva generación y análisis de datos en el Instituto de Medicina Tropical de Amberes, Bélgica.
- NAMRU-6, que proporcionó reactivos para los procedimientos de secuenciación.
- Proyecto “Transmission dynamics of residual malaria in the amazon: defining A roadmap to malaria elimination” International Centers of Excellence for Malaria Research (ICEMR) program (U19AI089681)”, NIH–USA, que brindó financiamiento de mi estipendio como asistente de investigación.
- Fundación Cobián, la cual financió las pensiones del programa de Maestría en Bioquímica y Biología molecular.

IDENTIFICACIÓN DE HUELLAS DE SELECCIÓN EN *Plasmodium falciparum* DE LA AMAZONÍA PERUANA

INFORME DE ORIGINALIDAD

13%	12%	5%	2%
INDICE DE SIMILITUD	FUENTES DE INTERNET	PUBLICACIONES	TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	repositorio.upch.edu.pe Fuente de Internet	5%
2	Submitted to Universidad Peruana Cayetano Heredia Trabajo del estudiante	1%
3	hdl.handle.net Fuente de Internet	1%
4	bdigital.unal.edu.co Fuente de Internet	<1%
5	www.paho.org Fuente de Internet	<1%
6	www.biorxiv.org Fuente de Internet	<1%
7	repositorio.unapiquitos.edu.pe Fuente de Internet	<1%
8	Castellano Esteve, David, Universitat Autònoma de Barcelona. Departament de Genètica i de Microbiologia. "Estimación de la	<1%

TABLA DE CONTENIDOS

LISTADO DE TABLAS

LISTADO DE FIGURAS

GLOSARIO

LISTADO DE ABREVIATURAS

RESUMEN

ABSTRACT

I. INTRODUCCIÓN	1
II. ANTECEDENTES	4
2.1. EPIDEMIOLOGÍA DE LA MALARIA	4
2.2. ETIOLOGÍA DE LA MALARIA	5
2.2.1. Ciclo de vida del parásito	5
2.3. DIAGNÓSTICO DE MALARIA	7
2.3.1. Microscopía.....	7
2.3.2. Diagnóstico basado en tecnologías moleculares	8
2.3.3. Pruebas de diagnóstico rápido	8
2.4. EVOLUCIÓN POR SELECCIÓN NATURAL.....	9
2.4.1. Modos de Selección	9
2.4.2. Hitchhiking y Barridos selectivos (Selective Sweeps)	10
2.4.3. Métodos para detección de huellas de selección	11

2.5.	DETECCIÓN DE HUELLAS DE SELECCIÓN EN <i>P. falciparum</i>	14
2.6.	DELECIÓN DE LOS GENES <i>pfhrp2</i> y <i>pfhrp3</i> COMO UN CASO POSIBLE DE SELECCIÓN NATURAL EN <i>P. falciparum</i>	16
2.7.	FUNCIÓN DE LA PROTEÍNA HRP2.....	17
2.8.	<i>P. falciparum</i> CON LA DELECIÓN DE LOS GENES <i>pfhrp2</i> Y <i>pfhrp3</i>	18
2.8.1.	Distribución de los parásitos con las deleciones de los genes <i>pfhrp2</i> y <i>pfhrp3</i>	18
2.8.2.	Diferencias a nivel fenotípico en parásitos con la deleción de <i>pfhrp2</i> y <i>pfhrp3</i>	22
III.	PLANTEAMIENTO DE LA INVESTIGACIÓN	24
IV.	JUSTIFICACIÓN	27
V.	OBJETIVOS.....	29
5.1.	OBJETIVO GENERAL.....	29
5.2.	OBJETIVOS ESPECÍFICOS	29
VI.	METODOLOGÍA	30
6.1.	DISEÑO DE INVESTIGACIÓN	30
6.2.	POBLACIÓN DE ESTUDIO	33
6.2.1.	Área de estudio	33
6.2.2.	Población de estudio.....	33
6.2.3.	Muestras Biológicas	34

6.2.4.	Operacionalización de Variables	35
6.3.	PROCEDIMIENTOS Y TÉCNICAS	37
6.3.1.	Caracterización de las deleciones de los genes <i>pfhrp2</i> , <i>pfhrp3</i> y genes flanqueantes PF3D7_0831700, PF3D7_0831900, PF3D7_1372100 y PF3D7_1372400.....	37
6.3.2.	Secuenciación de genoma completo (WGS) de <i>Plasmodium falciparum</i>	39
6.4.	CONSIDERACIONES ÉTICAS	44
6.5.	PLAN DE ANÁLISIS.....	45
6.5.1.	Primera etapa: Estimación de la frecuencia de las deleciones ..	45
6.5.2.	Segunda etapa: Identificación de barridos selectivos a partir de datos de genoma completo	46
VII.	RESULTADOS.....	54
7.1.	Caracterización de las muestras <i>P. falciparum</i> para la deleción de los genes <i>pfhrp2</i> , <i>pfhrp3</i> , y genes flanqueantes PF3D7_0831700, PF3D7_0831900, PF3D7_1372100 y PF3D7_1372400.....	54
7.2.	Amplificación Selectiva de Genoma Completo (sWGA) y Secuenciación de Genoma completo	58
7.3.	Análisis de datos de genoma completo para la caracterización de la población de estudio	60
VIII.	DISCUSIÓN.....	77
IX.	CONCLUSIONES.....	90

X. LIMITACIONES	91
XI. RECOMENDACIONES	92
XII. REFERENCIAS BIBLIOGRÁFICAS	94
XIII. ANEXOS	

LISTADO DE TABLAS

Tabla 1. Perfiles de delección de <i>pfhrp2</i> , <i>pfhrp3</i> y genes flanqueantes de <i>P. falciparum</i> en 227 muestras	19
Tabla 2. Tabla resumen de corridas de Secuenciación	43
Tabla 3. Lugar de procedencia de las muestras analizadas para perfil de la delección de los genes <i>pfhrp2</i> y <i>pfhrp3</i> y flanqueantes.....	55
Tabla 4. Tabla resumen de los distritos de origen de los genomas incluidos en el análisis.....	61
Tabla 5. Periodo de colecta y perfil de la delección de <i>pfhrp2</i>	65

LISTADO DE FIGURAS

Figura 1. Número de casos de malaria entre 2019 – 2022 en la región Loreto.	5
Figura 2. Ciclo de vida de <i>Plasmodium</i>	7
Figura 3. Ubicación genómica del gen <i>pfhrp2</i> en el cromosoma 8 y del gen <i>pfhrp3</i> en el cromosoma 13	17
Figura 4. Flujograma de trabajo.	32
Figura 5. Mapa de procedencia de las muestras incluidas en el estudio.	34
Figura 6. Esquema de placa para preparación de bibliotecas.....	43
Figura 7. Perfil de las deleciones de los genes <i>pfhrp2</i> y <i>pfhrp3</i> en 172 muestras.	56
Figura 8. Comparación de las proporciones de las deleciones de los genes <i>pfhrp2</i> y <i>pfhrp3</i> entre los distritos de Iquitos y Nauta.	57
Figura 9. Resultados de WGS en base a sWGA.	58
Figura 10. Cobertura de las corridas de WGS.....	60
Figura 11. Análisis de la complejidad de infección mediante Fws.	62
Figura 12. Análisis de componentes principales (PCA)	64
Figura 13. Análisis de ancestría ejecutado con Admixture	66
Figura 14. Decaimiento del Desequilibrio de ligamiento (LD).	68
Figura 15. Manhattan plot de los valores de Fst de Hudson a lo largo del genoma de los parásitos.	69
Figura 16. Manhattan plot del D de Tajima.	70
Figura 17. Manhattan plot del $-\log_{10}P$ -value del $ iHS $	71
Figura 18. Manhattan plot del $-\log_{10}P$ -value para las pruebas XP-EHH y rsb	73
Figura 19. Resultados de análisis de ontología genética para los parásitos con la doble deleción.	75

GLOSARIO

Alelo: Cada una de las dos o más versiones de un gen.

Amplificación Selectiva de Genoma Completo (sWGA): Técnica para amplificar un genoma blanco utilizando secuencias de ADN comunes en el genoma, pero raras en los genomas humano.

Barrido Selectivo: Proceso por el cual una nueva mutación beneficiosa aumenta su frecuencia y se fija (es decir, alcanza una frecuencia de 1) en la población, y conduce a la reducción o eliminación de la variación genética entre las secuencias de nucleótidos cercanas a la mutación.

Desequilibrio de ligamiento: La ocurrencia de combinaciones de loci o genes ligados en proporciones no aleatorias.

Frecuencia alélica: Representa la incidencia de una variante genética en una población.

Haplotipo: Un conjunto de variaciones de ADN, o polimorfismos, que tienden a heredarse juntos.

Hitchhiking genético: Proceso mediante el cual un loci o gen con valor neutro alcanza un valor alto, o incluso fijación, en una población porque está estrechamente relacionado con un gen que se está seleccionando.

Locus: Lugar específico del cromosoma donde está localizado un gen u otra secuencia de ADN. El plural de locus es "loci".

Secuenciación de Genoma Completo (WGS): Proceso de determinar la totalidad, o casi la totalidad, de la secuencia de ADN del genoma de un organismo.

LISTADO DE ABREVIATURAS

ADN: ácido desoxirribonucleico

Dd2: cepa de *Plasmodium falciparum* recolectada de Indochina/Laos

dNTP: deoxinucleotidos trifosfato

Fst: coeficiente de diferenciación intra-poblacional

HB3: cepa de *Plasmodium falciparum* recolectada de Honduras

HRP2: proteína rica en histidina tipo 2

HRP3: proteína rica en histidina tipo 3

Mb: mega bases

p/μl: parásitos por microlitro

pb: pares de bases

PCR: reacción en cadena de la polimerasa

PDR: pruebas de diagnóstico rápido

PF3D7_0831700: gen putativo HSP-70

PF3D7_0831900: gen putativo PHIST-x

PF3D7_1372100: gen putativo PHIST-b

PF3D7_1372400: gen putativo acyl-CoA sintetasa

Pfhrp2: gen codificante de la proteína rica en histidina tipo 2

Pfhrp3: gen codificante de la proteína rica en histidina tipo 3

pfhrp2/pfhrp3: gen codificante de la proteína HRP2 y gen codificante de la proteína
HRP3

sWGA: Amplificación Selectiva de Genoma Completo

WGS: Secuenciación de genoma completo

RESUMEN

Plasmodium falciparum en la Amazonía de Perú se ha descrito en años recientes como una población clonal, presentándose en casos de infecciones monoclonales, y que los parásitos que actualmente circulan serían producto de un reemplazo clonal. Adicionalmente, la frecuencia de los parásitos portadores de la delección de los genes *pfhrp2* y *pfhrp3* ha aumentado, predominando en la población de *P. falciparum*. Por otro lado, no se conoce que factores genéticos podrían subyacer a la persistencia de las infecciones por *P. falciparum* y al incremento en el tiempo de los parásitos con la delección de los genes *pfhrp2/3*. Por ello, el objetivo principal del presente estudio fue identificar regiones genómicas de *P. falciparum* que muestren evidencia de selección positiva reciente para esta especie; y de manera específica, en el grupo de parásitos con la delección de los genes *pfhrp2* y *pfhrp3*. En ese sentido, se ha caracterizado una muestra de 172 parásitos colectadas entre 2013 – 2017, estimándose la frecuencia de la delección de los genes *pfhrp2* y *pfhrp3*. Es así que la delección de *pfhrp2* ocurre en 76.7% de las muestras, la de *pfhrp3* en 63.4%, y la doble delección en 61.6%. Posteriormente, se analizaron los datos de genoma completo de 100 muestras colectadas entre 2006 – 2018 con un máximo de 25% de datos faltantes. Se incluyeron 41 de 54 genomas secuenciados de las muestras procesadas para el perfil *pfhrp2/3*, junto con 59 genomas secuenciadas por colaboradores. A partir de la caracterización de la estructura poblacional de los 100 genomas de *P. falciparum*, se observó la agrupación en base al periodo en el que fueron colectadas y que las muestras del periodo de colecta más recientes (Periodo 2: 2012 – 2018) se agrupaban en base a la presencia del gen *pfhrp2*. Asimismo, se identificó que las muestras del Periodo 2 serían producto de una expansión clonal

posterior a un probable cuello de botella causado por las intervenciones del programa PAMAFRO. En los parásitos del Periodo 2 se identificaron regiones candidatas a estar bajo selección en loci con genes de la familia *var*, codificantes para proteínas que interactúan con el sistema inmunitario del hospedero; que indicarían que la especie *P. falciparum* podría persistir por la evasión de sistema inmunitario. Por otro lado, no se identificó evidencia de selección en los cromosomas 8 y 13, en la región cercana a las deleciones de *pfhrp2* y *pfhrp3*, respectivamente. Pero, en los parásitos con la deleción *pfhrp2/3*, se identificaron loci candidatos a estar bajo selección reciente. Entre los genes que se identificaron, se incluyen *AP2-G* y *FP2A*, el primero codificante de un factor de transcripción que desencadena la diferenciación a las etapas sexuales del parásito para su transmisión, y el segundo codifica una proteína involucrada en la digestión de hemoglobina y ha sido reportado como candidato a conferir resistencia a la artemisinina a partir de estudios *in vitro* y *ex vivo*. Los hallazgos del presente trabajo complementan los reportes previos sobre el cambio en el tiempo en la población de *P. falciparum* en la Amazonía peruana, y brinda indicios de los mecanismos biológicos que explican la permanencia del parásito y de forma específica del incremento del grupo de parásitos con la deleción de *pfhrp2/3*, resultados que pueden ser de relevancia en las estrategias de vigilancia molecular de la especie.

Palabras clave: *Plasmodium falciparum*, deleción, *pfhrp2*, *pfhrp3*, genoma, selección

ABSTRACT

Plasmodium falciparum in the Amazon of Peru has been described in recent years as a clonal population, presenting in cases of monoclonal infections, and that the parasites would be the product of a clonal replacement. In addition, the frequency of parasites carrying the deletion of the *pfhrp2* and *pfhrp3* genes has increased, predominating in the *P. falciparum* population. On the other hand, it is not known what genetic factors could underlie the persistence of *P. falciparum* infections and the increase over time of the parasites with the deletion of the *pfhrp2/3* genes. Therefore, the main objective of the present study was to identify genomic regions of *P. falciparum* that show evidence of recent positive selection for this species; and specifically, in the group of parasites with the deletion of the *pfhrp2* and *pfhrp3* genes. In this sense, a sample of 172 parasites collected between 2013 - 2017 has been characterized, estimating the frequency of deletion of the *pfhrp2* and *pfhrp3* genes. Thus, the deletion of *pfhrp2* occurs in 76.7% of the samples, that of *pfhrp3* in 63.4%, and the double deletion in 61.6%. Subsequently, the whole genome data of 100 samples collected between 2006 – 2018 with a maximum of 25% missing data were analyzed. Forty-one of 54 sequenced genomes from samples processed for the *pfhrp2/3* profile were included, along with 59 collaborator-sequenced genomes. From the characterization of the population structure of the 100 *P. falciparum* genomes, the grouping was observed based on the period in which they were collected and that the samples from the most recent collection period (Period 2: 2012 - 2018) were grouped based on the presence of the *pfhrp2* gene. Likewise, it was identified that the samples from Period 2 would be the product of a clonal expansion after a probable bottleneck caused by the interventions of the

PAMAFRO program. In Period 2 parasites, candidate regions to be under selection were identified in loci with protein-coding genes of the var family, which would indicate that the species *P. falciparum* could persist by evading the immune system. On the other hand, no evidence of selection was identified on chromosomes 8 and 13, in the region close to *pfhrp2* and *pfhrp3* deletions, respectively. But, in parasites with the *pfhrp2/3* deletion, loci candidates for recent selection were identified. The genes that were identified were *AP2-G* and *FP2A*, the first encoding a transcription factor that triggers the differentiation to the sexual stages of the parasite for its transmission, and the second encoding a protein involved in the digestion of hemoglobin and has been reported as a candidate to confer resistance to artemisinin from in vitro and ex vivo studies. The findings of the present work complement previous reports on the change over time in the population of *P. falciparum* in the Peruvian Amazon, and provide indications of the biological mechanisms that explain the permanence of the parasite and specifically the increase in the group of parasites. with the deletion of *pfhrp2/3*, results that may be relevant in the molecular surveillance strategies of the species.

Keywords: *Plasmodium falciparum*, deletion, *pfhrp2*, *pfhrp3*, genome, selection

I. INTRODUCCIÓN

La malaria es una enfermedad tropical que sigue siendo un problema patente en el mundo y en regiones tropicales como Perú. Esta enfermedad es considerada reemergente, debido al incremento en el número de casos que ha experimentado en distintos periodos luego de las intervenciones del Ministerio de Salud (1). El principal vector transmisor de la infección es el mosquito hembra del género *Anopheles* infectado con parásitos del género *Plasmodium*. *Plasmodium vivax* es responsable de cerca del 84% de casos en Perú, mientras que *Plasmodium falciparum* es responsable de la forma más severa de malaria y causa cerca del 16% de casos (1).

Una etapa crucial en el control de la enfermedad es el diagnóstico. La técnica estándar de oro para el diagnóstico de malaria es la microscopía; sin embargo, su principal limitación es su baja sensibilidad en casos de baja parasitemia, además de ser una técnica laboriosa y de requerir de personal especializado. Un método diagnóstico alternativo es mediante las pruebas de diagnóstico rápido (PDRs) que detectan antígenos producidos por los parásitos, y se caracterizan por su mayor sensibilidad y fácil implementación para su aplicación en campo (2).

En el caso de las PDRs para diagnóstico de *P. falciparum*, el antígeno blanco de interés es la proteína rica en histidina tipo 2 (HRP2). Esta proteína es secretada en grandes cantidades en la etapa sanguínea del parásito, y debido a que presenta secuencias repetidas de alanina e histidina en su estructura, su detección mediante PDRs es potenciada. Asimismo, la proteína homóloga HRP3, codificada por el gen *pfhrp3*, también es detectada por las PDRs dirigidas a HRP2. Sin embargo, la

presencia de parásitos con la delección de los genes *pfhrp2* y *pfhrp3*, puede conducir a diagnósticos con resultados errados (3,4).

La delección de los genes *pfhrp2* y *pfhrp3* ocurren en alta frecuencia en la selva de Perú. El año 2010 se reportó la presencia de parásitos colectados en campo con la delección del gen *pfhrp2*, posteriormente se ha reportado el aumento de su frecuencia, así como de la delección del gen *pfhrp3* y de la doble delección de *pfhrp2* y *pfhrp3*, en muestras colectadas durante los años 1998 – 2001, 2003 – 2005, 2009 – 2010, 2012 – 2014, y en un estudio que abarcó muestras del 2011 al 2018 (3,5–8). Actualmente, la delección de *pfhrp2* se ha detectado en países como Colombia y Brasil (9,10); en este último país en 159 muestras de 2016 – 2017, 82.3% portaban la delección de *pfhrp2*, 96.2% de *pfhrp3* y 79.8% la delección de ambos genes (11). Asimismo, la delección de *pfhrp2* se reporta en países de África (12), donde *P. falciparum* causa más del 90% de casos de malaria.

Por otro lado, estudios de genética de poblaciones con marcadores microsatélites y de genoma han sugerido que la población de *P. falciparum* ha venido sufriendo reducciones en el número de clonas circulantes como consecuencia de las intervenciones de programas como el Proyecto Control de la Malaria en Zonas Fronterizas de la Región Andina: Un Enfoque Comunitario - PAMAFRO (13), y que en la actualidad estaría predominando una población clonal que a su vez correspondería con los parásitos con la delección de *pfhrp2/3* (5,7,8,14).

A partir de estudios de genoma completo se han detectado en regiones del genoma del parásito, *loci* que están bajo presión de selección y que contienen genes involucrados en resistencia a fármacos empleados en el tratamiento de la

enfermedad, así como otros procesos biológicos como la evasión del sistema inmunitario del hospedero y la diferenciación a las formas sexuales (15,16). Recientemente se ha reportado que la delección del gen *pfhrp2* también estaría siendo seleccionada en muestras colectadas en Etiopía y analizadas mediante la secuenciación de la región del cromosoma 8 y 13 (17). Sin embargo, estos estudios se concentran en la región de África, donde la enfermedad y el parásito se desarrollan en diferentes circunstancias. Para empezar, la transmisión de la malaria en África es elevada, lo que es contrastante a lo observado en la Amazonía peruana (18–20). También se ha sugerido que el uso de PDRs y la búsqueda de tratamiento a la enfermedad serían algunos de los factores que ejercerían presión de selección en este grupo de parásitos (17,21), algo que en Perú no es extrapolable dado el limitado uso de las PDRs.

En dicho contexto, en este estudio se ha llevado a cabo la caracterización de una muestra de parásitos de la especie *Plasmodium falciparum* con el fin de estimar la frecuencia de la delección de los genes *pfhrp2* y *pfhrp3* en la Amazonía peruana. Asimismo, a partir de datos de genoma completo se ha descrito la población de los parásitos con la delección y sin la delección de dichos genes. Finalmente, se han identificado regiones del genoma con evidencia de selección reciente que pueden dar indicios de los factores genéticos que explican porque los parásitos con la delección de los genes *pfhrp2/3* habría aumentado en el tiempo.

II. ANTECEDENTES

2.1. EPIDEMIOLOGÍA DE LA MALARIA

La malaria es un problema persistente en países tropicales en desarrollo. El año 2020 se reportaron alrededor de 241 millones de casos de malaria a nivel mundial y 627 mil muertes, de estos, el 93% de casos y el 94% de muertes se registraron en África. Alrededor del 95% de los casos fue causado por el parásito *Plasmodium falciparum*, ocurriendo principalmente en África. En contraste, en regiones de América central y del sur y regiones de clima templado, la causa de malaria se adjudica principalmente a *Plasmodium vivax* (22).

En Perú el número de casos de malaria reportados a finales del 2022 fue de 24701. Del total de casos de malaria en Perú, alrededor del ~84% fue causado por *P. vivax* y ~16% por *P. falciparum* (1). Actualmente, los casos de malaria por *P. vivax* y *P. falciparum* en Perú se concentran en la región Amazónica, principalmente en Loreto (>80%) (1). El alto porcentaje de casos en la región amazónica se debería a la reportada heterogeneidad en la transmisión de la malaria (18), factores sociodemográficos, como la edad, educación, ocupación y movilidad humana, así como las características de los hogares y las condiciones ecológicas específicas de la zona, estarían influenciando en el patrón de transmisión (18,23). Por otro lado, el alto número de infecciones asintomáticas y submicroscópicas en la Amazonía peruana, portadores de gametocitos, constituyen el reservorio que mantiene la transmisión en esta región ya que por sus características, escapan a la vigilancia rutinaria de la malaria por parte del Ministerio de Salud (MINSA) (24,25). A pesar de los factores mencionados anteriormente, el número de casos de malaria ha disminuido en los últimos años (Figura 1), a consecuencia de las medidas de control

adoptadas en la región como parte del Plan Malaria Cero ejecutado entre 2017-2021, y recientemente por el Plan Hacia la Eliminación de la malaria con periodo de ejecución 2022-2030 (26–28).

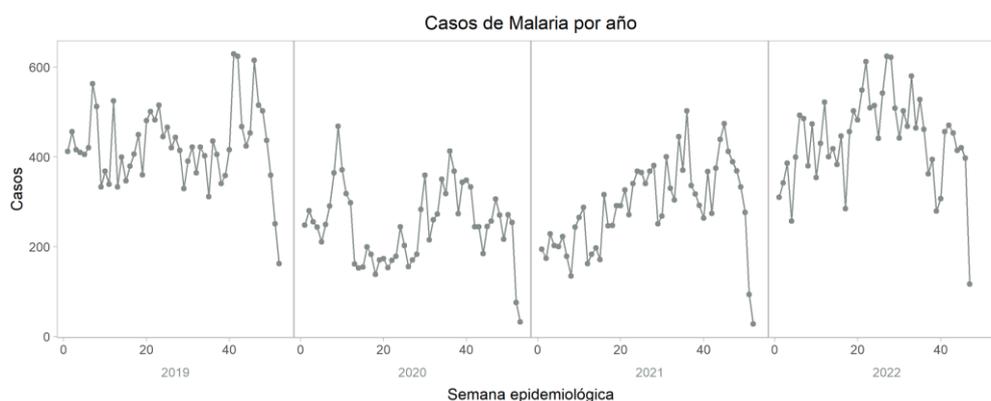


Figura 1. Número de casos de malaria entre 2019 – 2022 en la región Loreto. Figura generada a partir de datos obtenidos de CDC Perú (1).

2.2. ETIOLOGÍA DE LA MALARIA

La malaria es causada por parásitos del género *Plasmodium*. Los parásitos de *Plasmodium* pertenecen al grupo de protozoarios Apicomplexa, que se distinguen morfológicamente por la presencia de orgánulos apicales especializados involucrados en la invasión de la célula huésped (micronemes, rhoptries y gránulos densos) (2). Existen cinco especies de *Plasmodium* que pueden infectar al humano: *P. falciparum*, *P. vivax*, *P. malariae*, *P. knowlesi* y *P. ovale*. El vector que facilita la transmisión de los parásitos es el mosquito hembra del género *Anopheles*.

2.2.1. Ciclo de vida del parásito

El ciclo de vida de *Plasmodium* es similar en todas las especies del parásito. La transmisión de los parásitos ocurre por la picadura de los mosquitos hembra del género *Anopheles* e involucra a dos huéspedes. Cuando el *Anopheles* hembra

infectado se alimenta de sangre, inocula esporozoitos en el huésped humano. Los esporozoitos infectan las células hepáticas y maduran en esquizontes, los cuales al romperse liberan merozoitos (ciclo exoeritrocítico). Después de la replicación inicial en el hígado, los merozoitos infectan los glóbulos rojos y experimentan una multiplicación asexual (ciclo eritrocítico). Dentro de los eritrocitos, los merozoitos atraviesan un proceso de maduración pasando por los estadios de trofozoíto joven, adulto y esquizonte, estos últimos se rompen liberando merozoitos. Las manifestaciones clínicas de la enfermedad son adjudicadas a los parásitos en la etapa sanguínea. Asimismo, algunos merozoitos se diferencian en las formas sexuales del parásito, gametocitos masculinos (microgametocitos) y femeninos (macrogametocitos). Los gametocitos son ingeridos por el mosquito cuando se alimenta, dentro del mosquito tiene lugar la fase sexual del parásito (ciclo esporogónico). En el estómago del mosquito ocurre la formación del cigoto por la unión de los gametocitos. Los cigotos se vuelven móviles y alargados (oocinetos) que invaden la pared del intestino medio del mosquito donde se convierten en ooquistes. Los ooquistes crecen, se rompen, liberando esporozoitos que llegan a las glándulas salivales del mosquito. El ciclo de vida del parásito inicia nuevamente con la inoculación de los esporozoitos en un nuevo huésped humano. Por otro lado, en las especies *P. vivax* y *P. ovale* la forma esquizonte hepática puede permanecer latente como hipnozoito hasta su reactivación, la cual puede ocurrir incluso años después de la infección (2).

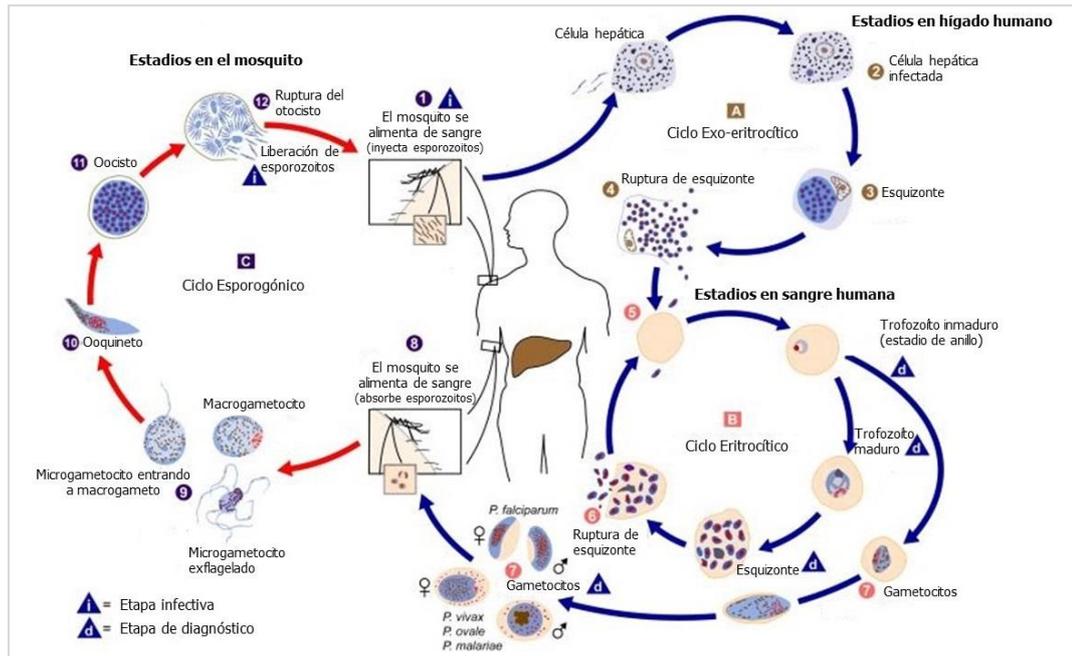


Figura 2. Ciclo de vida de *Plasmodium*. Adaptado de (28).

2.3. DIAGNÓSTICO DE MALARIA

Clínicamente, la malaria se puede diagnosticar en base a los síntomas como fiebre, escalofríos y dolor de cabeza. Adicionalmente, se consideran algunos factores de riesgo como el residir en zonas endémicas, la historia de viajes a zonas endémicas con presencia de vector. Sin embargo, el diagnóstico definitivo es mediante la demostración de la presencia de parásitos o sus componentes mediante pruebas de laboratorio (29).

2.3.1. Microscopía

El examen de la gota gruesa, teñida con Giemsa, es el método estándar y rutinario que se emplea para la identificación de parásitos de malaria. Entre las principales ventajas de la microscopía es que es una técnica de bajo costo, permite diferenciar especies de *Plasmodium*, se puede determinar la densidad parasitaria y es útil para monitorear la respuesta a la terapia (29). Esta técnica puede alcanzar una

sensibilidad de hasta 10 parásitos por microlitro de sangre cuando es realizada por un microscopista experto; sin embargo, la sensibilidad en condiciones de campo es de alrededor de 100 parásitos por microlitro de sangre (29,30).

2.3.2. Diagnóstico basado en tecnologías moleculares

El diagnóstico mediante pruebas moleculares se basa en la amplificación por reacción en cadena de la polimerasa (PCR) de secuencias de ADN de parásitos del género *Plasmodium*. La principal ventaja que ofrecen es su alta sensibilidad y especificidad. Este grupo de pruebas puede alcanzar a detectar, en condiciones óptimas, infecciones por 0.02 parásitos/ μ L (31), este es el caso de la PCR en tiempo real. Sin embargo, tiene como principal desventaja la dificultad de implementación para su aplicación en campo. Como alternativa se han desarrollado pruebas basadas en PCR como la amplificación isotérmica mediada por bucles, LAMP (*Loop-mediated isothermal DNA amplification*). Esta técnica no requiere de un equipo termociclador y su implementación es sencilla ya que no requiere de mayor experticia (32); asimismo, llega a alcanzar una especificidad y sensibilidad promedio en muestras de campo de 91.9% y 91.8%, respectivamente, respecto a un procedimiento estándar de PCR en tiempo real (33).

2.3.3. Pruebas de diagnóstico rápido

Las pruebas de diagnóstico rápido (PDR) son una alternativa de herramienta de diagnóstico, en situaciones en las que el examen por microscopía no sea factible de realizar. Las PDR se basan en la detección inmunológica de antígenos del parásito en la sangre de un individuo con sospecha de infección por *Plasmodium* (30).

Las PDR disponibles para el diagnóstico de malaria se basan en la detección de la proteína HRP2, lactato deshidrogenasa (LDH) y aldolasa. Las PDR que detectan LDH y aldolasa permiten detectar diferentes especies de parásito, o de manera general al género *Plasmodium* (pan LDH o pLDH) (4,31). Por otro lado, las PDR que detectan HRP2 en sangre son específicas para *Plasmodium falciparum* (4,31).

La principal ventaja de las PDR que detectan HRP2 es la alta sensibilidad de la prueba debido a las características de la proteína. La proteína HRP2 es una proteína de la etapa asexual del parásito, asimismo las secuencias repetidas de su estructura contribuyen a que el antígeno presente múltiples epítomos (4). Sin embargo, el hallazgo de parásitos que no tienen el gen *pfhrp2* en muestras obtenidas en campo, puede conducir a un diagnóstico de falso negativo de la enfermedad (3).

2.4. EVOLUCIÓN POR SELECCIÓN NATURAL

El principio de la evolución por selección natural se basa en que los rasgos que aumentan las posibilidades de supervivencia y reproducción tienen una mayor tendencia a transmitirse a la siguiente generación. En la era genómica, la selección se refiere a cualquier propagación diferencial no aleatoria de un alelo como consecuencia de su efecto fenotípico.

2.4.1. Modos de Selección

La forma más simple en la que la selección actúa es aquella en la que un alelo es favorecido y propagado, lo que se conoce como selección positiva. Por otro lado, cuando ocurre la aparición de un alelo perjudicial este disminuye en frecuencia en la población al punto de ser eliminado por selección negativa (34,35). En este sentido, es más probable que las mutaciones que ocurren de manera aleatoria sean

perjudiciales, por lo cual se someten a una selección negativa para ser eliminadas del acervo genético antes de que alcancen frecuencia detectable. Este tipo de selección negativa es denominado selección de fondo (*background selection*) (35). Otro fenómeno en el cual un fenotipo depende de la interacción de múltiples alelos de un locus es la selección de equilibrio (*balancing selection*), en el cual múltiples alelos son mantenidos a una frecuencia apreciable dentro de un conjunto génico. A pesar de esta diversidad de modos de selección, la investigación actual se ha centrado en desarrollar métodos genómicos para identificar evidencias de selección positiva. Ese interés se debe a que este modo de selección deja huellas que son detectables en contraste a los otros modos de selección. Asimismo, al ser la selección positiva el mecanismo primario de adaptación, su identificación toma gran interés teórico en investigación (34,35). Además, identificar estas huellas de selección es de utilidad en el entendimiento de la biología de organismos de relevancia en salud, como es en el caso de las enfermedades infecciosas parasitarias.

2.4.2. Hitchhiking y Barridos selectivos (Selective Sweeps)

El término *Hitchhiking* se le atribuye a Maynard Smith y Haigh para describir el aumento en la frecuencia de un alelo neutro que está ligado a un alelo que está bajo selección positiva. En consecuencia, la variabilidad genética de la vecindad genómica del alelo bajo selección se ve afectada.

El término barrido selectivo, se refiere al barrido de la variación que ocurre alrededor de un sitio seleccionado después que un alelo favorable aumenta en frecuencia hasta fijarse. El término barrido parcial se refiere a un evento en el que un alelo aumenta rápidamente desde una frecuencia baja, pero sin alcanzar la fijación (debido a que el barrido sigue en curso, o a que la ventaja selectiva del alelo

ventajoso se ha debilitado). Por otro lado, el término barrido fuerte (*Hard sweep*) hace alusión al modelo clásico de barrido selectivo en el que una nueva mutación ventajosa surge y se propaga rápidamente hasta fijarse por selección natural. Bajo este modelo, la variación neutra cerca del sitio que es favorecido se modifica junto con el alelo favorecido. Esto se refleja en los patrones de variación alrededor del sitio seleccionado que son detectables mediante una variedad de pruebas de selección (36,37).

Otro concepto relacionado al barrido selectivo es el de barrido suave (*Soft sweep*), y describe dos escenarios que contrastan con el barrido fuerte. En un barrido suave de origen único, la mutación eventualmente favorable ya existía previamente al inicio de la selección, siendo inicialmente neutro o incluso ligeramente perjudicial. El alelo se desplaza por la población y se propaga potencialmente con diferentes orígenes de haplotipos a través de la recombinación, hasta que finalmente un cambio en el entorno lo favorece. En contraste, bajo un barrido suave de origen múltiple, el alelo favorable fijo no descende de una sola mutación, sino de una colección de múltiples eventos independientes, un escenario que es más probable en poblaciones más grandes. Los barridos suaves se caracterizan por la dificultad para ser detectados (36–38).

2.4.3. Métodos para detección de huellas de selección

2.4.3.1. Métodos para detectar macroevolución

Los métodos de detección de selección a nivel macroevolutivo dependen de comparaciones de rasgos o secuencias homólogos entre taxas relacionadas.

2.4.3.1.1. Métodos basados en genes

Las sustituciones sinónimas se asumen como selectivamente neutras. De este modo, informan sobre la tasa de evolución de fondo (*background selection*). Se considera indicativo de selección si la tasa de sustitución no sinónima difiere significativamente. Ente estos métodos se incluyen la relación dN/dS y la prueba de McDonald-Kreiman (34).

2.4.3.1.2. Métodos basados en otras tasas

Estos métodos detectan las regiones que experimentan un cambio acelerado en un linaje, pero que se conservan en linajes relacionados. Estas regiones son probables candidatas para la selección.

2.4.3.2. Métodos para detectar microevolución

La selección positiva hace que un alelo ventajoso alcance una alta prevalencia o fijación (100% de prevalencia) rápidamente dentro de una población. Cuando un alelo favorable y las variantes circundantes de ese haplotipo alcanzan una alta prevalencia en conjunto, se produce una reducción de la diversidad genética (heterocigosidad, polimorfismo o variabilidad) que rodea al alelo causal. Esta reducción, persiste hasta que la recombinación y las mutaciones restauran la diversidad. Los métodos descritos a continuación se orientan a detectar estas huellas de selección (34).

2.4.3.2.1. Métodos basados en frecuencias

En un barrido selectivo, una variante genética logra una elevada frecuencia junto con las variantes vinculadas cercanas (alelos derivados de alta frecuencia). De este

background homogéneo surgen nuevos alelos, pero inicialmente tienen una frecuencia baja (exceso de alelos raros). Entre las pruebas que detectan este tipo de huellas se incluyen la prueba de Ewens-Waterson, el cálculo de D de Tajima, H de Fay y Wu (34,39).

2.4.3.2.2. Métodos basados en desequilibrio de ligamiento

Los barridos selectivos conllevan a que una región genética alcance una alta prevalencia en una población, esta región incluye la variante causal y sus alelos vecinos. Las asociaciones entre estos alelos definen un haplotipo, que persiste en la población hasta que la recombinación rompe estas asociaciones. Entre los métodos basados en desequilibrio de ligamiento se pueden mencionar la prueba de haplotipo de rango amplio (LRH), el *score* de haplotipo integrado (iHS), la evaluación de haplotipo homocigoto extendido entre poblaciones (XP-EHH), el decaimiento de desequilibrio de ligamiento (LDD) (34,37,39).

2.4.3.2.3. Métodos basados en diferenciación poblacional

La selección ocurre sobre un alelo en una población, pero no en otra, esto crea una marcada diferencia en la frecuencia de ese alelo entre las dos poblaciones. Este efecto de diferenciación destaca en comparación a la diferenciación entre poblaciones en relación a los alelos neutros no seleccionados (36). La prueba F_{st} , la prueba de Lewontin-Krakauer (LKT), y su variante hapFLK se incluyen en este grupo (34,37,39).

2.4.3.2.4. Métodos compuestos

Se incluyen dos aproximaciones. La primera considera que la combinación de las puntuaciones de las pruebas para varios sitios en una región contigua puede reducir

la tasa de falsos positivos. En la segunda estrategia la combinación de múltiples pruebas independientes en un sitio puede mejorar la resolución, permitiendo distinguir variantes causales (34,37,39). En este sentido, diferentes pruebas pueden proporcionar información complementaria. En el primer grupo se incluyen la relación de verosimilitud compuesta (*Composite likelihood ratio* - CLR) y relación de verosimilitud compuesta entre poblaciones (XP-CLR). En el segundo grupo se incluye la prueba DH (34).

2.5. DETECCIÓN DE HUELLAS DE SELECCIÓN EN *P. falciparum*

Los parásitos como *P. falciparum* sufren presión de selección por distintos factores. El sistema inmunitario del hospedero y del vector ejerce presión de selección en los parásitos. De igual forma, las estrategias de control que tienen como objetivo reducir el número de infecciones también ejercen presión selectiva en los parásitos. De este modo, los tratamientos farmacológicos han llevado a la aparición e incremento de mutaciones asociadas con resistencia a antimaláricos. Análisis de la región genómica cercana a las mutaciones en los genes *pfdhps* y *pfdhfr*, empleando microsatélites, han permitido identificar barridos selectivos asociados a la mutación en *pfdhps* en muestras de Venezuela. Esto fue evidenciado a través de la disminución de la heterocigosidad en la región genómica, y que aparentemente se mantendría incluso después del retiro del fármaco sulfadoxina-pirimetamina (40). Por otro lado, en Kenia se reportó un efecto de presión de selección en el gen *pfmdr1* distinto entre la zona costera y oeste que guardaría relación con la diferente transmisión entre ambas zonas (41).

Los escaneos de genoma completo en África han identificado selección en diferentes partes del genoma de *P. falciparum*. Al comparar el efecto de las

estrategias de control en regiones con diferente grado de endemicidad en África (Guinea, alta transmisión y Gambia, menor transmisión), han permitido identificar huellas de selección en otros genes. Adicionalmente a los genes asociados a la resistencia a antimaláricos (*pfert*, *pfdhps*, *pfdhpr*, *pfmdr1*), se detectó evidencia de selección en genes involucrados en la invasión y codificantes de antígenos como *msp1*, *glurp* y *ama* (15,42). Asimismo, se reportó un alto grado de diferenciación entre ambas poblaciones con respecto al gen *gdv1*, el cual se expresaría en las etapas sexuales del parásito y que se podría relacionar a diferencias en la transmisión en ambas regiones (42). Finalmente, en un estudio en el que se compararon grupos de parásitos diferenciados espacial y temporalmente, se destaca la presencia de un SNP (65K) en el gen *pfafs* que para la fecha de las muestras (2014), presentaría evidencia de reciente selección asociada con la introducción y resistencia al fármaco lumefantrina (16).

La selección de los parásitos de la malaria varía en regiones con diferentes intensidades y frecuencia de transmisión o incidencia de infección. Los parásitos en áreas altamente endémicas experimentan competencia dentro del hospedero en la etapa asexual a causa de la superinfección (43). Por el contrario, en áreas de baja endemicidad dentro del hospedero existe menor competencia o selección por el sistema inmunitario y las oportunidades de transmisión son limitadas debido a que las poblaciones de mosquitos son estacionales y de baja densidad (42).

2.6. DELECIÓN DE LOS GENES *pfhrp2* y *pfhrp3* COMO UN CASO POSIBLE DE SELECCIÓN NATURAL EN *P. falciparum*

Los genes que codifican para las proteínas HRPs se ubican en regiones cromosómicas subteloméricas altamente variables. El gen *pfhrp2* está ubicado en el cromosoma 8 (PlasmoDB ID: PF3D7_081800) y el gen *pfhrp3* en el cromosoma 13 (PlasmoDB ID: PF3D7_1372200). El gen *pfhrp2* está flanqueado por los genes PF3D7_0831900 y PF3D7_0831700, en las posiciones *upstream* y *downstream* respectivamente. Asimismo, el gen *pfhrp3* está flanqueado por los genes PF3D7_1372400 y PF3D7_1372100, en las posiciones *upstream* y *downstream* respectivamente (Figura 3).

El gen *pfhrp2* tiene una longitud de 1064 pares de bases (pb) y el gen *pfhrp3* de 977 pb. Ambos genes constan de dos exones, el primer exón y el inicio del segundo codifica la secuencia señal y de corte de la proteína. Dentro del segundo exón se ubican las secuencias ricas en histidina de los genes *pfhrp2* y *pfhrp3* (4). Las regiones flanqueantes a la secuencia repetitiva muestran un porcentaje de similitud de 85-90% a nivel de ADN; esto último indicaría que ambas secuencias descienden de una secuencia común que habría sufrido un proceso de duplicación y luego divergió (4).

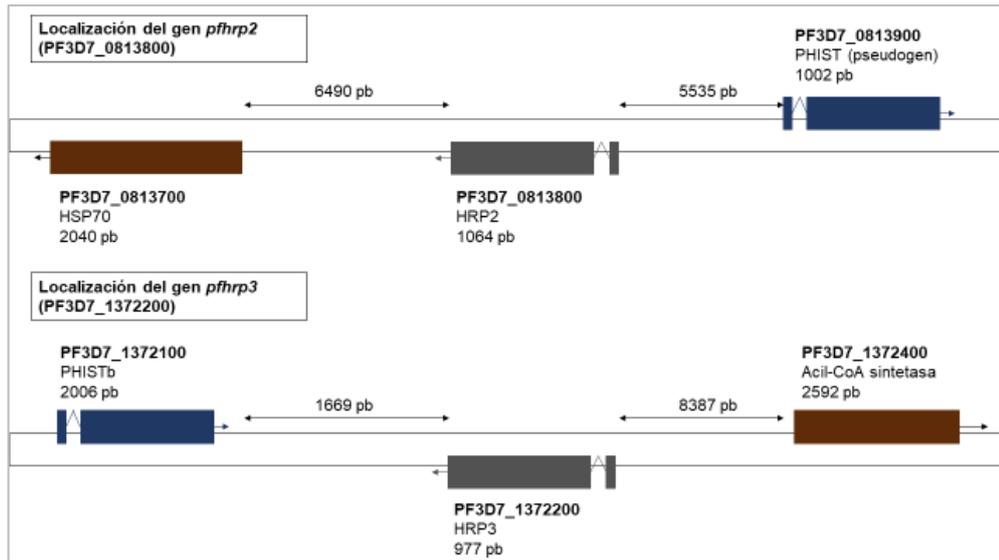


Figura 3. Ubicación genómica del gen *pfhrp2* en el cromosoma 8 (superior) y del gen *pfhrp3* en el cromosoma 13 (inferior). Adaptado de <https://plasmodb.org/>.

2.7. FUNCIÓN DE LA PROTEÍNA HRP2

Se ha postulado que la proteína HRP2 participa en la detoxificación de moléculas del grupo hemo. Los grupos hemo son el producto de la digestión de la hemoglobina. La proteína HRP2, *in vitro*, se une al grupo hemo y media su cristalización a un producto menos tóxico denominado hemozoína (4). Por otro lado, la proteína HRP3 también participaría en la polimerización del hemo a hemozoína (44). Sin embargo, los parásitos con la delección de *pfhrp2* y *pfhrp3* siguen produciendo hemozoína, por lo que se sugiere que habrían otros mecanismos independientes de HRP2 (4,44,45). Entre estos puede mencionarse la participación de la proteína HDP (heme detoxification protein) que se encontraría conservada funcionalmente en el género *Plasmodium* (46), por lo que el parásito tendría proteínas con funciones redundantes.

2.8. *P. falciparum* CON LA DELECCIÓN DE LOS GENES *pfhrp2* Y *pfhrp3*

2.8.1. Distribución de los parásitos con las deleciones de los genes *pfhrp2* y *pfhrp3*

El primer reporte de parásitos con la deleción de los genes *pfhrp2* y *pfhrp3* en muestras de campo fue de muestras de la Amazonía peruana. Gamboa y colaboradores reportaron una prevalencia de 41% y 70% de la deleción del gen *pfhrp2* y *pfhrp3*, respectivamente y 21.6% para la deleción de ambos genes, en 148 parásitos colectados en el periodo 2003- 2007 (3). Posteriormente se ha reportado que la frecuencia de los parásitos con la doble deleción ha aumentado en frecuencia, siendo así que entre muestras de los años 2011 y 2018 el gen *pfhrp2* está delecionado en 77.2% y *pfhrp3* en 71.6% y la doble deleción en 67% (14).

Asimismo, los parásitos con las deleciones de los genes *pfhrp2* y *pfhrp3* y sus genes flanqueantes se han reportado en otras regiones en frecuencias variables. En Sudamérica las regiones donde las deleciones se han reportado incluyen a los estados de Acre, Rondonia y Pará en Brasil, con frecuencias de 13.6% *pfhrp2*-, 35.9% *pfhrp3*- y 11.6% *pfhrp2*-/*pfhrp3*-; en la Amazonía de Colombia con 53.6% *pfhrp2*-, 71.4% *pfhrp3*- y 53.6% *pfhrp2*-/*pfhrp3*-; y en Surinam con 14% *pfhrp2*-, 4% *pfhrp3*- y 2.6% *pfhrp2*-/*pfhrp3*- (9,10,47). En Centroamérica (Honduras, Nicaragua y Guatemala) se reporta una frecuencia global de 25.8% *pfhrp2*-, 91.4% *pfhrp3*- y 21% *pfhrp2*-/*pfhrp3*- (48). Asimismo, en regiones del continente africano, la frecuencia reportada toma valores desde 0.4% en Angola, 2.4% en Senegal, 6.4% en la República del Congo, 36.2% en Ghana, hasta 60 % en Eritrea (12,49–51).

Las deleciones de los genes *pfhrp2* y *pfhrp3* con sus genes flanqueantes evidencian un aumento en su frecuencia en la selva de Perú. El análisis de Akinyi y colaboradores (47) con muestras colectadas en los periodos 1998-2001 y 2003-2005 reporta un incremento en la deleción del gen *pfhrp2* de 20% a 40%, de manera similar se observó un aumento de la deleción de su gen flanqueante P3D7_0831900 de 39% a 44%, lo que no se observó con el gen flanqueante P3D7_0831700 (5). Posteriormente, el estudio de Quispe del 2017 (6), empleando muestras colectadas entre los años 2012 y 2014 reportó una frecuencia de 53% de la deleción del gen *pfhrp2*, de 48% del gen *pfhrp3* y de 37% de la deleción de ambos genes. En este último estudio la frecuencia reportada de la deleción del gen flanqueante P3D7_0831900 fue de 31.3%; cabe resaltar el perfil genético variado de deleciones que puede o no incluir a los genes *pfhrp2* y *pfhrp3* y sus genes flanqueantes (6) (Tabla 1).

Tabla 1. Perfiles de deleción de *pfhrp2*, *pfhrp3* y genes flanqueantes de *P. falciparum* en 227 muestras. Quispe, 2017 (6)

Comunidad de estudio	Cahuide	San José de Lupuna	Santa Emilia
Variable			
Número de muestras	76	58	93
Perfil <i>pfhrp2</i> (PF3D7_0831700, <i>pfhrp2</i>, PF3D7_0831900)			
+++	81 (35.7%)		
+---	71 (31.3%)		
----	23 (10.1%)		
+--+	20 (8.8%)		
++--	17 (7.5%)		
---+	8 (3.5%)		
--++	4 (1.8%)		
-+-	3 (1.3%)		
Perfil <i>pfhrp3</i> (PF3D7_1372100, <i>pfhrp3</i>, PF3D7_1372400)			
+++	104 (45.8%)		
---+	52 (22.9%)		
++--	34 (15%)		
----	14 (6.2%)		
+---	10 (4.4%)		
--++	8 (3.5%)		
++--	3 (1.3%)		
-+-	2 (0.9%)		

+: Presencia del gen

—: Ausencia del gen

PF3D7_0831700: Gen HSP-70

La delección de *pfhrp2* podría tener múltiples orígenes. En un trabajo del 2013, Akinyi y colaboradores identificaron que los parásitos con la delección del gen *pfhrp2* colectados en los periodos de 1998-2001 y 2003-2005, se agrupan en diferentes *clusters* poblacionales. Los 4 *clusters* identificados en el primer periodo corresponderían a los linajes A-D de *P. falciparum*, mientras que en el segundo periodo se observaron *clusters* híbridos de los linajes identificados previamente. Por otro lado, el análisis de ancestría local con haplotipos construidos con marcadores microsatélites cercanos a la región del gen *pfhrp2* demostraron que los haplotipos del primer periodo seguían presentes en el segundo periodo (5). Sin embargo, los hallazgos de Quispe indican que los parásitos *pfhrp2* negativos aparecen en mayor frecuencia y que se agruparían en un único *cluster* (6). Posteriormente, dos estudios publicados en 2022 describen algunos aspectos poblacionales de los parásitos con las delecciones *pfhrp2/3*, donde se menciona que los parásitos habrían estado conformados por distintas clonas durante los años de 2009-2010 y que en años más recientes habría una clona predominante, la cual porta la delección (7,8).

A pesar de los hallazgos descritos, se desconoce si la aparición y aumento en frecuencia de parásitos con la delección de *pfhrp2* y de la delección de *pfhrp3* en la Amazonía peruana sería producto de un efecto de presión selectiva específico. En el caso de África se ha identificado evidencia de selección de la delección del gen *pfhrp2* posiblemente por el uso predominante de PDRs que permitiría que las clonas con la delección no sean diagnosticadas y por lo tanto, no tratadas, a diferencia de los casos con las clonas portadoras del gen que son detectadas. Algo que había sido

sugerido previamente mediante modelos matemáticos, junto a otros factores como el nivel de transmisión (20,21).

La delección de los genes que flanquean a *pfhrp2* no tendría un efecto en el desarrollo de *P. falciparum*. Entre los genes que flanquean a *pfhrp2*, se incluye al gen que codifica para la proteína hsp70-x, una proteína chaperona exclusiva de *P. falciparum* y la única que se ha reportado que es exportada al eritrocito del hospedero durante el desarrollo asexual del parásito (52). Sin embargo, se ha demostrado mediante estudios *in vitro* que no es indispensable para el desarrollo normal del parásito (53,54). El otro gen que flanquea a *pfhrp2* es PHIST, el cual codifica una proteína parte de una familia de 89 proteínas poco estudiada que se expresan en distintas etapas del desarrollo del parásito (55,56). Sin embargo, se desconoce la función de la proteína producto del gen PHIST. De manera similar, el gen *pfhrp3* está flanqueado también por un gen de la familia PHIST, y por el extremo 5' por un gen codificante de una enzima de la familia de las acil Coenzima A sintetasa (Acil-CoA-sintetasa). *P. falciparum* tiene 11 genes de esta familia distribuidos en los cromosomas 2, 3, 4, 5, 7, 12, 13, y 14, los cuales se expresarían en la etapa eritrocítica del parásito y participarían en el metabolismo de lípidos. Esta familia de genes estaría sufriendo recombinación y conversión genética, y también se ha sugerido que estarían bajo selección que favorecería los cambios de la proteína; sin embargo, se desconoce si la delección del gen que flanquea a *pfhrp3* en específico pueda tener algún efecto sobre el parásito (57,58).

2.8.2. Diferencias a nivel fenotípico en parásitos con la delección de *pfhrp2* y *pfhrp3*

Entre los rasgos genéticos que caracterizan a los parásitos con la delección del gen *pfhrp2* está la presencia de mutaciones de resistencia a fármacos antimaláricos.

Las mutaciones de resistencia descritas en este grupo de parásitos se ubican en los genes *pfcr1*, *pfdhfr*, *pfdhps*, *pfmdr1*, ubicados en los cromosomas 5, 7, 4, y 13 (59,60). Sin embargo, estas mutaciones han sido descritas igualmente en parásitos sin la delección, por lo que no es un rasgo exclusivo.

Por otro lado, recientemente se ha reportado en estudios *in vitro* que el costo de *fitness* de portar las delecciones de *pfhrp2* y *pfhrp3* es alto para el parásito en comparación con los silvestres (61), esto a pesar que no se han descrito a cabalidad las funciones biológicas en las que participa la proteína. Igualmente, distintos estudios han descrito algunas características de *P. falciparum* cuando no presenta el gen *pfhrp2*. De este modo, se ha reportado que la cepa Dd2, portadora de la delección, tiene un mecanismo de invasión eritrocítica dependiente de ácido siálico (62). Además, se ha sugerido que los parásitos dependientes de ácido siálico podrían modificar su mecanismo de invasión por uno independiente, en condiciones de agitación del cultivo (denominado por los autores como cultivo en suspensión) (62). El perfil de expresión en estas condiciones del estudio muestra una sobre-expresión de genes relacionado con la invasión de eritrocitos (proteínas de unión a eritrocitos, proteínas de exportación, proteínas homólogas de unión a eritrocitos) (62,63). Se ha postulado que las condiciones en “suspensión” guardan similitud con las condiciones en la etapa sanguínea dentro del hospedero.

Por otro lado, Yang y colaboradores mediante la tecnología CRISPR-Cas9 alteraron la expresión de *pfhrp2* en la cepa 3D7 al eliminar su segundo exón. La disrupción del gen *pfhrp2* tendría como consecuencia la sobre-expresión de genes mitocondriales que participan en el metabolismo del grupo hemo, lo que compensaría la pérdida de la vía metabólica de hemo mediada por HRP2 (45).

A pesar de que los estudios descritos fueron realizados en cultivos *in vitro* y que se desconoce si los hallazgos ocurren de manera similar *in vivo*, se pueden considerar indicios de que existen diferencias entre los parásitos con y sin la delección de *pfhrp2* y *pfhrp3*.

III. PLANTEAMIENTO DE LA INVESTIGACIÓN

Los casos de malaria en Perú han disminuido en los últimos años. Al 2022 se registraron un total de 24701, concentrándose ~90% de los casos en la región Loreto, de los cuales alrededor del 16% fueron causados por *P. falciparum* causante de la forma severa de malaria (64). Uno de los factores que contribuyen a la prevalencia de estos parásitos es la presencia de mecanismos de resistencia que son seleccionados positivamente (16,65). Por otro lado, en los últimos años los parásitos *P. falciparum* que se reportan en mayor frecuencia tienen la delección del gen *pfhrp2* y de su homólogo *pfhrp3* (5,6). Sin embargo, se desconocen los mecanismos que han permitido su incremento.

La delección de los genes *pfhrp2* y *pfhrp3* fue inicialmente reportada en muestras de la selva peruana. En Perú se ha reportado la delección de *pfhrp2* entre 20 y 53% de muestras evaluadas, los estudios realizados con muestras colectadas en los periodos 1998-2001, 2003-2005 y 2012-2014 obtenidas de distintas localidades de Loreto muestran un aumento en la frecuencia de la delección del gen *pfhrp2* (20%, 40% y 53% en cada periodo) (5,6).

Las delecciones se han reportado también en otras regiones de Sudamérica como los estados de Acre, Rondonia y Pará en Brasil (10,11), la Amazonía de Colombia (9) y en Surinam (47). También se ha reportado la delección de *pfhrp2* en diferentes países de África en una frecuencia que oscila desde 0.4% en Angola, hasta 62% en Eritrea (12,49–51,66). Sin embargo, se desconoce qué factores contribuyeron a la aparición y expansión de estos parásitos. En África se ha postulado que el uso de PDRs puede haber conducido a la expansión de estos parásitos, sin embargo no hay evidencia previa de un efecto similar por un método diagnóstico (21). En el caso de

Perú el uso de PDRs en el diagnóstico de malaria es menos frecuente que la microscopía, por lo cual no se podría sugerir una relación similar (5).

La presión de selección en *P. falciparum* ocurre por diversos factores. Los genes asociados a resistencia *pfcr1*, *pfmdr1*, *pfdhfr* y *pfdhps* muestran evidencia de haber sufrido presión de selección por el tratamiento antimalárico en África y en otras regiones (15,40,41). De igual forma, el sistema inmunitario del hospedero puede ejercer presión de selección (47). Se ha reportado barridos selectivos en genes codificantes de antígenos, candidatos a vacunas, proteínas que participan en la interacción con el eritrocito y en genes involucrados en el desarrollo de gametocitos, y en consecuencia en la transmisión (15,65,67). Este tipo de análisis no se ha realizado en *P. falciparum* en Perú. Además, los estudios de barridos selectivos no reportan si la presencia de la delección de los genes *pfhrp2* y *pfhrp3* se relaciona con alguna diferencia en el genoma debido a selección.

P. falciparum con y sin la delección tienen similitudes y diferencias entre sí a nivel genético y fenotípico. Los parásitos *P. falciparum* tienen mutaciones de resistencia a fármacos antimaláricos, lo cual podría explicar su aumento en frecuencia; sin embargo, es un rasgo compartido entre ambos grupos de parásitos con o sin la delección (60). Análisis de estructuración con microsatélites han agrupado a los parásitos con la delección junto con parásitos no delecionados en más de dos clusters, sugiriendo un origen múltiple (5,47), pero también se ha encontrado que los parásitos con la delección forman un cluster separado de los parásitos sin la delección (6,9). A nivel fenotípico, se han identificado diferencias en la expresión de genes relacionados a invasión o metabolismo del grupo hemo a partir de estudios *in vitro* con la cepa Dd2 (62,63) o con la cepa 3D7 sin el gen *pfhrp2* obtenida mediante

CRISPR-Cas9 (45). De este modo no se puede descartar que *in vivo* los parásitos *Plasmodium* con la deleción se distinguen de aquellos sin la deleción lo que podría o no reflejarse en diferencias genómicas por selección.

En ese sentido, en el presente estudio se propone la búsqueda de regiones genómicas que muestren evidencia de haber sufrido selección positiva reciente o estar en proceso de selección, en parásitos *P. falciparum* de la selva peruana. Para este análisis se abordarán dos aproximaciones, la primera partirá del total de parásitos *P. falciparum*. En la segunda aproximación se evaluarán por separado al grupo de parásitos con la deleción de *pfhrp2* y *pfhrp3*, y al grupo de parásitos sin la deleción. De esta forma, se determinarán los barridos selectivos presente en la muestra evaluada de *Plasmodium falciparum*, asimismo se identificarán los barridos selectivos específicos de los parásitos que tienen la deleción de *pfhrp2* y *pfhrp3* que podrían explicar su aumento en frecuencia.

IV. JUSTIFICACIÓN

La malaria sigue representando un problema de salud de importancia en Perú y el mundo. A pesar de representar solo alrededor del 20% de casos en Perú, la malaria por *Plasmodium falciparum* es un reto en su control y erradicación. Durante los años 2012 y 2015 se reportó un incremento en los casos de malaria en la Amazonía peruana (24). Asimismo, se reportaron dos brotes de malaria por *P. falciparum*, el primero entre los años 2010-2012 en Tumbes, y el segundo el 2013 en Cusco (59,60). En ambos casos los parásitos portaban la delección de *pfhrp2*.

No se conocen que mecanismos biológicos se están seleccionando positivamente y contribuyen a la persistencia de *P. falciparum* en la selva peruana. Se sabe que factores sociodemográficos como la movilidad por diversas actividades, el tipo de vivienda, entre otros, favorecen la transmisión de la malaria en la selva peruana (18). Por otro lado, los estudios de genética de poblaciones permiten ver el efecto de las estrategias de control sobre el parásito. Es así que la diversidad genética de *P. falciparum* es baja a moderada, su reproducción es clonal y la multiplicidad de infección reportada es baja (menor al 8%) (6); sin embargo, no se ha logrado su erradicación. Identificar las regiones del genoma del parásito bajo selección positiva puede explicar los mecanismos que favorecen la permanencia de *P. falciparum*, y también el aumento en frecuencia de los parásitos portadores de las delecciones de los genes *pfhrp2* y *pfhrp3*.

En resumen, los resultados de este estudio contribuirán al entendimiento de la genética de *P. falciparum*, y los mecanismos biológicos que favorecen su transmisión. Asimismo, la información generada podría contribuir a la explicación del aumento en frecuencia de los parásitos con la delección de los genes *pfhrp2* y

pfhrp3. Esta información puede ser de relevancia en la vigilancia molecular de la malaria por *P. falciparum* en la Amazonía peruana.

V. OBJETIVOS

5.1. OBJETIVO GENERAL

Identificar las regiones del genoma de *Plasmodium falciparum* de la selva peruana que presenten evidencia de haber sufrido selección positiva reciente o que estén en proceso de selección.

5.2. OBJETIVOS ESPECÍFICOS

- Estimar la frecuencia de las deleciones de los genes *pfhrp2* y *pfhrp3* en una muestra de parásitos *P. falciparum* colectadas durante los años 2013 – 2017 en dos centros de salud de Loreto.
- Determinar la estructura poblacional de los parásitos *P. falciparum* a partir de información de genoma completo en muestras colectadas entre 2006 – 2018
- Detectar barridos selectivos recientes o en proceso en el genoma de *P. falciparum* de la selva peruana mediante dos aproximaciones para el análisis.
- Determinar si el aumento de *P. falciparum* con la deleción de los genes *pfhrp2* y *pfhrp3* es producto de procesos de presión de selección específicos.

VI. METODOLOGÍA

6.1. DISEÑO DE INVESTIGACIÓN

Este es un estudio de tipo observacional descriptivo. Se realizó la caracterización genética de una muestra de parásitos de la especie *Plasmodium falciparum*. Se evaluaron muestras colectadas en el periodo 2006-2018 en la región Loreto, con el fin de identificar barridos selectivos a partir de datos de genoma completo.

El estudio fue ejecutado en dos etapas:

- La primera etapa consistió en la caracterización de la delección de los genes *pfhrp2* y *pfhrp3* y sus genes flanqueantes. La caracterización se llevó a cabo en 206 muestras colectadas durante el periodo 2013-2017 en Loreto, disponibles a partir de estudios previos del laboratorio de malaria de UPCH y de colaboradores del Departamento de Parasitología del NAMRU-6, como se describe en la sección 6.2.2. Población de estudio.
- La segunda etapa comprendió el análisis de datos de genoma completo. Se secuenciaron 54 muestras caracterizadas en la primera etapa, y a estas muestras se adicionaron genomas almacenados en el repositorio de datos del laboratorio de Malaria-UPCH (n=13), genomas secuenciados por la unidad de malaria del Instituto de Medicina Tropical de Amberes (ITM) (n=17), y genomas depositados en bases de datos públicas secuenciados por el consorcio MalariaGEN (n=38), y por colaboradores de NAMRU-6 (n=24). Los análisis de genética de poblaciones se realizaron luego de aplicar los filtros de calidad de muestras y SNPs, conservándose 59 genomas de muestras con la delección del gen *pfhrp2* y 41 muestras sin la

delección (incluidas 41 de las 54 muestras secuenciadas en esta tesis). Las muestras contaron con información de la delección del gen *pfhrp2*, región de procedencia y año de colecta (2006-2018).

Se propusieron dos aproximaciones para la búsqueda de barridos selectivos mediante escaneo del genoma completo del parásito.

1. En la primera aproximación se propuso el análisis del total de parásitos *P. falciparum*, con el fin de detectar barridos selectivos característicos de la especie presente en la Amazonía peruana.
2. En la segunda aproximación se propuso el análisis de los parásitos agrupados en base a la presencia o ausencia de la delección de *pfhrp2/3*, a fin de identificar barridos selectivos específicos de cada grupo de parásitos definido.

La segunda aproximación fue la que se ejecutó debido a las características de la muestra evaluada.

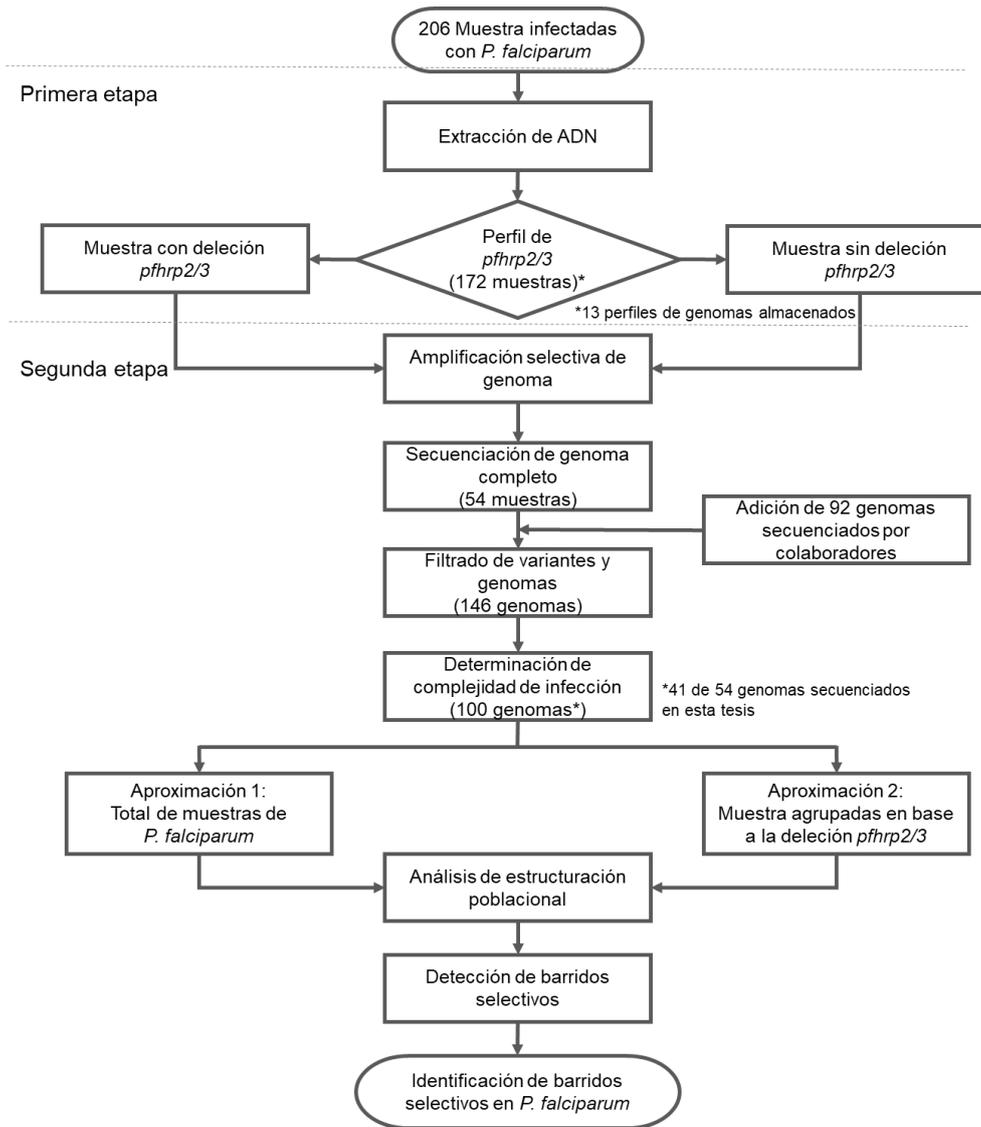


Figura 4. Flujograma de trabajo.

6.2. POBLACIÓN DE ESTUDIO

6.2.1. Área de estudio

Las muestras incluidas fueron colectadas en los años 2006 – 2018, en las provincias de Loreto, Maynas, Ramón Castilla y Datem del Marañón. La colecta pasiva de las muestras fue en el centro de salud de Padre Cocha (3° 41' 54" S, 73° 16' 43" W) y en el Hospital de apoyo de Iquitos (3° 45' 36" S, 73° 15' 07" W), y comprendieron muestras provenientes de los distritos de Alto Nanay, Iquitos, Mazan, Morona, Punchana, Tigre, Torres Causana, Trompeteros y Yavari. Las muestras colectadas activamente tenían como origen los distritos de Iquitos, San Juan y Nauta.

6.2.2. Población de estudio

Las muestras seleccionadas tienen confirmada la infección por *P. falciparum*, y se obtuvieron previa firma de un consentimiento informado. Las muestras se colectaron como parte de un estudio de vigilancia pasiva (Protocolo NMRC. 2007.0004, Enfermedad de la Malaria en Perú) por el Departamento de Parasitología del NAMRU-6. Las muestras fueron caracterizadas para la delección de los genes *pfhrp2/pfhrp3*. A partir de estas muestras se seleccionaron aquellas con la delección de *pfhrp2/pfhrp3* y muestras sin las delecciones, para la secuenciación de genoma completo.

El laboratorio de Malaria de la Universidad Peruana Cayetano Heredia (UPCH) tenía la información genómica de 10 muestras de parásitos portadores de la delección *pfhrp2/pfhrp3*. Asimismo, se contó con la información genómica de 3 muestras sin la delección. Estas muestras fueron colectadas de forma activa por el proyecto de investigación “Impacto de los portadores asintomáticos en la epidemiología y

control de la malaria en la Amazonía Peruana” (Código SIDISI: 57395), del Laboratorio de Malaria de UPCH.

Las secuencias genómicas obtenidas de bases de datos públicos, y por colaboradores, fueron procedentes de los distritos de Iquitos, Lagunas, Mazan, Napo, Punchana, San Juan y Urarinas.

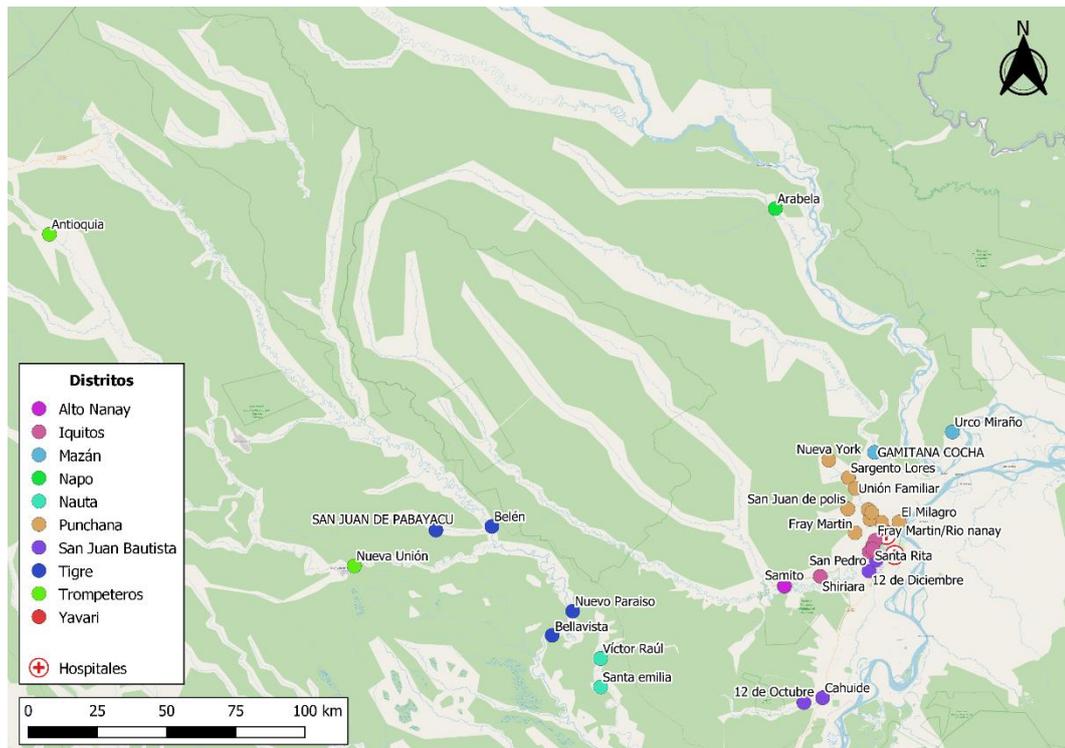


Figura 5. Mapa de procedencia de las muestras incluidas en el estudio.

6.2.3. Muestras Biológicas

Unidad de análisis: Cromosoma

Unidad de muestreo: Muestra infectada con *Plasmodium falciparum* de acuerdo a los criterios de inclusión.

Este es un estudio de tipo retrospectivo descriptivo, por lo cual no se llevó a cabo el cálculo del tamaño muestral.

6.2.3.1. Criterios de inclusión

- Muestras positivas a *P. falciparum* confirmado por microscopía o PCR.
- Muestras de *P. falciparum* con una parasitemia no menor a 200 parásitos por uL de sangre.
- Muestras obtenidas de individuos que hayan firmado el consentimiento informado.
- Muestras de *P. falciparum* con suficiente cantidad de ADN para la secuenciación genómica.

6.2.3.2. Criterios de exclusión

- Muestras con infección mixta por *P. vivax*.
- Muestras de *P. falciparum* con parasitemia menor a 200 parásitos por uL de sangre.
- Muestras de ADN de *P. falciparum* degradadas.

6.2.4. Operacionalización de Variables

6.2.4.1. Variables independientes

- Año de recolección de las muestras
- Deleción de los genes *pfhrp2* y *pfhrp3*

6.2.4.2. Variables dependientes

- Fst: indica la diversidad genética entre poblaciones. Es el grado de diferenciación genética entre pares de poblaciones. Su valor oscila entre 0 y 1. Poblaciones altamente diferenciadas presentan valores cercanos a 1; mientras que poblaciones semejantes tienen valores cercanos a 0.
- Desequilibrio de ligamiento (LD): asociación no aleatoria entre los alelos de marcadores ligados.

- Haplotipos largos: regiones de LD mucho más largas de lo esperado dada la frecuencia observada de un alelo. Los alelos de alta frecuencia son en promedio alelos más antiguos bajo neutralidad, por lo tanto, han experimentado más recombinación, lo que resulta en haplotipos más cortos. En contraste a alelos recientes que se seleccionan positivamente y aumentan en frecuencia rápidamente.
- Cluster: número de subpoblaciones de *P. falciparum* determinado por los marcadores SNPs en este estudio.

6.3. PROCEDIMIENTOS Y TÉCNICAS

6.3.1. Caracterización de las deleciones de los genes *pfhrp2*, *pfhrp3* y genes flanqueantes PF3D7_0831700, PF3D7_0831900, PF3D7_1372100 y PF3D7_1372400

6.3.1.1. Extracción de ADN

La extracción de ADN se realizó a partir de sangre total o de papel filtro impregnado con sangre. Se empleó el protocolo de extracción del laboratorio de malaria de UPOCH, adaptado a partir del manual del kit de extracción E.Z.N.A blood DNA mini kit (68). El protocolo de extracción se basa en el uso de columnas de sílica. El protocolo adaptado consistió en la adición de 40 μ L de sangre total o 72mm² de papel filtro impregnado con sangre en un microtubo de 2 mL.

A la muestra se le adicionó 250 μ L de buffer de lisis TEN 2X (Tris HCl 1M pH 8.0, NaCl 5M, EDTA 0.5 M pH 8.0). Se incubó la mezcla a 65°C por una hora en agitación constante de 450 rpm. Concluida la incubación se adicionó 25 μ L de proteasa K y se incubó por treinta minutos adicionales en agitación constante. Paso seguido, se agregó 250 μ L de buffer de lisis (BL) y de etanol absoluto; luego de agitar en vortex, se transfirió la mezcla a las mini columnas de sílica y se procedió a centrifugar a 8000 rpm por 1 min. Posteriormente se adicionó el buffer HBC y se centrifugó a 8000 rpm por 1min. El lavado de las muestras fue con el buffer correspondiente, dos veces, y se eliminaron residuos líquidos mediante centrifugación a 12000 rpm por 2 minutos. Finalmente, se resuspendió el ADN en 50 μ L de buffer de elución previa incubación a 65°C por 5 minutos y centrifugación a 12000 rpm por 1 minuto.

El ADN se almacenó a -20°C para su posterior procesamiento.

6.3.1.2. Detección de las deleciones de los genes de interés mediante PCR convencional

El protocolo de detección de las deleciones incluye la amplificación por PCR convencional del exón 1 de los genes *msp1*, *msp2* y *glurp*, en base al protocolo de laboratorio adaptado a partir de reportes previos (69). Este paso sirve como control de calidad de la muestra de ADN. Cada mezcla de reacción comprendió buffer de reacción 1X, MgCl₂ 2.5 mM, dNTPs 0.2 mM, 1U de enzima polimerasa “hot start”, Taq AccuStart II (Quantabio) o platinum Taq DNA polimerasa (Invitrogen). La concentración de cada cebador para la amplificación del gen *msp1* fue de 0.24 μM, y de 0.31 μM para los genes *msp2* y *glurp*. A la mezcla de reacción se adicionó 2.5 μL de ADN. El volumen final por reacción fue de 12.5 μL.

Las muestras que amplificaron los genes control fueron evaluadas para detectar la deleción de los genes *pfhrp2*, *pfhrp3* y sus genes flanqueantes, empleando el protocolo de laboratorio adaptado (6). Se realizó la amplificación de la secuencia del exón 2 de los genes *pfhrp2*, *pfhrp3* y de los genes flanqueantes. La mezcla de reacción de cada protocolo incluyó buffer de reacción 1X, MgCl₂ 1.5 mM, dNTPs 0.2 mM, 1U de enzima polimerasa “hot start”, Taq AccuStart II (Quantabio) o platinum Taq DNA polimerasa (Invitrogen). La concentración de cada cebador para la amplificación de cada gen fue la siguiente: para el gen *pfhrp2* fue de 0.2 μM, para *pfhrp3* fue de 0.32 μM, en el caso de PF3D7_0831700 fue 0.28 μM, de PF3D7_0831900 0.2 μM, de los genes PF3D7_1372100 y para PF3D7_1372400

fue de 0.16 μ M. A cada reacción se agregó 3 μ L de ADN. El volumen final de cada reacción fue de 25 μ L.

La amplificación se llevó a cabo en un termociclador T100™ Thermal Cycler de la marca Bio-Rad. La visualización de los productos de PCR fue mediante electroforesis en gel de agarosa al 2% y TBE 0.5 X (Tris-borato 0.045 M y EDTA 0.001 M), a un voltaje de 90V min por 60-90 minutos. En cada corrida de electroforesis se incluyó un marcador de peso molecular de 100 pb (Thermo Scientific). Los geles fueron revelados en un fotodocumentador Gel Doc™ XR+System Bio-Rad. Las temperaturas y ciclos de amplificación se resumen en Anexo 2.

Los procedimientos descritos fueron ejecutados en presencia de muestras control. Se utilizó una muestra de la cepa 3D7 como control positivo para todos los genes, la cepa Dd2 fue empleada como control negativo del gen *pfhrp2*, la cepa HB3 fue incluida como control negativo del gen *pfhrp3*, y una muestra no infectada como control negativo para *Plasmodium falciparum*, además de una reacción sin ADN como blanco. En parte de los ensayos se usó una muestra previamente reportada como portadora de la delección de los genes *pfhrp2* y *pfhrp3* como control negativo del gen *pfhrp3*.

6.3.2. Secuenciación de genoma completo (WGS) de *Plasmodium falciparum*

Se seleccionaron dos grupos de muestras portadoras y no portadoras de las delecciones de los genes *pfhrp2* y *pfhrp3*, con la finalidad de llevar a cabo la comparación entre ambos grupos de parásitos presentes en la población de *P.*

falciparum de estudio. Treinta muestras con la delección de los genes *pfhrp2* y *pfhrp3* (*pfhrp2/3* (-/-)) y 24 sin la delección (*pfhrp2/3* (+/+)), fueron seleccionadas para secuenciación de genoma completo (WGS).

6.3.2.1. Protocolo Base de Secuenciación de Genoma Completo

6.3.2.1.1. PCR cuantitativa de muestras

Las muestras fueron cuantificadas empleando los cebadores propuestos por Mangold y colaboradores (70) para la detección y cuantificación del número de copias del gen ARN ribosomal 18S, en una reacción con el colorante fluorescente Sybr Green. Este procedimiento fue empleado para la selección de las muestras que se secuenciaron en base a su parasitemia medida en parásitos por μL . Se preparó la mezcla de reacción con una concentración 1X de PerfeCTa SYBR Green FastMix (QuantaBio), 0.3 μM de cebador *forward* y *reverse*, se utilizó 5 μL de ADN, y se completó con H_2O hasta un volumen final de 25 μL . La amplificación se realizó en el equipo termociclador CFX96TM Bio-Rad. El programa de amplificación consistió de la denaturación inicial a 95°C por 2 minutos, y 45 ciclos que consistieron en denaturación a 95°C por 20 segundos, hibridación a 52°C por 20 segundos y elongación a 68°C por 30 segundos; luego del ciclado continúa la elongación final a 68°C por 3 minutos. El protocolo incluye una curva de variación de temperatura de “melting”, posterior a la amplificación, que consiste en un aumento escalonado de la temperatura desde 65 a 85°C en incrementos de 0.5°C cada 5 segundos.

6.3.2.1.2. Amplificación Selectiva de Genoma Completo (sWGA)

El protocolo de sWGA fue una adaptación del propuesto por Sundararaman y colaboradores (71). De acuerdo al trabajo de referencia, la amplificación se realiza

en dos rondas; cada ronda con un set de 10 cebadores diferentes (set de cebadores 6A y 8A), que se unen de manera aleatoria al genoma de *P. falciparum*.

6.3.2.1.2.1. sWGA de una ronda

Para la estandarización, el protocolo sWGA fue ejecutado inicialmente con solamente una ronda, similar a lo desarrollado por Mavila, A. (72). La mezcla de reacción de la ronda de amplificación consistió de buffer de reacción 1X, Seroalbúmina bovina (BSA) 0.1 mg/mL, dNTPs 1 mM y 3.5 μ M del set de cebadores, 30 U de la enzima phi29 (New England Biolabs) y 1-20 ng de ADN genómico cuantificado por kit Qubit™ dsDNA HS Assay (Invitrogen). El volumen final de reacción fue de 50 μ L.

La amplificación se llevó a cabo en un termociclador T100™ Thermal Cycler de la marca Bio-Rad. El programa de amplificación consistió de 10 minutos a cada uno de las siguientes temperaturas, 35°C, 34°C, 33°C, 32°C, 31°C, y 16h a 30°C, con un paso final de inactivación de la enzima a 65°C por 10 minutos.

6.3.2.1.2.2. sWGA de dos rondas

Luego de la prueba de amplificación genómica con solo una ronda, se utilizó el protocolo descrito por Sundararaman y se incluyó una segunda ronda de amplificación (71). La primera ronda de amplificación se ejecutó sin ningún cambio. Para la segunda ronda se preparó una mezcla de reacción con las mismas concentraciones de cada componente. A diferencia de la primera ronda de sWGA, se usó como molde 4 μ L del producto de la primera ronda de amplificación, y se ajustó el volumen de agua. El programa de amplificación fue el mismo que el de la primera ronda.

6.3.2.1.2.3. Lavado de productos de sWGA

El producto de cada ronda de sWGA fue lavado usando *beads* magnéticos del reactivo Ampure XP (Beckman Coulter). Para esto, se adicionó a cada pocillo con producto un volumen de *beads* igual al volumen de producto, se mezcló por pipeteo y mediante una centrifugación breve se recolectaron las gotas que hubieran quedado en las paredes o tapas de cada pocillo. La mezcla se dejó incubar a temperatura ambiente por 5 minutos, para posteriormente colocar la placa con muestras en un soporte magnético (Invitrogen) por 5 minutos o hasta que la solución se aclare. Se descartó el sobrenadante, y a cada pocillo se le agregó 150 μL de etanol al 70% preparado en el momento y se dejó por un minuto. Se eliminó el alcohol y se repitió el paso. Los residuos de alcohol se eliminaron con la ayuda de una pipeta de 10 μL , para luego dejar secar por 2 minutos, sin retirar del soporte magnético. Se retiró del soporte magnético y se adicionó 30 μL de buffer TE *low* o *buffer* de elución del kit de extracción de ADN, se homogenizó por pipeteo y se centrifugó brevemente para recuperar cualquier gota que quedara en las tapas o paredes de los pocillos, y se dejó incubar por 5 minutos. Luego, la placa fue colocada nuevamente en el soporte magnético hasta que la solución se aclaró. Finalmente, se transfirieron 27 μL de sobrenadante a una nueva placa o *strip* de 8 pocillos. El producto recuperado se almacenó a -20°C .

Posterior a cada ronda de amplificación se realizó una PCR cuantitativa para evaluar el rendimiento de la sWGA en las muestras. Se siguió el protocolo descrito previamente. Se empleó como ADN molde una dilución 1:10 del producto de la ronda de sWGA respectiva.

6.3.2.1.3. Secuenciación de Genoma Completo (WGS)

El protocolo empleado fue el del kit Nextera XT DNA Library prep de Illumina (73). Dependiendo del kit de secuenciación, se prepararon entre 8 y 12 muestras. Para las ejecuciones de WGS, las muestras se organizaron como se muestra en el esquema de la Figura 6.

El protocolo de preparación de bibliotecas utilizado en esta tesis se detalla en el Anexo 3.

	1	2	3	4	5	6	7	8	9	10	11	12
A	M3784 N502 N701	M3797 N502 N702	M4077 N502 N703	M4315 N502 N704	M4317 N502 N705	M4354 N502 N706						
B	M4805 N517 N701	M5048 N517 N702	V44490 N517 N703	V44766 N517 N704	V44947 N517 N705	V45099 N517 N706						
C												
D												
E												
F												
G												
H												

Figura 6. Esquema de placa para preparación de bibliotecas. Se indica en color rojo la combinación de indexes por muestra.

6.3.2.2. Modificaciones en protocolo de WGS

Se llevaron a cabo cinco corridas de secuenciación de genoma completo, los detalles de cada corrida se indican en el Anexo 4. Las principales diferencias en los ensayos realizados se resumen en la tabla 2.

Tabla 2. Tabla resumen de corridas de Secuenciación

Corrida	Número de muestras	Kit de Secuenciación	Carga de bibliotecas
1	8	MiSeq v2	8 pM
2	12	MiSeq v3	8 pM
3	10	MiSeq v3	8 pM
4	12	MiSeq v3	10 pM
5	12	MiSeq v3	10 pM

6.3.2.3. Protocolo Final de Secuenciación de Genoma Completo

Se evaluó la integridad de las muestras a ser secuenciadas mediante electroforesis en un gel de agarosa al 0.7% a 80V por 40 minutos.

Para el enriquecimiento de las muestras se utilizó el protocolo de sWGA de dos rondas descrito líneas arriba. Se evaluó la presencia de contaminantes en los productos de sWGA, mediante el ratio 260/230 obtenido por espectrofotometría usando el equipo nanodrop. Un valor > 2 indica ausencia de contaminantes orgánicos. En presencia de contaminantes se procedió con un paso adicional de lavado con los *beads* Ampure XP.

Los productos finales de sWGA se procesaron siguiendo las indicaciones del fabricante, descrito en el Anexo 3. Se consideró que de acuerdo a la capacidad del kit MiSeq v2, se pueden secuenciar 8 muestras y la concentración final del pool de bibliotecas es de 8pM. En el caso del kit MiSeq v3 se pueden secuenciar 12 muestras y la carga final del pool de bibliotecas es de 10pM.

6.4. CONSIDERACIONES ÉTICAS

Las muestras incluidas para este estudio fueron colectadas previamente y cuentan con aprobación del Comité de Ética del NAMRU-6 (Número de protocolo NMRC.D.2007.0004) y de la Universidad Peruana Cayetano Heredia (Código SIDISI 57395). Todas las muestras que se han procesado en este estudio fueron colectadas luego de la aprobación y firma de un consentimiento informado de acuerdo a las Normas del Ética y Bioseguridad. Todos los procedimientos siguen los lineamientos establecidos en la Declaración de Helsinki.

6.5. PLAN DE ANÁLISIS

6.5.1. Primera etapa: Estimación de la frecuencia de las deleciones

La primera etapa buscó caracterizar una muestra de parásitos *P. falciparum* colectados en la última década; en este sentido se evaluó la frecuencia de las deleciones de los genes *pfhrp2*, *pfhrp3* y sus genes flanqueantes, y si la frecuencia de las deleciones de *pfhrp2/3* habría aumentado en comparación a datos previos.

Para el cálculo de la frecuencia de las deleciones solo se consideraron las muestras que fueron procesadas en el laboratorio de malaria (n=219). El cálculo de la proporción de la deleción se realizó dividiendo el número de muestras con la deleción, con respecto al total de muestras en las cuales la amplificación de los genes control fue exitosa. La comparación de la frecuencia de las deleciones de los genes *pfhrp2/3* se hizo incluyendo datos de un estudio previo (6), y mediante una prueba exacta de Fischer.

Luego de estimar la frecuencia de las deleciones de los genes *pfhrp2/3* y sus flanqueantes, se seleccionaron las muestras para secuenciación de genoma completo (WGS). La selección posterior a la estimación de la frecuencia de las deleciones se hizo para incluir la mayor cantidad posible de parásitos sin la deleción de los genes *pfhrp2/3*, a fin de realizar la comparación de los grupos de parásitos con y sin las deleciones de *pfhrp2/3*. Esto se tomó en consideración dado que en los años más recientes la doble deleción *pfhrp2/3* habría seguido aumentando.

6.5.2. Segunda etapa: Identificación de barridos selectivos a partir de datos de genoma completo

La segunda etapa de esta tesis tuvo como finalidad caracterizar la genética poblacional de una muestra de *P. falciparum* colectada en un periodo de tiempo mayor y evaluar su cambio en el tiempo. Considerando el aumento en frecuencia de parásitos con la delección del gen *pfhrp2*, la delección fue incluida como una variable en los análisis de estructuración. A partir de los análisis de estructuración poblacional se agruparon a los parásitos en base a la presencia de la delección. Esto último se consideró para los análisis de búsqueda de barridos selectivos, comparando los grupos de parásitos con y sin la delección de los genes *pfhrp2/3* identificados previamente. Teniendo en cuenta que inicialmente se propusieron dos aproximaciones para el análisis de genética de poblaciones de las muestras (Figura 4), se terminó ejecutando la aproximación 2, lo cual se muestra en la sección de resultados.

Para esta segunda etapa, los datos de genoma obtenidos en esta tesis fueron analizados en conjunto con secuencias genómicas disponibles por colaboraciones. A los 54 genomas obtenidos en esta tesis, se le adicionaron los 13 genomas disponibles en la base de datos del laboratorio de malaria, 17 genomas secuenciados por el ITM, 24 genomas secuenciados por NAMRU-6 y 38 genomas secuenciados por el consorcio MalariaGEN, de los cuales 25 fueron proporcionados por el laboratorio de malaria-UPCH; los genomas secuenciados por NAMRU-6 y MalariaGEN fueron descargados de la base de datos pública ENA (<https://www.ebi.ac.uk/ena/browser/>). En total, inicialmente se contaron con 146 secuencias de genoma completo de *P. falciparum*.

6.5.2.1. Procesamiento de datos crudos de secuenciación y generación de archivo de variantes (VCF)

La calidad de los archivos fastq fue evaluada con el programa FastQC (74). Con el programa FastQ Screen se realizó la verificación del porcentaje de genomas secuenciados en cada muestra, el programa selecciona ~100000 *reads* y los alinea con los genomas de *P. falciparum* y otras especies de *Plasmodium*, además del genoma humano. Se creó un reporte resumen usando el programa multiQC (75).

Los *reads* pasaron por un filtro de calidad. Se usó el programa BBduk (76) para eliminar cualquier secuencia del genoma PhiX. Luego con el programa Trimmomatic (77) se eliminaron las secuencias de los adaptadores y se estableció como umbral mínimo un promedio de calidad por base de 20, y una longitud mínima de 50 pb. Se verificó la calidad *post-trimming* de los *reads* con FastQC y se generó un nuevo reporte resumen con los resultados de *trimming* y FastQC *post-trimming*.

Las *reads* que pasaron el control de calidad fueron mapeados al genoma de referencia de *Plasmodium falciparum* versión 3, descargado de la base de datos de PlasmoDB (<https://plasmodb.org>). Se usó el programa Burrows-Wheeler Aligner (bwa) con su función mem (78). Se generaron las estadísticas del mapeo con el programa Picardtools (79). Adicionalmente, se estimó la cobertura promedio con el programa Bedtools (80), para posteriormente estimar la cobertura en los genes *pfhrp2* y *pfhrp3* con un script *in-house* de Python.

Se realizó el llamado de variantes con el programa GATK4 para generar archivos gvcfs, usando la herramienta HaplotypeCaller (81,82). La información de genotipos

de los archivos gvcfs de cada muestra se juntaron en un archivo de variantes (vcf) con la información de genotipos de todas las muestras. Se usó la herramienta GenotypeGVCFs.

A partir del archivo vcf con todas las muestras, se continuó con los pasos de filtrado de variantes. Se generó un nuevo archivo seleccionando solo la información de polimorfismos de nucleótido simple (SNP). Luego se realizó el *Hard Filtering* siguiendo los parámetros sugeridos en el manual de GATK4 (83). Los SNPs se filtraron manteniendo aquellas con un mínimo de calidad de profundidad de 2 ($QD < 2$), un mínimo de calidad de mapeo de 40 ($MQ < 40$), Calidad mínima de 30 ($QUAL < 30$).

El archivo vcf con SNPs que pasó el primer filtrado fue reducido, limitando a determinadas regiones del genoma del parásito. Se excluyeron las variantes ubicadas en el genoma mitocondrial y del apicoplasto. Asimismo, se limitó a las variantes ubicadas en la región “*core*” del genoma parasitario, excluyendo regiones subteloméricas, según lo sugerido en (84). Adicionalmente, se eliminaron 6 muestras que se encontraban duplicadas y otras 6 muestras que fueron colectadas en Tumbes, según lo indicado en el artículo por el cual se secuenciaron.

El paso siguiente consistió en el filtrado de las variantes en base a parámetros de cada variante y de las muestras analizadas. Se usó el programa VCFtools (85) para obtener los valores de profundidad, de calidad, y frecuencias alélicas por SNP; también los valores de cobertura promedio, y porcentajes de datos faltantes por SNPs y por muestra. Se utilizó un script *in-house* de R para la visualización de la

distribución de estos parámetros, y en base a ello se escogieron los valores mínimos para cada parámetro.

Con el programa VCFtools se seleccionaron solo los SNPs bialélicos, asimismo se filtraron aquellas variantes con cobertura promedio menor a 5, y también las variantes con más de 25% de datos faltantes. Luego, se eliminaron las muestras con más de 25% de datos faltantes.

Finalmente, se eliminó una muestra secuenciada por MalariaGEN (<https://www.malariagen.net/>), que correspondería a una posible contaminación con cepa de referencia.

El archivo VCF generado contenía un total de 100 muestras y 39164 SNPs.

6.5.2.2. Caracterización genética de la muestra de *P. falciparum*

6.5.2.2.1. Estimación de la complejidad de infección

La complejidad de infección se obtuvo para cada muestra de los 100 genomas mediante el índice de fijación intra-individuo (Fws). El cálculo de Fws se hizo con el paquete moimix de R (86), y se obtiene mediante la fórmula siguiente:

$$Fws = 1 - \frac{Hw}{Ht}$$

Donde Hw es la heterocigocidad promedio para una muestra, y Ht la heterocigocidad total de la muestra de parásitos analizada. Fws toma valores entre 0 y 1, un valor cercano a 1 (> 0.95) indica que la muestra tiene más de una clona.

Adicionalmente, se complementó el resultado de Fws con el análisis usando el programa estMOI (87). El cual estima el número probable de clonas presentes en una muestra a partir de datos de mapeo y de cobertura.

6.5.2.2.2. Análisis de estructuración poblacional

La caracterización de la estructuración de la muestra se realizó para identificar la presencia de *clusters* o agrupaciones de parásitos. La identificación de *clusters* de parásitos influyó en la aproximación de análisis a seguir (Figura 4). Si se identifican agrupaciones de parásitos, los análisis siguientes se realizarían para cada grupo por separado.

6.5.2.2.2.1. Análisis de componentes principales (PCA)

Para el análisis de estructuración poblacional se realizó un filtrado de las variantes, en base al valor de desequilibrio de ligamiento. Las variantes en desequilibrio de ligamiento, representado por un $r^2 > 0.5$, fueron filtradas usando el programa plink 1.9 (88). Las variantes que se conservaron fueron incluidas en el análisis de componentes principales con el programa plink 1.9.

La visualización de los resultados fue mediante un script *in-house* en el programa R.

6.5.2.2.2.2. Análisis de ancestría con el programa Admixture

El archivo con los genotipos obtenido luego del filtrado con plink, en formato ped, fue usado para el análisis de ancestría usando el programa Admixture (89). Se evaluaron valores de número de *clusters* desde 1 a 10, con 10000 pseudoreplicas (*bootstrap*), un valor de 10 de *cross-validation*, y 10 iteraciones.

El programa pophelper (90) se usó para la visualización de los resultados de ancestría. Para la elección del número de *cluster* más probable, se graficó en R los valores de error de *cross-validation* que se obtiene en cada análisis que el programa

Admixture realiza para cada número de *cluster* indicado (1 – 10). El menor valor de error de *cross-validation* indica el número de *clusters* más probable.

En base a los resultados obtenidos de los análisis de estructuración poblacional las muestras se agruparon en base a la presencia/ausencia de la delección de *pfhrp2/3*. Asimismo, se agruparon las muestras en base al Periodo de colecta, las muestras colectadas entre 2006-2011 correspondieron al Periodo 1, y de 2012-2018 al Periodo 2. Los análisis siguientes se llevaron a cabo en las muestras del Periodo 2 y considerando dos grupos de parásitos diferentes si los parásitos portaban o no la delección *pfhrp2/3*.

6.5.2.2.3. Análisis de decaimiento del desequilibrio de ligamiento

El análisis de decaimiento del desequilibrio de ligamiento proporciona información respecto al grado de recombinación en las muestras evaluadas. Este análisis se realizó con el programa plink 1.9, el desequilibrio de ligamiento se estimó mediante el cálculo del valor de r^2 entre pares de SNPs en una ventana de máximo 100 kb. Luego, se graficó la variación del r^2 promedio conforme aumentaba la separación en pares de bases entre SNPs.

6.5.2.3. Escaneo de genoma para detectar barridos selectivos

Para la detección de barridos selectivos se emplearon tres tipos de pruebas. El test de neutralidad de Tajima (D de Tajima) se basa en el espectro de frecuencias por posición, el índice de fijación F_{st} se basa en las diferencias en frecuencias alélicas, y las pruebas basadas en desequilibrio de ligamiento para la detección de regiones con evidencia de haplotipos extendidos.

Estas pruebas fueron aplicadas solamente a las muestras del Periodo 2 siguiendo la Aproximación 2 propuesta (Figura 4), comparando los grupos de los parásitos con y sin la delección de los genes *pfhrp2/3*.

6.5.2.3.1. Prueba de neutralidad con D de Tajima

Se utilizó el paquete *scikit-allel* para la prueba estadística D de Tajima a nivel genómico (91). Para esto, se indicó que estime el valor del D en ventanas de 500 pb sin sobreponerse. El cálculo de D de Tajima se realizó para cada grupo de parásitos por separado.

6.5.2.3.2. Análisis de diferenciación poblacional por Fst

El valor de Fst se estimó confrontando ambos grupos de parásitos. Se realizó el cálculo de Fst de Hudson, usando el programa *plink 2.0* (<https://www.cog-genomics.org/plink/2.0/>). El Fst de Hudson es recomendado en casos donde los grupos en comparación difieren en el tamaño de individuos, por lo cual su cálculo es menos proclive a errores por el tamaño de la población (92).

6.5.2.3.3. Detección de barridos selectivos mediante pruebas basadas en desequilibrio de ligamiento y haplotipos extendidos

Se utilizó el paquete *rehh* de R para las pruebas estadísticas basadas en desequilibrio de ligamiento (93,94). Se estimó el índice *iHS* para la identificación de loci candidatos a estar o haber estado bajo presión de selección. Se categorizaron como loci candidato aquellas regiones que presentasen un $-\text{Log}_{10}(\text{p-value}) \geq 5$, o aquellas que presentasen $|iHS| \geq 4$ y $-\text{Log}_{10}(\text{p-value}) \geq 4$.

Adicionalmente, se realizó el análisis XP-EHH para la identificación de regiones genómica que presenten haplotipos extendidos presentes en el grupo de parásitos

sin la delección del gen *pfhrp2* y que difieran con el grupo sin la delección. Las regiones candidatas presentaron valores positivos o negativos si era candidato a estar selección en el grupo sin la delección o con la delección, respectivamente. Asimismo, el valor de $-\text{Log}_{10}(\text{p-value}) \geq 5$ fue el mínimo para considerar el loci como significativo.

6.5.2.4. Análisis *in silico* de ontología genética de las regiones candidatas identificadas

Las regiones candidatas que presentaron valores superiores a los mínimos fueron enlistadas. Se identificaron los genes presentes contrastando las posiciones genómicas con la información presente en la base de datos PlasmoDB. La lista de genes identificados fue utilizada para el análisis de Gene Ontology (GO) Enrichment con la corrección de bonferroni, usando la herramienta de PlasmoDB. El análisis de GO brinda información acerca de las funciones biológicas, funciones moleculares o componentes celulares enriquecidos por el conjunto de genes candidatos a estar seleccionados.

VII. RESULTADOS

7.1. Caracterización de las muestras *P. falciparum* para la delección de los genes *pfhrp2*, *pfhrp3*, y genes flanqueantes PF3D7_0831700, PF3D7_0831900, PF3D7_1372100 y PF3D7_1372400

7.1.1. Procedencia de las muestras incluidas en el análisis

De 219 muestras analizadas, se obtuvo el perfil para un total de 172 muestras colectadas entre los años 2013 y 2017. Las 47 muestras que no se incluyeron en los cálculos no amplificaron los genes control, por factores como calidad de ADN, baja parasitemia o por ser muestras con infección por *P. vivax*. Asimismo, no se incluyeron las muestras que fueron obtenidas de bases de datos públicas al no tener la información completa. Las muestras analizadas provenían de las provincias de Maynas (103), Loreto (67), Datem del Marañón (1) y Ramón Castilla (1) (Tabla 3). La mayor cantidad de muestras tienen como origen los distritos de Nauta e Iquitos. De las 62 muestras de Nauta, 59 fueron colectadas en la comunidad de Santa Emilia, y el resto en Víctor Raúl y 12 de octubre. En el caso de las muestras de Iquitos, 33 fueron colectadas en el hospital de Apoyo de Iquitos, 28 en la comunidad de Lupuna, y el resto en las zonas de Amazon Hope (4), Huaturi (1) y San Pedro (2).

Tabla 3. Lugar de procedencia de las muestras analizadas para perfil de la
delección de los genes *pfhrp2* y *pfhrp3* y flanqueantes

Distrito de Origen	n
ALTO NANAY	2
IQUITOS	68
MAZAN	1
MORONA	1
NAUTA	62
PUNCHANA	17
SAN JUAN	14
TIGRE	4
TORRES CAUSANA	1
TROMPETEROS	1
YAVARI	1

7.1.2. Frecuencia de la delección de los genes *pfhrp2*, *pfhrp3* y los genes flanqueantes

La muestra de *P. falciparum* incluida en esta tesis presentó una frecuencia de 76.7% y 63.4%, para la delección del gen *pfhrp2* y del gen *pfhrp3*, respectivamente; mientras que la frecuencia de la doble delección fue de 61.6%. Asimismo, se identificó que el 15.1% de las muestras tenían deleccionado solo el gen *pfhrp2*, y el 1.7% solo tenían deleccionado el gen *pfhrp3* (Figura 7, Anexo 5a).

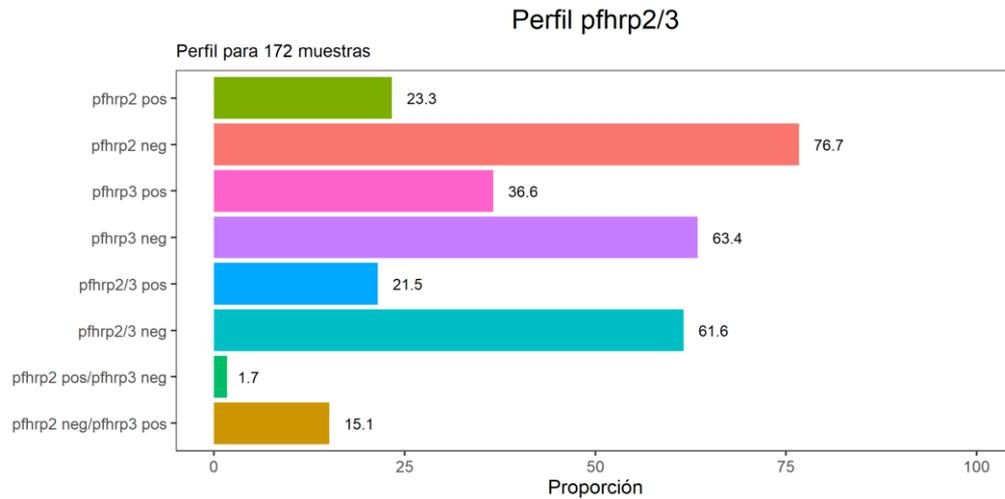


Figura 7. Perfil de las deleciones de los genes *pfhrp2* y *pfhrp3* en 172 muestras. Los porcentajes presentados se obtuvieron tomando por separado a las muestras *pfhrp2* positivas vs negativas, *pfhrp3* positivas vs negativas, y finalmente tomando ambos genes de forma simultánea (cuatro barras inferiores).

Adicionalmente, se obtuvo el perfil de las deleciones y sus respectivos genes flanqueantes. Se encontraron 6 perfiles diferentes para *pfhrp2* y sus genes flanqueantes, y 7 para *pfhrp3* y sus flanqueantes (Anexo 5b). El perfil que consiste en la deleción del gen *pfhrp2* y del gen PF3D7_0831900 fue el más frecuente de los identificados, representando el 71.3% del total. Por otro lado, el perfil conformado por la deleción de *pfhrp3* y su gen flanqueante PF3D7132100 estuvo presente en 55.6% del total de muestras con información completa. El perfil combinado de los dos perfiles descritos líneas arriba, estuvo presente en el 54.4% de las muestras, seguido en frecuencia por el perfil conformado por todos los genes presentes (21.1%); y finalmente por el perfil compuesto por la deleción del gen *pfhrp2* y del gen PF3D7_0831900, y el resto de los genes positivos (14.6%).

Al ser las provincias de Iquitos y Nauta, las que tienen mayor número de muestras en este estudio, se comparó mediante test exacto de Fisher las proporciones de la delecciones. Se observó que hay una diferencia significativa entre ambas provincias, observándose una mayor proporción de muestras *pfhrp3* positivas en Nauta.

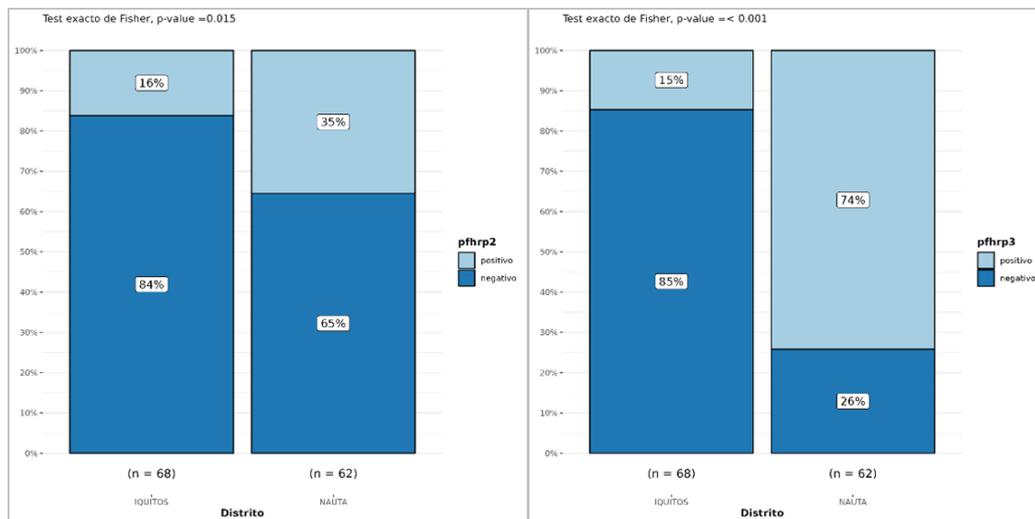


Figura 8. Comparación de las proporciones de las delecciones de los genes *pfhrp2* y *pfhrp3* entre los distritos de Iquitos y Nauta.

Por otro lado, para confirmar si la delección de los genes *pfhrp2* y *pfhrp3* habría seguido aumentando en frecuencia con el tiempo en las muestras de *P. falciparum*, se llevó a cabo la comparación de las proporciones por test exacto de Fisher, de las muestras colectadas en el periodo 2012 – 2014 (incluyendo los resultados de trabajo de Quispe y este estudio) y las muestras más recientes incluidas en este estudio (2015 – 2017). Se encontró una diferencia significativa ($p < 0.001$) en la proporción de la delección de *pfhrp2* entre las muestras más recientes y las del periodo anterior, siendo mayor en el periodo 2015 – 2017; algo que no ocurre con la proporción de la delección del gen *pfhrp3* (Anexo 6).

7.2. Amplificación Selectiva de Genoma Completo (sWGA) y Secuenciación de Genoma completo

7.2.1. Modificaciones en la estandarización del protocolo de sWGA y WGS

El protocolo de sWGA de una y dos rondas fue evaluado a partir de los resultados obtenidos luego de ejecutar el WGS. Se observó que el porcentaje de *reads* alineados con el genoma de referencia fue significativamente mayor cuando se emplearon dos rondas de sWGA, asimismo presentó una menor dispersión de datos; en contraste con lo observado cuando el genoma del parásito fue enriquecido utilizando solamente una ronda de sWGA.

Solo se emplearon muestras con parasitemia mayor a 200 parásitos/ μ L. Asimismo se observó que a mayor parasitemia el porcentaje de *reads* alineados con el genoma de referencia es mayor. El porcentaje promedio de *reads* alineados fue mayor al 50% en las corridas que fueron enriquecidas con dos rondas de sWGA, a diferencia de lo observado con una ronda, donde el porcentaje fue de mínimo 29.4% a 85.6% (Figura 9).

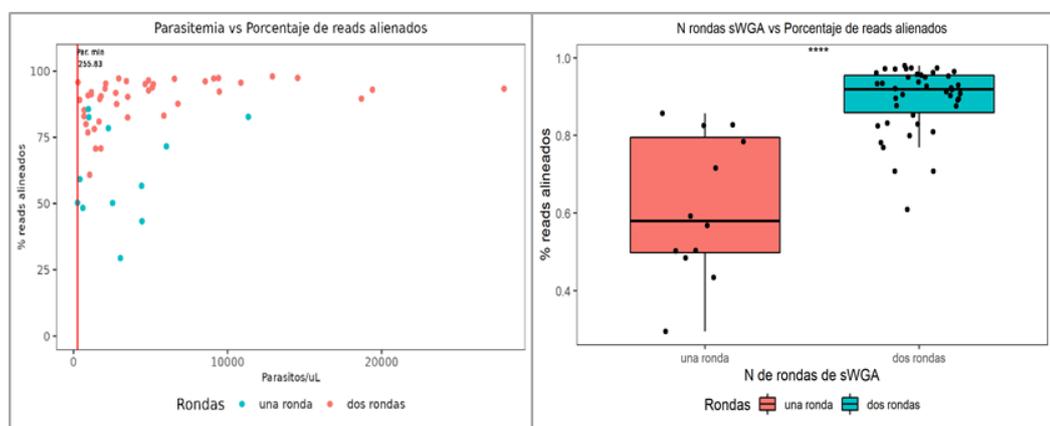


Figura 9. Resultados de WGS en base a sWGA. Izquierda, relación de la parasitemia con el porcentaje de *reads* alineados con el genoma de referencia. Derecha.

Distribución del porcentaje de *reads* alineados luego de sWGA con una o dos rondas.

En la figura 10 se observa que la cobertura promedio de las cinco corridas ejecutadas fue diferente, presentándose la mayor dispersión en la primera y tercera corrida. Asimismo, la corrida tres presentó los menores valores de cobertura promedio. En la corrida uno solo se ejecutó una ronda de sWGA y en la tercera corrida cuatro de las diez muestras analizadas fueron enriquecidas con solo una ronda de sWGA, de las cuales 3 presentaron valores de cobertura promedio menor a 10. De igual forma, en la figura 10 se muestran en diagramas de cajas los porcentajes de genoma con cobertura de 5X obtenidos con cada corrida, donde también se observa que la primera corrida presenta mayor dispersión de datos, junto con la tercera corrida.

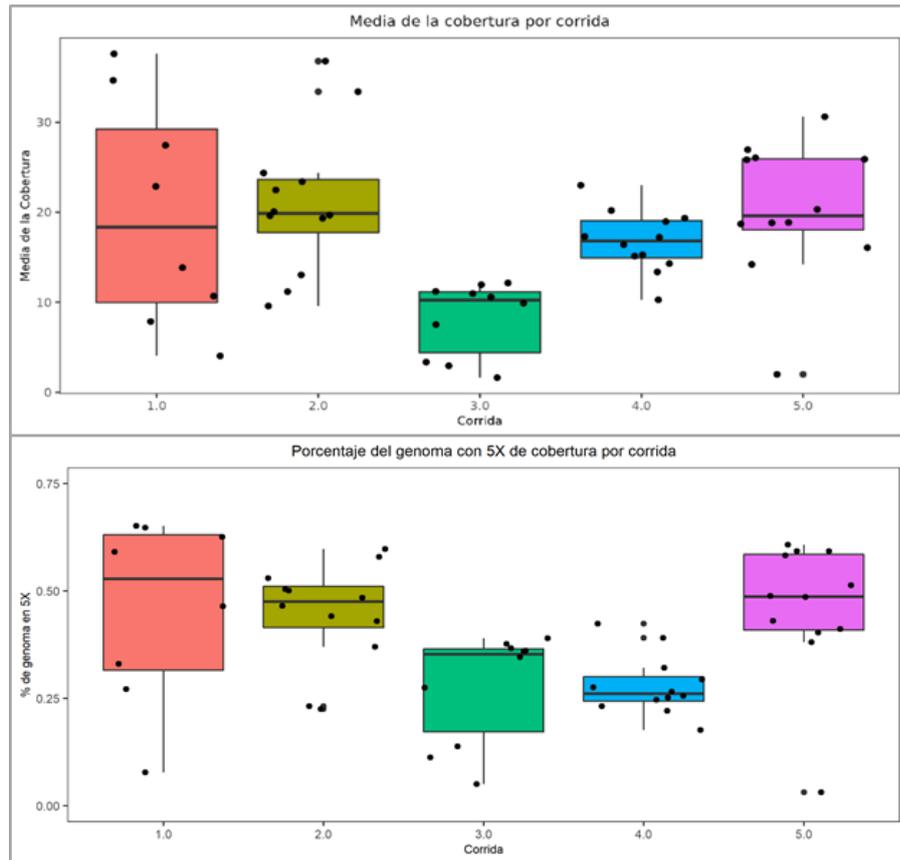


Figura 10. Cobertura de las corridas de WGS. Imagen superior. Media de la cobertura de cada una de las corridas de WGS ejecutadas. Imagen inferior. Porcentaje promedio del genoma que tiene una cobertura mínima de 5X, por corrida de WGS.

7.3. Análisis de datos de genoma completo para la caracterización de la población de estudio

7.3.1. Origen de los genomas analizados

Luego de aplicar los filtros a los archivos de variantes (VCFs), se conservaron un total de 100 muestras y 39164 SNPs, los cuales fueron incluidos en la segunda etapa del análisis. Las muestras tenían como lugar de origen principalmente los distritos de Iquitos (43/100), seguido de Nauta (17/100), San Juan (15/100), y Mazan

(10/100), como se indica en la tabla 4. Asimismo, los años de colecta de las muestras fueron entre 2006-2018.

Tabla 4. Tabla resumen de los distritos de origen de los genomas incluidos en el análisis

Distrito de origen	n
ALTO NANAY	2
IQUITOS	43
LAGUNAS	1
MAZAN	10
MORONA	1
NAPO	2
NAUTA	17
PAUCARES	1
PUNCHANA	6
SAN JUAN	15
TORRES CAUSANA	1
URARINAS	1

Adicionalmente, en la tabla del Anexo 7 se detallan las cantidades de las muestras con y sin la delección del gen *pfhrp2*. Se consideró solo la delección del gen *pfhrp2*, dado que no se contaba con la información de la delección del gen *pfhrp3* para todas las muestras obtenidas de datos públicos.

7.3.2. Caracterización genómica de la muestra analizada

7.3.2.1. Complejidad de la infección

El análisis con Fws sugieren que de los 100 genomas analizados sólo 1 presentaría más de una clona (Figura 11), se observó que dicha muestra correspondería a un parásito con la delección del gen *pfhrp2* que fue colectada en el 2018. El análisis con el programa estMOI guardó concordancia con los resultados de Fws de las 100 muestras analizadas.

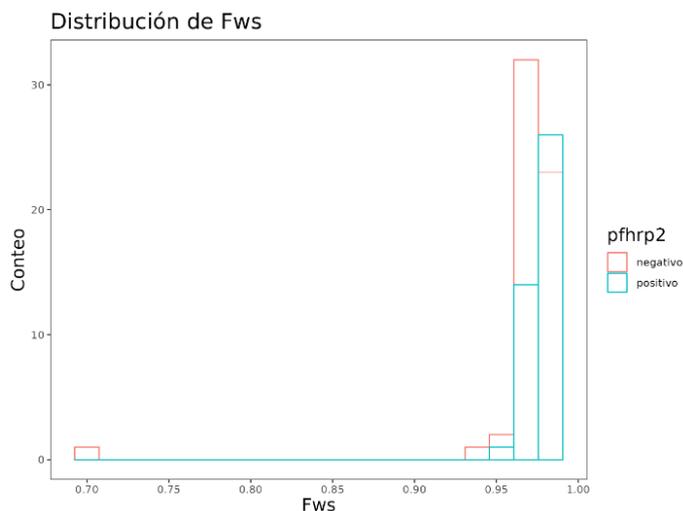


Figura 11. Análisis de la complejidad de infección mediante Fws. Un valor de Fws ≥ 0.95 , sugiere la presencia de una única clona en la muestra analizada.

7.3.2.2. Análisis de estructuración poblacional

7.3.2.2.1. Análisis de componentes principales (PCA)

Luego de aplicar el filtro por desequilibrio de ligamiento se conservaron 9037 SNPs, que fueron incluidos en el análisis PCA. Se graficaron los componentes principales (PC) 1 y 2. Por definición, cada PC explica los mayores patrones de variabilidad de las muestras, luego de que por este análisis el set de más de 9 mil SNPs fuera reducido dimensionalmente a unos 20 PCs (95,96). Adicionalmente, en el Anexo 8 se muestran las tablas con los primeros 20 SNPs que contribuyen a cada PC, y que se distribuyen en diferentes regiones del genoma.

En el gráfico de PCs se observa que el PC1 explica el 17.8% de variabilidad en la muestra, frente al 10.5% de PC2. De manera semejante, el PC3 solo explicaría el 9.41% de variabilidad de la muestra (Anexo 9b). No obstante, se puede apreciar la distribución de las muestras en base a las variables de estudio correspondientes al año de colecta y la presencia de la delección del gen *pflhrp2*.

Las muestras de *P. falciparum* se agruparon de acuerdo a la presencia de la delección del gen *pfhrp2* y al año en que se colectaron (Anexo 9a). Primero, las muestras que fueron colectadas entre el periodo 2006 – 2011 se encuentran disgregadas y apartadas de las muestras colectadas en años posteriores. Por otro lado, las muestras colectadas entre 2012 – 2018, forman dos agrupaciones. Estas agrupaciones corresponden a los grupos de parásitos negativos (con la delección) y positivos (sin la delección) para el gen *pfhrp2*. Asimismo, se observan algunas muestras negativas y positivas del 2009 – 2010 agrupando entre las muestras negativas y también entre las positivas del 2012 – 2018. Lo que podría deberse a una cercanía genética entre las muestras de ambos periodos. Otro aspecto a resaltar es la presencia de muestras (positivas y negativas) de 2016 ubicadas en la región intermedia, entre los grupos negativos y positivos del periodo correspondiente, que podría deberse a una composición mixta de estas muestras.

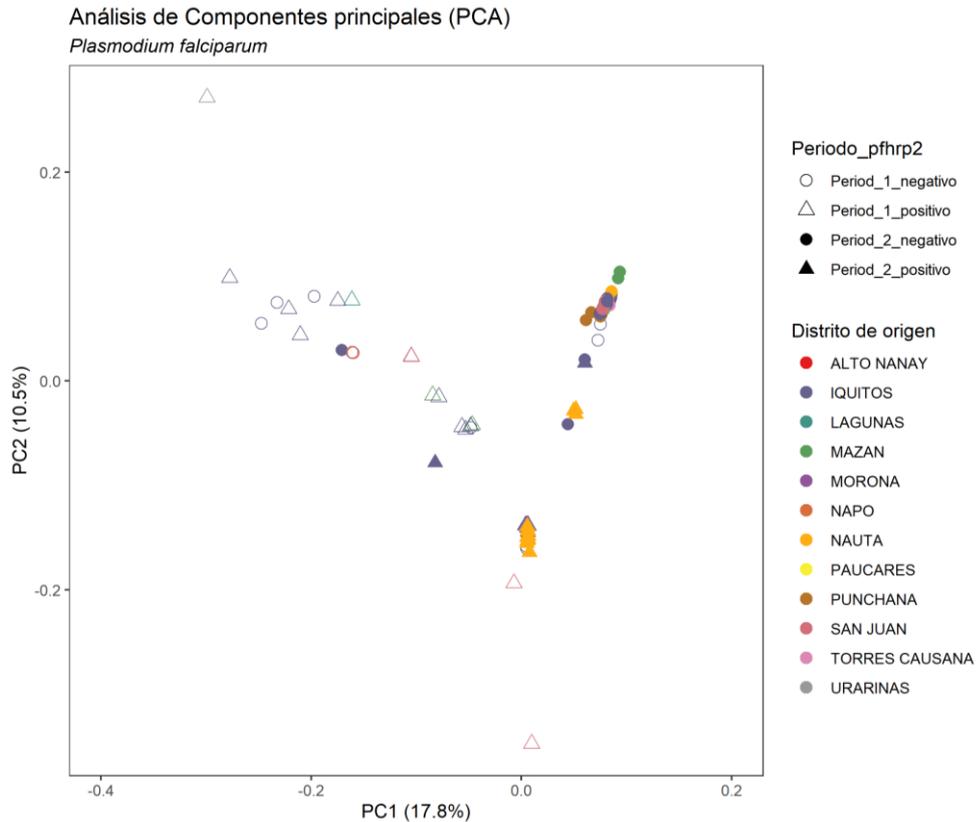


Figura 12. Análisis de componentes principales (PCA). Resultado de las 100 muestras. Se muestran las muestras agrupadas en periodos de colecta. Periodo 1: 2006 – 2011. Periodo 2: 2012 – 2018. Asimismo, las muestras se colorearon de acuerdo al distrito de origen.

En base a lo observado en el PCA, las muestras de *P. falciparum* se pudieron agrupar considerando dos periodos de colecta: Periodo 1: de 2006 – 2011; y Periodo 2: 2012 – 2018 (Tabla 5). En la figura 12 se graficó el lugar de colecta de las muestras y la diferenciación temporal junto con la presencia o ausencia de la delección del gen *pfhrp2*. Las muestras positivas para *pfhrp2* del Periodo 2, provienen en su mayoría de los distritos de Nauta (Comunidad de Santa Emilia) e Iquitos y regiones aledañas. En el caso de las muestras negativas, el lugar de

procedencia es más diverso. Por otro lado, las muestras del Periodo 1 también fueron colectadas en distintas regiones, en su mayoría de Iquitos y zonas aledañas.

Tabla 5. Periodo de colecta y perfil de la deleción de *pfhrp2*

Periodo	<i>pfhrp2</i>	Cantidad
Period_1 (2006 – 2011)	negativo	14
	positivo	19
Period_2 (2012 – 2018)	negativo	45
	positivo	22

Mediante una prueba exacta de Fisher se obtuvo que la diferencia en la proporción de parásitos con la deleción del gen *pfhrp2* era significativamente mayor en el Periodo 2 (Anexo 10). Si bien es un resultado que debe tomarse con cautela por la aproximación utilizada en la selección de las muestras, no deja de ser destacable la coincidencia con lo observado en la comparación que se realizó en la primera etapa y lo reportado previamente.

7.3.2.2.2. Análisis de ancestría

A partir del resultado del error de *cross-validation* obtenido con el programa Admixture, no fue posible concluir el número de clusters presentes en la muestra de parásitos analizados. Sin embargo, se puede sugerir que el número de clusters se encuentra entre 4 u 8, al presentar los valores más bajos de error (Anexo 11). Teniendo en cuenta esto, se graficó el resultado de admixture para números de clusters de 4, 8 y 9 (Figura 13).

A partir de los gráficos se observa que se comparte un patrón común. En el caso de las muestras correspondientes al periodo 2 (2012 – 2018), éstas estarían compuestas principalmente por dos clonas distintas entre sí, que a su vez corresponderían a los

parásitos con y sin la delección; que en los gráficos se distingue como los colores predominantes en cada plot. Por otro lado, las muestras del periodo 1 (2006 – 2011) presentarían una composición mixta tanto en el grupo de los parásitos con la delección y aquellos que no tienen la delección.

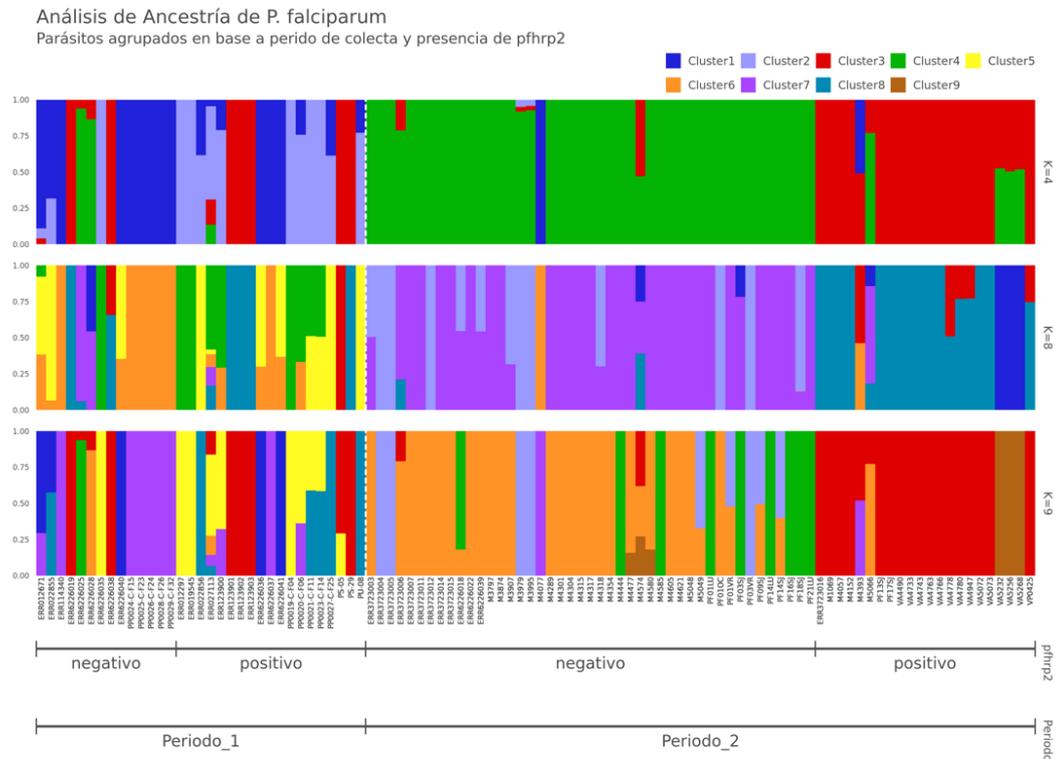


Figura 13. Análisis de ancestría ejecutado con Admixture. Se muestra la agrupación de los clusters, de acuerdo al periodo de colecta y a la presencia del gen *pfhrp2*.

Adicionalmente, se llevó a cabo un análisis de estructuración poblacional incluyendo solo las muestras de los distritos de Iquitos y Nauta (Anexo 12). Este análisis complementario se hizo considerando que la mayoría de las muestras provenían de ambos distritos. A pesar de que las muestras de Nauta eran en su mayoría positivas para el gen *pfhrp2* y que fueron colectadas en el Periodo 2, el agrupamiento del PCA es similar al observado con el total de muestras. El análisis

de ancestría muestra un patrón de clusters similar, sugiriendo en este caso la presencia de solo 4 clusters, y que los parásitos con la delección de *pfhrp2* son diferentes a los parásitos sin la delección. También se observó que, en ambos casos, la población clonal del Periodo 2 estaba presente en el Periodo 1. En dicho periodo ambos grupos estuvieron compuestos por más de una clona, y para el Periodo 2 la complejidad habría disminuido (como se sugirió con el total de muestras). Cabe resaltar que a pesar que en el Periodo 1 no se tenían muestras de Nauta, el análisis de ancestría (Anexo 12b.2) muestra que la población de Nauta presente en el Periodo 2 (*pfhrp2* positivas) está genéticamente relacionada con las muestras de Iquitos del Periodo 1 y del Periodo 2 (*pfhrp2* positivas) al observarse que el cluster predominante en Nauta también estaba entre los clusters del Periodo 1. Esto es un indicio de que el agrupamiento no está influenciado por factores geográficos.

7.3.2.3. Decaimiento del desequilibrio de ligamiento (LD)

A partir de los resultados de estructuración poblacional se continuaron los análisis con las muestras del periodo 2 por cada grupo de parásitos por separado, las cuales fueron *pfhrp2/3* negativas o positivas. El decaimiento del LD promedio con respecto a la distancia que separa dos SNP se muestra en la figura 14, tanto para las muestras con la delección y sin la delección. En ambos casos el LD promedio tiene un valor mayor a 0.2, siendo en el caso de las muestras positivas para *pfhrp2/3* las que presenten un mayor valor. Asimismo, se observa que en el caso de las muestras con la delección el valor promedio decae cuando los SNPs se encuentran separados a una distancia de ~200 Kb y se mantiene mayor a 0.4, mientras que en las muestras positivas el valor de R^2 se mantiene elevado hasta ~400Kb. En este último grupo,

el valor promedio de LD se mantiene por encima de 0.2 hasta pasado los 500 mil pares de bases.

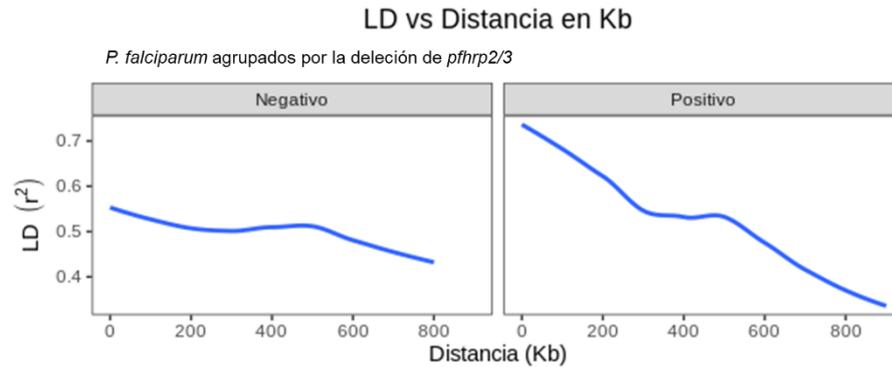


Figura 14. Decaimiento del Desequilibrio de ligamiento (LD). Izquierda: Decaimiento del desequilibrio de ligamiento vs la distancia en pb que separa los SNPs, del grupo de muestras con la delección de *pfhrp2/3* del periodo 2 (45). Derecha: Decaimiento del LD vs distancia en pb, para los parásitos positivos para *pfhrp2/3*.

7.3.3. Búsqueda de barridos selectivos en los genomas del periodo 2

De manera similar al análisis previo, las pruebas para búsqueda de barridos selectivos se realizaron considerando que la muestra incluida de *P. falciparum* del Periodo 2 está conformada por dos grupos de parásitos (*pfhrp2/3* positivos y negativos). Los resultados descritos a continuación corresponden a lo obtenido de manera específica para cada grupo, así como aquellos barridos que se identificaron en común para ambos grupos de parásitos.

7.3.3.1. Análisis de diferenciación poblacional por Fst

El cálculo de Fst a nivel de genoma se realizó confrontando los parásitos con y sin la delección de *pfhrp2/3* del periodo 2. En la figura 15 se muestra que los valores de

Fst oscilaron desde 0 alcanzando incluso la unidad, en todos los cromosomas. No se distinguió algún locus que pueda representar una región candidata a diferenciar ambos grupos y por ende candidato a estar o haber estado bajo selección.

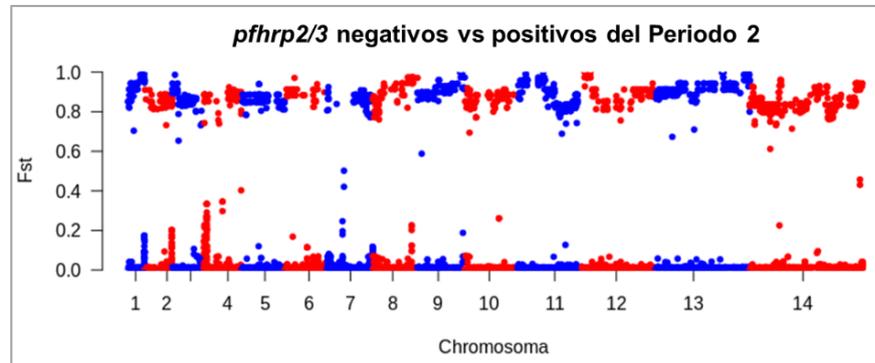


Figura 15. Manhattan plot de los valores de Fst de Hudson a lo largo del genoma de los parásitos.

7.3.3.2. Prueba de neutralidad D de Tajima

Se ejecutó la prueba D de Tajima por cada cromosoma para las muestras con y sin la delección de los genes *pfhrp2/3*. En las muestras negativas para *pfhrp2/3*, el valor obtenido del estadístico fue principalmente negativo a lo largo del genoma, observándose algunos loci en los que el valor fue positivo y mayor o igual a 2 en los cromosomas 4, 6, 7, 10, 11 y 12, que podrían corresponder a eventos de selección de balance. Por otro lado, las muestras positivas para *pfhrp2/3*, también presentaron valores negativos o cercanos a cero a lo largo del genoma, y de manera similar presentaron valores mayores o iguales a 2 en regiones dentro de los cromosomas 4, 6, 7, 10, 12 y 14.

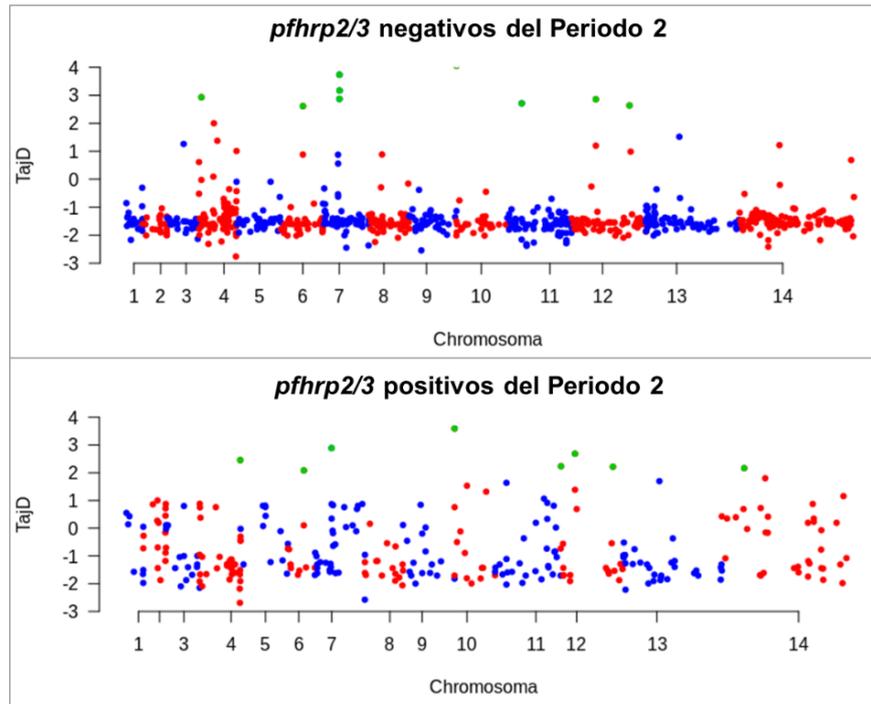


Figura 16. Manhattan plot del D de Tajima. Superior. Parásitos portadores de la deleción de *pfhrp2/3*. Inferior. Parásitos positivos para el gen *pfhrp2/3*.

Al explorar las regiones donde el D de Tajima alcanzó valores mayores a 2, se identificaron tanto en el grupo con y sin la deleción en los cromosomas 6, 7 y 12, los genes *VAR* que codifican a la proteína PfEMP1; y el gen *ACBP1*, codificante de la proteína de unión a Acyl-CoA en el cromosoma 10. Por otro lado, en el grupo de parásitos con la deleción se identificó una región candidata en el cromosoma 11 que contenía al gen *FP2A*, codificante de la proteína falcipaina 2a.

7.3.3.3. Detección de barridos selectivos mediante la estimación de iHS

Se consideraron como regiones candidatas significativas aquellas que presentaron un valor de $-\log_{10}P\text{-value} \geq 5$. También se han considerado de interés aquellos loci que presentaron un valor de $iHS \geq 4$ y un $-\log_{10}P\text{-value} \geq 4$.

Teniendo en cuenta esto, mediante la prueba basada en desequilibrio de ligamiento, iHS, no se identificaron regiones candidatas en el cromosoma 8 ni en el cromosoma 13, cerca de la región donde ocurre la delección de los genes *pfhrp2* y *pfhrp3*. Sin embargo, se identificaron regiones candidatas a selección en otros cromosomas.

En los parásitos con la delección *pfhrp2/3* algunas de las regiones candidatas se ubicaron en los cromosomas 4, 5, 7, 11, 12, 13 y 14. Entre los genes dentro de las regiones se incluyeron los genes *var* de PfEMP1, el gen *PTPA* y el gen *AP2-G*. De manera similar, en el grupo de parásitos positivos para *pfhrp2/3*, se identificaron barridos selectivos en loci en los cromosomas 4, 5, 11, 12, 13 y 14. Se identificó que los barridos se ubican en los genes *CYP19C*, *var* (PfEMP1), de L-lactato deshidrogenasa y de una proteína rica en leucina.

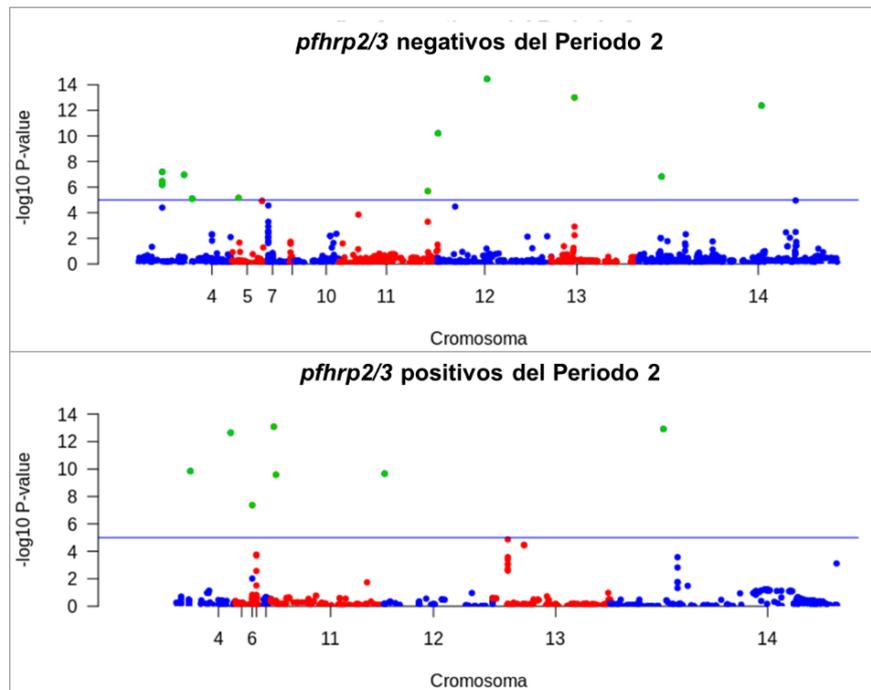


Figura 17. Manhattan plot del $-\log_{10}P$ -value del |iHS|. Resultado obtenido para todo el genoma de los parásitos con la delección de *pfhrp2/3* (superior), y sin la delección (inferior). En verde se indican los valores que son significativos, al superar

el umbral de 5 de $-\log_{10}P$ -value, o que fueron considerados de interés al tener $|iHS| \geq 4$ y $-\log_{10}P$ -value ≥ 4 .

7.3.3.4. Detección de barridos selectivos mediante las pruebas de comparación de poblaciones XP-EHH y rsb

La prueba XP-EHH, que permite detectar barridos selectivos candidatos específicos comparando grupos de parásitos, identificó señales significativas ($-\log_{10}P$ -value ≥ 5) en loci ubicados en los cromosomas 4, 11 y 12 de los parásitos *pfhrp2/3* positivos. Los genes contenidos en los loci fueron el gen *RIF*, una ribonucleoproteína putativa y *VAR* (PfEMP1). Asimismo, solo se identificó un locus en el cromosoma 11 de los parásitos con la delección de *pfhrp2/3*. La señal candidata se encontraba en el gen *FP2A*, gen que codifica a la proteína falcipaina 2a.

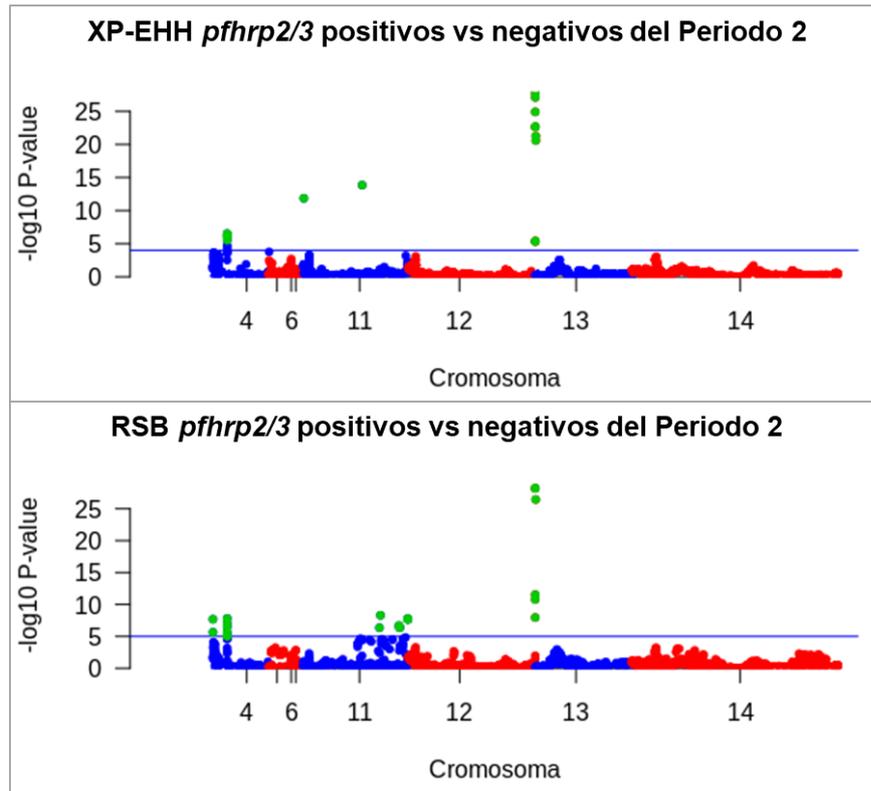


Figura 18. Manhattan plot del $-\log_{10}P$ -value para las pruebas XP-EHH y rsb. En verde se resaltan los valores que superaron el límite de significancia estadística de 5. Superior: XP-EHH, inferior: rsb

La prueba de comparación rsb, no identificó alguna señal de selección significativa ($-\log_{10}P\text{-value} \geq 5$) en genes conocidos en el grupo de parásitos con la delección de *pfhrp2/3*. Sin embargo, en el caso de los parásitos positivos para el gen *pfhrp2/3* si identificaron señales en el gen *RIF* y *VAR* (PfEMP1).

7.3.3.5. Análisis de ontología genética (Gene Ontology) de los genes identificados

El análisis de Gene ontology (GO) realizado con la herramienta en línea de PlasmODB dio resultados no significativos de acuerdo con el p-value ajustado de Bonferroni. Se procedió a evaluar la tasa de descubrimientos falsos (*False*

Discovery rate – FDR) para los términos de GO de los genes identificados con señales de selección candidatas por iHS, XP-EHH y rsb en ambos grupos de parásitos. En el grupo de parásitos con la doble delección, se incluían términos de GO de procesos catabólicos de componentes proteicos celulares y organonitrógenos, además de procesos de adhesión celular, entre otros; no obstante, no alcanzaron significancia al tener $FDR > 0.05$ (Anexo 19). Por otro lado, en el set de genes identificados con señales de selección en el grupo de parásitos sin la delección de *pfhrp2/3*, el FDR fue < 0.05 para procesos de interacción con el hospedero, incluyendo adhesión celular, evasión de la respuesta inmunitaria del hospedero, respuesta a otros organismos y evasión antigénica, principalmente (Anexo 20). En la figura 19 se muestra una representación gráfica de los términos de Gene ontology identificados en el conjunto total de genes de las muestras con la doble delección, obtenido con la herramienta online Revigo (<http://revigo.irb.hr/>). El tamaño de los círculos refleja la frecuencia con la que aparecen los términos relacionados a un proceso biológico como es el caso de los procesos catabólicos proteicos y adhesión al hospedero. Asimismo, el color refleja el \log_{10} del valor p obtenido con la herramienta de PlasmoDB. Un gráfico similar fue generado para los parásitos positivos para ambos genes y se encuentra adjunto en el Anexo 21.

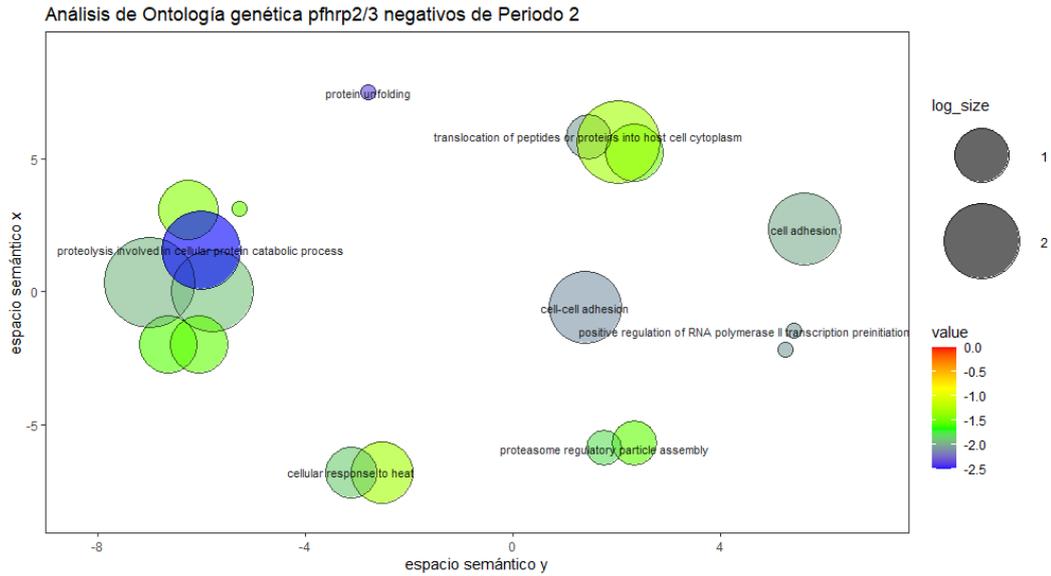


Figura 19. Resultados de análisis de ontología genética para los parásitos con la doble deleción. Procesos biológicos identificados en el set de genes totales identificados con regiones candidatas a selección. Gráfico obtenido con Revigo a partir de los resultados del análisis de GO con PlasmoDB.

7.3.4. Análisis de genes identificados con barridos selectivos en parásitos *pfhrp2/3* negativos

De los 24 SNPs en *FP2A*, 13 tenían como efecto el cambio del aminoácido codificado, mientras que los 11 restantes eran variantes sinónimas. Asimismo, de las 13 variantes con cambio de sentido o cambio del aminoácido, 3 polimorfismos estaban fijados con una frecuencia de 100% en la muestra (Anexo 22). En las 10 variantes restantes se observó que la frecuencia variaba entre periodos y grupos según la presencia de la deleción. Este es el caso del el SNP identificado mediante la prueba XP-EHH (11_593409_G_A.A), el cual tenía una frecuencia de 59.4% y 71.4% en el grupo positivos y negativos para *pfhrp2/3* en el periodo 1; mientras que en el periodo 2 la frecuencia de este SNP aumentó a 86.4% y 97.8% en el grupo

positivo y negativo para *pfhrp2/3* del periodo 2. Es decir, el SNP estaría prácticamente fijado en las muestras con la delección del periodo 2. Esta variante se encontraría en una región no definida de la proteína, cerca de un dominio transmembrana. Asimismo, se observa que otras variantes también localizadas en regiones no definidas de la proteína se habrían fijado al pasar del Periodo 1 al Periodo 2 en ambos grupos de parásitos.

Por otro lado, en el gen *AP2-G* se identificaron 30 SNPs que generaban un cambio en la secuencia de aminoácidos de la proteína (Anexo 23). De estos se observó que ocho SNPs aumentaron de frecuencia entre el Periodo 1 al Periodo 2 en el grupo de parásitos con la delección. Lo que se evidenció también en el SNP core de la región donde se detectó el barrido selectivo (12_909914_C_T.C), que pasó de 46.2% a ~100% en los parásitos con la delección, mientras en los parásitos sin la delección el aumento fue de 71.1% a 100%. De manera similar a lo observado en *FP2A*, esta variante se habría fijado en el Periodo 2. En el caso de este gen que codifica para un factor de transcripción, los SNPs identificados se localizan en regiones no definidas de la molécula.

VIII. DISCUSIÓN

Plasmodium falciparum de la región Amazónica de Perú ha sido descrito en estudios recientes como una población que ha sufrido un reemplazo y expansión clonal. Asimismo, se ha reportado que las deleciones de los genes *pfhrp2* y *pfhrp3* han seguido aumentando en frecuencia en esta población, sin alguna explicación subyacente. Factores como la presión de selección en regiones del genoma del parásito han explicado anteriormente la persistencia del parásito mediante el desarrollo de resistencia a fármacos (16). De manera similar, el hospedero y el vector pueden ejercer selección en genes envueltos en procesos biológicos como la interacción con el hospedero, o en las etapas sexuales del parásito (15,67). En ese sentido, en este estudio se llevó a cabo la caracterización genética y genómica de una muestra de *P. falciparum* de la región Loreto; y la búsqueda de regiones del genoma que muestren evidencia de haber sufrido presión de selección que pueda explicar la persistencia de *P. falciparum* en la selva de Perú, y adicionalmente el aumento de los parásitos con la deleción de los genes *pfhrp2* y *pfhrp3*.

La muestra de parásitos *P. falciparum* caracterizados en esta tesis presentan una frecuencia de la deleción del gen *pfhrp2* mayor con respecto a lo reportado en otros estudios con muestras de la Amazonía peruana colectadas en años anteriores. Los resultados de este estudio demuestran que la frecuencia de la deleción del gen *pfhrp2* es significativamente mayor en comparación a lo reportado anteriormente, siendo estadísticamente significativo a lo reportado en un estudio previo que empleó el mismo protocolo de caracterización de la deleción (6). El aumento de la frecuencia de la deleción del gen *pfhrp2*, *pfhrp3* y de la doble deleción de ambos genes ha sido reportado recientemente en un estudio por Valdivia y colaboradores.

En dicho estudio se observa que en 324 muestras colectadas entre 2011 y 2018 el gen *pfhrp2* está delecionado en 77.2% y *pfhrp3* en 71.6% y la doble deleción en 67%. Asimismo, los investigadores reportan que la doble deleción habría cambiado su prevalencia de 14.3% en 2011, alcanzando un pico de 88.39% en 2016 y manteniéndose en 65% en las muestras de 2017-2018 (8). No obstante, en algunos de los periodos considerados el número de muestras fue bajo, por lo cual podría haber un sesgo en estimación de la frecuencia de la deleción en dichos periodos. De manera similar se ha observado la tendencia de aumento de ambas deleciones en diferentes estudios con muestras de diferentes periodos (3,5–8).

Por otro lado, en el presente estudio no se observó una variación significativa de la frecuencia de la deleción del gen *pfhrp3* en el tiempo, con respecto al estudio previo de Quispe (Anexo 6). Esto se debería a la alta proporción de muestras *pfhrp3* positivas provenientes del distrito de Nauta colectadas entre 2015 y 2016 (74%, Figura 8), la cual contrasta con lo reportado en el estudio citado, en el cual también se incluyeron muestras de la comunidad de Santa Emilia del distrito de Nauta. En dicho estudio las muestras positivas para *pfhrp3* se detectaron en ~51% de 227 muestras totales, y en 57% de las 93 muestras que provenían de Santa Emilia (6). El distrito de Nauta, en específico la comunidad de Santa Emilia, de donde proviene la mayoría de muestras de dicho distrito; se encuentra en una localización remota, ubicada a 150 km de la ciudad de Iquitos, es únicamente accesible por bote, y no tiene ningún centro de salud. Lo observado en Santa Emilia sería similar a lo observado en otros estudios que reportan proporciones variables de las deleciones según el lugar de colecta de las muestras (7,8,97).

La proporción de la delección contrasta con reportes de otras regiones de Sudamérica. En el caso de Brasil se reportó que la delección de *pfhrp2* estaba presente en 71.7% (71/99) y 100% (60/60) de las muestras del estado de Acre y Amazonas, respectivamente; dando una frecuencia total de 82.4% de 159 muestras totales (98). Asimismo, las muestras fueron colectadas entre los años 2016 – 2017, en ese periodo los parásitos con la delección probablemente han seguido aumentando hasta casi fijarse. Esto último se sugiere, dado que previamente se reportó para el estado de Acre que la frecuencia de la delección de *pfhrp2* era de 31.6% (25/79) (10). Por otro lado, en muestras colectadas en Colombia en los periodos 2003 – 2010 y 2011 – 2012, la delección de *pfhrp2* se reportó solo en 15 muestras del primer periodo (5.9% de 253) colectadas en uno de los siete departamentos incluidos; mientras que la delección de *pfhrp3* fue identificada en ambos periodos, en frecuencias de 45.5% de 112 y 41.9% de 253, en cada periodo respectivamente (9).

Adicionalmente, en las muestras con la delección predomina la delección de uno de los genes flanqueantes a *pfhrp2* y *pfhrp3*. El perfil identificado en mayor frecuencia está conformado por la delección del gen *pfhrp2* y su gen flanqueante PF3D7_0831900, ubicado en la región *upstream*; algo similar se observa con el patrón de la delección de *pfhrp3* y PF3D7_1372100, ubicado en la región *downstream* a *pfhrp3*. Ambos perfiles de delección habrían aumentado en frecuencia, conformando el 93% y 88% del total de parásitos portadores de las delecciones de *pfhrp2* y *pfhrp3*, respectivamente. La coincidencia de la delección de los genes y los flanqueantes es consistente con los estudios previos en muestras de Perú (6–8) y otras regiones como Brasil y Colombia (9–11).

Por otro lado, en este estudio se ha caracterizado solamente la presencia del exón 2 tanto de *pfhrp2* y *pfhrp3*, para definir la delección de los mismos. Debe tenerse en cuenta que la delección de *pfhrp2* puede ocurrir de forma parcial, por lo que el exón 1 podría estar delecionado cuando el exón 2 está presente total o parcialmente (97).. En ese sentido, en muestras de Brasil se ha descrito un escenario que apuntaría a la presencia de delecciones parciales. Los investigadores de dicho trabajo compararon un protocolo de PCR que detecta solo al exón 2, con otro que detecta tanto al exón 1 como al exón 2. De este modo, se observó que, con el primer protocolo, 60 de las 129 muestras analizadas presentaban la delección; mientras que con el segundo protocolo la cantidad de muestras con la delección se reducía solo a 13 (99). A pesar que los autores no mencionan la existencia de las delecciones parciales, esto podría ser un caso. Sin embargo, esto debería ser confirmado por otros métodos como la secuenciación de nueva generación, dado que el patrón contrasta con lo sugerido previamente (97). La baja frecuencia de las delecciones en dicho estudio también se puede atribuir al amplio periodo en el que las muestras fueron colectadas (2002 - 2020).

Previo al análisis genómico se llevó a cabo la optimización del protocolo de sWGA y WGS. El resultado variable en el rendimiento del proceso de WGS se debería a factores técnicos como operativos. Uno de los factores determinantes fue el empleo de dos rondas de amplificación genómica previa al WGS, obteniéndose mejor cobertura que con una sola ronda, como se ha visto en estudios previos (71). Asimismo, la calidad e integridad del ADN también impactó en el rendimiento, dado que se observó una mejora evidente cuando se empleó ADN recién extraído junto con un paso de limpieza adicional del producto de sWGA antes de la

preparación de bibliotecas para WGS, para cumplir con los criterios de calidad sugeridos por el fabricante (73). Además, la reducción en el tiempo de manipulación para detener la digestión enzimática en el protocolo de preparación de bibliotecas también habría influido en la mejora de los resultados. Esto se debería a que un mayor tiempo de exposición a la enzima puede conducir a un problema denominado *overtagmentation* que afecta negativamente la calidad de los productos de secuenciación (73).

Los análisis de estructuración poblacional con los datos de genoma completo de *P. falciparum* muestran que los parásitos colectados en años más recientes serían producto de una expansión clonal de dos poblaciones diferenciadas. El patrón clonal observado en las muestras del Periodo 2, sugieren que los parásitos del Periodo 1 (2006 – 2011) habría sufrido un evento de cuello de botella, y las clonas sobrevivientes se habrían expandido. Las dos poblaciones clonales que se distinguen en el Periodo 2 corresponden a los parásitos con y sin la delección del gen *pfhrp2* (2012 – 2018). Esto es semejante a los resultados obtenidos en el trabajo de Quispe, para el cual incluyó muestras de 2012 – 2014, reportando la presencia de dos clusters y entre los cuales la proporción de parásitos con y sin la delección de *pfhrp2/pfhrp3* era diferente. Asimismo, la diferencia temporal en la población de *P. falciparum* de Perú ha sido reportado anteriormente (8,100). La característica clonal de los dos grupos de parásitos se respalda con el alto valor promedio de desequilibrio de ligamiento a lo largo del genoma, esto coincide con una baja tasa de recombinación en ambos grupos de parásitos del Periodo 2 (*pfhrp2/pfhrp3* negativos y positivos) (101).

La población de *P. falciparum* habría pasado de una población de mayor complejidad, conformada por más de dos clonas, a una población reciente más homogénea o clonal, lo que probablemente correspondería a un remplazo clonal en el tiempo. A pesar que con el análisis de ancestría del presente estudio no se ha concluido el número de clusters, el resultado con ocho clusters, tendría semejanza con los hallazgos de estudios previos que reportan que la población de *P. falciparum* habría estado conformado por varias clonas en muestras de antes del 2011, y en muestras más recientes se ha reportado menor número de clonas o haplotipos (7,8). En el trabajo de Valdivia y colaboradores se reportó un haplotipo construido a partir de siete microsatélites (H8) presente en 144 de 203 muestras, que habría incrementado en el tiempo desde 2011 hasta ser el haplotipo predominante en 2018 (88.9%). Asimismo, 120 de las muestras portaban la doble delección de *pfhrp2/3* y solo 7 muestras eran positivas, por lo cual los autores sugieren que la expansión de dicho haplotipo explicaría el incremento de los parásitos con la delección de *pfhrp2*. Por el contrario, Akinyi y colaboradores reportaron que en muestras colectadas en los periodos 1998 - 2001 y 2003 – 2005, se reportó que los parásitos del primer periodo estarían conformados por cinco clusters, y en el segundo periodo habría hasta ocho clusters, de los cuales algunos serían híbridos de los clusters del primer periodo (47). Una descripción similar de la conformación de varias clonas ha sido reportado también en 93 muestras colectadas durante 2009 – 2010, provenientes de comunidades de Requena, San Juan, Yurimaguas, Mazan, y otras cuatro localidades (7). En este último estudio se emplearon los mismos microsatélites de Akinyi, y se observaron algunas de las mismas clonas entre estudios. Adicionalmente, otro estudio con datos de genoma completo de muestras de 2006 – 2017 reportó que una

clona relacionada con Bv1 sería predominante desde 2011 en la Amazonía peruana, producto de un evento de reemplazo clonal (14). Una limitante del presente estudio es que no se ha evaluado la presencia de la clona Bv1 o de alguna relacionada, ya reportada en otros trabajos; sin embargo, los resultados obtenidos coinciden con el perfil de reemplazo y expansión clonal reportado.

Adicionalmente, al analizar solamente las muestras de los distritos de donde se colectaron la mayoría de las muestras (Iquitos y Nauta), se observó que el patrón de estructuración se mantenía. Se demostró de este modo que las muestras de Nauta del Periodo 2 (*pfhrp2* positivas) están genéticamente relacionadas con las de Iquitos, del Periodo 1 y 2. No obstante, la mayor frecuencia de parásitos positivos en Nauta y el análisis de ancestría indican que los parásitos *pfhrp2* positivos estuvieron presentes tanto en Iquitos y Nauta durante el Periodo 1. Para el Periodo 2 los parásitos con la delección se han expandido, casi reemplazando los positivos en Iquitos, mientras que en Nauta aún hay presencia de *pfhrp2* positivos. Esto sugiere que la expansión de los parásitos con la delección del gen *pfhrp2* en Nauta no ha sido igual a lo observado en Iquitos, lo que podría explicarse por ser una ciudad alejada y de difícil acceso.

Finalmente, el evento de cuello de botella que se sugiere que habría sido causado por las intervenciones del programa PAMAFRO, explicaría la característica clonal de la población del Periodo 2. Dicho programa estuvo en ejecución desde el 1 de octubre de 2005 hasta el 31 de diciembre de 2010, y abarcó actividades como el diagnóstico, el tratamiento de la enfermedad, así como el control del vector (102), algo sugerido también en algunos de los estudios previos. Esto coincide con la disminución de casos reportado hasta 2011 que también implicaría una disminución

en la diversidad genética y riqueza alélica, lo cual ocurriría de manera aleatoria. La expansión clonal descrita correspondería con el aumento de casos reportados en la región Amazónica, que alcanzó un pico en el año 2016. Al ser el cuello de botella un evento aleatorio, las clonas que se mantuvieron luego de PAMAFRO y que portarían la delección se expandieron, pero no se habrían seleccionado necesariamente por las intervenciones dado que no se han registrado casos de resistencia al tratamiento ofrecido, el cual consistió de un tratamiento combinado de artemisinina (mefloquina + artesunato) para infecciones por *P. falciparum* (102). Asimismo, no hubo reportes de mutaciones validadas asociadas con la resistencia al tratamiento ofrecido (14,100). Se sugiere más bien, que las variantes identificadas en este trabajo tendrían una ventaja que favorece la transmisión en un contexto general de baja transmisión, alta proporción de mono-infecciones y aspectos sociodemográficos característicos de la región, factores que en su conjunto explicarían la persistencia de las infecciones por *P. falciparum*, y en específico el aumento del grupo de parásitos con la delección de los genes *pfhrp2* y *pfhrp3* (23).

Mediante tres tipos de pruebas de selección se han identificado barridos selectivos en regiones genómicas de *P. falciparum* que podrían explicar su persistencia. Entre estos se encuentran los genes de la familia *var* que codifican para la proteína PfEMP1, y que se encuentran en el genoma de *P. falciparum* en número de ~60 (103). Los genes *var* se expresan en las etapas asexuales y etapas tempranas de la fase sexual gametocítica del parásito (104), y la proteína está involucrada en procesos biológicos de interacción con el hospedero humano, permitiéndole evadir el sistema inmunitario humano al intercambiar las variantes de PfEMP1 (105,106). Estos genes presentarían evidencia de selección de balance (*balancing selection*),

al ser expuestos al sistema inmunitario del humano, y a su vez podría explicar la persistencia de las infecciones por *P. falciparum* en la región Amazónica de Perú.

Adicionalmente, se identificaron barridos selectivos en ambos grupos de parásitos en una región que contiene al gen *ACBPI*, que codifica para una proteína de unión a acil-CoA. El gen *ACBPI*, codifica para la isoforma 1 de una proteína involucrada en metabolismo de moléculas lipídicas, y también presentaría evidencia de selección de balance, que podría explicarse por el uso de mefloquina en la terapia a malaria, dado que este fármaco inhibe la interacción entre las proteínas de unión a acil-CoA a las acil-CoAs en el parásito (107,108).

Por otro lado, el incremento de la frecuencia de los parásitos con las deleciones de los genes *pfhrp2* y *pfhrp3* no se debería a selección positiva de las deleciones. En este estudio no se encontró evidencia de selección de las deleciones, dado que ninguna de las pruebas detectó barridos selectivos en los cromosomas 8 y 13 cerca de la deleción del gen *pfhrp2* y *pfhrp3*, respectivamente. Estos resultados contrastan con lo reportado por Feleke y colaboradores (17), trabajo en el cual reportan una región >143 kb extendida en la región centromérica a la deleción de *pfhrp2*, sugiriendo que se habría seleccionado; en contraste en la región alrededor de *pfhrp3* donde no detectaron algo similar. El mismo artículo junto con otros, han sugerido que los parásitos con la deleción de los genes *pfhrp2* y *pfhrp3* habrían sido seleccionados en las regiones de África por el uso de las pruebas de diagnóstico rápido, ya que luego del diagnóstico las personas reciben el tratamiento para la enfermedad. Asimismo, no han descartado que además existan otras regiones del genoma del parásito siendo seleccionadas y que involucrarían otros procesos biológicos (17,21).

Entre los factores que se han sugerido que puedan favorecer el aumento de los parásitos con la delección de *pfhrp2*, se encuentran la edad, la baja transmisión y la búsqueda de tratamiento en base a un diagnóstico por PDRs (20,21). Sin embargo, esto ha sido sugerido para la región africana; en el caso de Perú no se puede extrapolar un efecto por uso de PDRs, y tampoco se ha reportado alguna relación con la edad; pero no puede descartarse que la baja transmisión de malaria en la región pueda favorecer su incremento. Las mono-infecciones son predominantes en las regiones de baja transmisión y podrían aumentar las oportunidades del parásito de sobrevivir, como se ha sugerido anteriormente, ya que sería un contexto menos competitivo para la población circulante y con menos eventos de recombinación entre linajes, lo que llevaría a un cambio de linajes más lento (8). Adicionalmente, se ha reportado que los parásitos con la delección de *pfhrp2/pfhrp3* tendrían menor parasitemia por microscopía en comparación con las muestras silvestres (20).

A pesar de que el análisis de *Gene Ontology* no dio resultados estadísticamente significativos, se pudo observar que los términos de ontología que aparecen en mayor frecuencia y con $FDR < 0.05$, refieren procesos biológicos de interacción con el hospedero y procesos metabólicos del parásito en el grupo *pfhrp2/3* positivos.

Por otro lado, se han identificado barridos selectivos que son específicos para los parásitos con la delección de los genes *pfhrp2/3* en el Periodo 2. De los genes identificados dentro de loci con barridos selectivos se puede destacar aquellos con funciones bien definidas. Entre los genes identificados específicamente en los parásitos con la doble delección se incluyen el gen *AP2-G*, que codifica para un factor de transcripción; y el gen *FP2A*, que codifica para la proteína falcipaina 2a.

El gen *AP2-G* es un gen de un solo exón que está involucrado en el compromiso del desarrollo a la forma sexual de los parásitos (109,110). Su expresión ocurre en etapas tempranas del proceso de diferenciación (111), siendo considerado por ello el desencadenante de la diferenciación a gametocitos; además de iniciar la activación de otros factores de transcripción que pueden dar lugar a la diferenciación en gametocitos masculinos o femeninos (111–113). Debido a la función en la diferenciación a gametocitos, la selección en este gen podría favorecer los procesos de transmisión de los parásitos con la delección de *pfhrp2/3*. Esto ha sido sugerido para el gen *gdv1* al encontrarse evidencia de selección en poblaciones de *P.falciparum* de África diferenciadas en el nivel de transmisión (15,67). Este gen facilita el inicio del compromiso sexual del parásito al permitir la activación del gen *AP2-G* específicamente (114,115). En este estudio se observó que la frecuencia del SNP *core* en la región del barrido selectivo, ha aumentado en frecuencia al punto de fijarse tanto en el grupo con la delección como sin la delección *pfhrp2/3*, a pesar que no se detectó barrido selectivo en el segundo grupo. Sin embargo, dentro del gen se observa que en el grupo de parásitos con la delección *pfhrp2/3*, distintas variantes habrían alcanzado una frecuencia de 1 en el Periodo 2, por lo que este grupo probablemente se podría haber favorecido por la presencia de dichas variantes en el gen.

Finalmente, *FP2A* es un gen de un solo exón que codifica para la falcipaína 2a, una cisteína proteasa producida por los parásitos de etapas sanguíneas como trofozoitos y esquizonte, y que participa en la degradación de la hemoglobina (116,117). Este gen ha sido asociado con reducida susceptibilidad a artemisinina *in vitro* y *ex vivo* al evaluarse el impacto de algunos SNPs en la susceptibilidad al fármaco (118–

120). Al ser la digestión de la hemoglobina un paso necesario en la activación del mecanismo de acción del fármaco, se ha descrito menor actividad enzimática cuando el gen es parte de un haplotipo con SNPs en el dominio de cisteína proteasa (121,122).

El cambio en la frecuencia del SNP core en el barrido selectivo identificado en *FP2A* demuestra su fijación en el tiempo en los parásitos con la delección *pfhrp2/3*. Sin embargo, a pesar que no se detectaron barrido selectivo en esta región en el grupo de parásitos sin la delección *pfhrp2/3*, también se vio un incremento en la frecuencia de dicho SNP, pero sin lograr fijarse. Esto último podría deberse a que ambos grupos de parásitos están sometidos a la misma presión por el esquema de tratamiento empleado contra la malaria, pero que la selección en el gen *FP2A* habría sido diferente en los parásitos con la delección de los genes *pfhrp2/3*, en comparación a los parásitos sin la delección, quienes también muestran un incremento en la frecuencia del SNP *core* pero sin llegar a fijarse. Igualmente, no se puede descartar que la diferencia pueda deberse a que la proteína cumple una función compensatoria en la digestión del grupo hemo en los parásitos con la delección de *pfhrp2/3*. Asimismo, se observa que en otras posiciones del gen algunos SNPs no sinónimos, aumentaron en frecuencia del Periodo 1 al 2, y estarían fijados en el Periodo 2. No obstante, lo mencionado para los genes con barridos selectivos son sugerencias que requieren de estudios complementarios que evalúen la dinámica en el tiempo de la diversidad genética, así estudios complementarios para evaluar los efectos estructurales y funcionales de las variantes identificadas en el gen de la falciparina, y comparar con lo reportado previamente para SNPs ubicados en regiones como el dominio cisteína proteasa y el pro-dominio de la proteína (120,121).

En contraste, en el caso de los parásitos sin la delección de los genes *pfhrp2/3* una señal específica fue en el gen *HSP101* que codifica para una proteína *heatshock* que participa en un complejo para la exportación proteica del parásito en la célula huésped (123,124). Asimismo, el gen LDH que codifica para el lactato deshidrogenasa se encontró en una región donde se identificó un barrido selectivo. Esta proteína es usada como blanco de las pruebas de diagnóstico rápido que se usan de forma alternativa a las que detectan la proteína *HRP2* (125).

En resumen, los hallazgos del presente estudio demuestran los cambios que ha sufrido la población de *Plasmodium falciparum* en la Amazonía peruana en el tiempo como consecuencia de las intervenciones del plan PAMAFRO. Además, los resultados del escaneo del genoma completo, sugieren regiones que explicarían la persistencia del parásito debido a diversos factores genéticos, entre ellos la evasión del sistema inmunitario del hospedero. De manera específica, se han identificado en el grupo de parásitos con la delección de *pfhrp2/3* barridos selectivos en loci con genes involucrados en la diferenciación a etapas sexuales y en un candidato a resistencia a artemisinina; lo que podría facilitar la transmisión de este grupo de parásitos. Se considera que esta información debe ser tomada en cuenta en las estrategias de vigilancia molecular de la especie en la región.

IX. CONCLUSIONES

- La muestra de *Plasmodium falciparum* evidencia un aumento significativo de la frecuencia de delección del gen *pfhrp2* con respecto a años anteriores, pero esto no ocurre para la delección del gen *pfhrp3*. Esto último sería debido a una diferencia específica en la comunidad del distrito de Nauta, donde el número de parásitos *pfhrp3* positivos ha aumentado.
- La población de *Plasmodium falciparum* de la Amazonía peruana ha sufrido una reducción de su complejidad por un cuello de botella ocasionado por las intervenciones del programa PAMAFRO ejecutado durante los años 2005 – 2010, y posteriormente la población se expandió, dando como resultado que la población más reciente sea clonal.
- Los parásitos con la delección de *pfhrp2/3* y los parásitos sin la delección están conformados por dos poblaciones clonales diferentes entre sí.
- La delección de los genes *pfhrp2* y *pfhrp3* no muestran evidencia de haber sufrido selección positiva.

X. LIMITACIONES

- Entre las limitaciones del presente estudio se puede mencionar el bajo número de muestras en el grupo de parásitos sin la delección, para el análisis genómico con muestras del Periodo 2.
- Otra limitación es que no se han empleado un número de muestras representativo de todos los distritos incluidos en el análisis, no obstante, no se observó alguna tendencia de *clustering* por lugar de origen.
- La alta cantidad de alelos raros característico de la poblacional clonal de *P. falciparum* en este estudio, dificultó la detección de selección positiva con las pruebas F_{st} y D de Tajima.
- Para los análisis se ha decidió incluir datos con un máximo de 25% de data faltante, tanto para los SNPs como para las muestras consideradas, con el fin de conservar la mayor cantidad de variantes y muestras para los análisis.
- La cobertura obtenida en las corridas de secuenciación realizadas en UPCH fue considerablemente menor a lo obtenido por MalariaGEN.

XI. RECOMENDACIONES

- Se recomienda que para estudios subsiguientes se incluyan mayor número de muestras por distrito.
- Se recomienda realizar la genotipificación con un protocolo que detecte tanto la delección del exón 1 como del exón 2 de los genes *pfhrp2* y *pfhrp3*, para proporcionar información más completa respecto al perfil de delección.
- Se recomienda evaluar la calidad e integridad de las muestras de ADN antes de la preparación de biblioteca, así como emplear, en la medida de lo posible, muestras de ADN de alta calidad y parasitemia para optimizar el rendimiento y calidad los *reads* y la cobertura.
- Se recomienda reforzar la evidencia de los hallazgos que sugieren que la persistencia de *Plasmodium falciparum* se debería a eventos de selección en genes *var*, involucrados en la interacción y evasión del sistema inmunitario del hospedero.
- Es recomendable realizar más estudios que permitan confirmar que el aumento en la frecuencia de los parásitos con la delección *pfhrp2/3* sería por barridos selectivos en regiones del genoma con genes como *AP2-G* y *FP2A*. *AP2-G* está involucrado en procesos biológicos de diferenciación hacia las etapas sexuales del parásito; *FP2A* en el metabolismo de hemoglobina, y en consecuencia en una posible alteración de manera indirecta de la susceptibilidad al tratamiento con artemisinina. Eventos de selección involucrando los procesos biológicos mencionados pueden favorecer la transmisión de este grupo de parásitos, así como también conferirle menor susceptibilidad al tratamiento.

- Se recomienda complementar los hallazgos con estudios de dinámica de la diversidad genética en los genes donde se encontraron los barridos selectivos.

XII. REFERENCIAS BIBLIOGRÁFICAS

1. Sala Situacional | Centro Nacional de Epidemiología, Prevención y Control de Enfermedades. CDC - Perú [Internet]. [citado 1 de diciembre de 2022]. Disponible en: <https://www.dge.gob.pe/salasisituacional/>
2. Phillips MA, Burrows JN, Manyando C, van Huijsduijnen RH, Van Voorhis WC, Wells TNC. Malaria. Nat Rev Dis Primer. 3 de agosto de 2017;3:17050.
3. Gamboa D, Ho MF, Bendezu J, Torres K, Chiodini PL, Barnwell JW, et al. A large proportion of *P. falciparum* isolates in the Amazon region of Peru lack *pfhrp2* and *pfhrp3*: implications for malaria rapid diagnostic tests. PloS One. 25 de enero de 2010;5(1):e8091.
4. Poti KE, Sullivan DJ, Dondorp AM, Woodrow CJ. HRP2: Transforming Malaria Diagnosis, but with Caveats. Trends Parasitol. febrero de 2020;36(2):112-26.
5. Akinyi S, Hayden T, Gamboa D, Torres K, Bendezu J, Abdallah JF, et al. Multiple genetic origins of histidine-rich protein 2 gene deletion in *Plasmodium falciparum* parasites from Peru. Sci Rep. 30 de septiembre de 2013;3:2797.
6. Mariella Quispe Carbajal. Análisis de la diversidad genética y estructura genética poblacional de *Plasmodium falciparum* con deleciones en los genes *Pfhrp2* y *Pfhrp3* en tres comunidades de la Amazonía Peruana [Tesis de Maestría]. [Lima]: Universidad Peruana Cayetano Heredia; 2017.
7. Bendezu J, Torres K, Villasis E, Incardona S, Bell D, Vinetz J, et al. Geographical distribution and genetic characterization of *pfhrp2* negative *Plasmodium falciparum* parasites in the Peruvian Amazon. PloS One. 2022;17(11):e0273872.
8. Valdivia HO, Anderson K, Smith D, Pasay C, Salas CJ, Braga G, et al. Spatiotemporal dynamics of *Plasmodium falciparum* histidine-rich protein 2 and 3 deletions in Peru. Sci Rep. 18 de noviembre de 2022;12(1):19845.
9. Dorado EJ, Okoth SA, Montenegro LM, Diaz G, Barnwell JW, Udhayakumar V, et al. Genetic Characterisation of *Plasmodium falciparum* Isolates with Deletion of the *pfhrp2* and/or *pfhrp3* Genes in Colombia: The Amazon Region, a Challenge for Malaria Diagnosis and Control. PLoS ONE [Internet]. 16 de septiembre de 2016 [citado 9 de agosto de 2020];11(9). Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5026348/>
10. Rachid Viana GM, Akinyi Okoth S, Silva-Flannery L, Lima Barbosa DR, Macedo de Oliveira A, Goldman IF, et al. Histidine-rich protein 2 (*pfhrp2*) and *pfhrp3* gene deletions in *Plasmodium falciparum* isolates from select sites in Brazil and Bolivia. PloS One. 2017;12(3):e0171150.

11. Góes L, Chamma-Siqueira N, Peres JM, Nascimento JM, Valle S, Arcanjo AR, et al. Evaluation of Histidine-Rich Proteins 2 and 3 Gene Deletions in *Plasmodium falciparum* in Endemic Areas of the Brazilian Amazon. *Int J Environ Res Public Health*. 26 de diciembre de 2020;18(1).
12. Beshir KB, Sepúlveda N, Bharmal J, Robinson A, Mwanguzi J, Busula AO, et al. *Plasmodium falciparum* parasites with histidine-rich protein 2 (*pfhrp2*) and *pfhrp3* gene deletions in two endemic regions of Kenya. *Sci Rep*. 07 de 2017;7(1):14718.
13. Soto-Calle V, Rosas-Aguirre A, Llanos-Cuentas A, Abatih E, DeDeken R, Rodriguez H, et al. Spatio-temporal analysis of malaria incidence in the Peruvian Amazon Region between 2002 and 2013. *Sci Rep*. 16 de 2017;7:40350.
14. Villena FE, Lizewski SE, Joya CA, Valdivia HO. Population genomics and evidence of clonal replacement of *Plasmodium falciparum* in the Peruvian Amazon. *Sci Rep*. 27 de octubre de 2021;11(1):21212.
15. Duffy CW, Assefa SA, Abugri J, Amoako N, Owusu-Agyei S, Anyorigiya T, et al. Comparison of genomic signatures of selection on *Plasmodium falciparum* between different regions of a country with high malaria endemicity. *BMC Genomics*. 16 de julio de 2015;16:527.
16. Amambua-Ngwa A, Jeffries D, Amato R, Worwui A, Karim M, Ceesay S, et al. Consistent signatures of selection from genomic analysis of pairs of temporal and spatial *Plasmodium falciparum* populations from The Gambia. *Sci Rep*. 26 de 2018;8(1):9687.
17. Feleke SM, Reichert EN, Mohammed H, Brhane BG, Mekete K, Mamo H, et al. *Plasmodium falciparum* is evolving to escape malaria rapid diagnostic tests in Ethiopia. *Nat Microbiol*. octubre de 2021;6(10):1289-99.
18. Rosas-Aguirre A, Guzman-Guzman M, Gamboa D, Chuquiyauri R, Ramirez R, Manrique P, et al. Micro-heterogeneity of malaria transmission in the Peruvian Amazon: a baseline assessment underlying a population-based cohort study. *Malar J*. 04 de 2017;16(1):312.
19. Rosas-Aguirre A, Guzman-Guzman M, Chuquiyauri R, Moreno M, Manrique P, Ramirez R, et al. Temporal and micro-spatial heterogeneity in transmission dynamics of co-endemic *Plasmodium vivax* and *Plasmodium falciparum* in two rural cohort populations in the Peruvian Amazon. *J Infect Dis*. 21 de agosto de 2020;
20. Watson OJ, Verity R, Ghani AC, Garske T, Cunningham J, Tshetu A, et al. Impact of seasonal variations in *Plasmodium falciparum* malaria transmission on the surveillance of *pfhrp2* gene deletions. *eLife*. 8:e40339.

21. Watson OJ, Slater HC, Verity R, Parr JB, Mwandagalirwa MK, Tshetu A, et al. Modelling the drivers of the spread of *Plasmodium falciparum* *hrp2* gene deletions in sub-Saharan Africa. *eLife*. 24 de agosto de 2017;6.
22. World malaria report 2021 [Internet]. [citado 1 de diciembre de 2022]. Disponible en: <https://www.who.int/publications-detail-redirect/9789240040496>
23. Carrasco-Escobar G, Gamboa D, Castro MC, Bangdiwala SI, Rodriguez H, Contreras-Mancilla J, et al. Micro-epidemiology and spatial heterogeneity of *P. vivax* parasitaemia in riverine communities of the Peruvian Amazon: A multilevel analysis. *Sci Rep*. 14 de 2017;7(1):8082.
24. Rosas-Aguirre A, Gamboa D, Manrique P, Conn JE, Moreno M, Lescano AG, et al. Epidemiology of *Plasmodium vivax* Malaria in Peru. *Am J Trop Med Hyg*. 28 de diciembre de 2016;95(6 Suppl):133-44.
25. Rovira-Vallbona E, Contreras-Mancilla JJ, Ramirez R, Guzmán-Guzmán M, Carrasco-Escobar G, Llanos-Cuentas A, et al. Predominance of asymptomatic and sub-microscopic infections characterizes the *Plasmodium* gametocyte reservoir in the Peruvian Amazon. *PLoS Negl Trop Dis*. julio de 2017;11(7):e0005674.
26. Ministerio de Salud. Resolución Ministerial N.º 244-2017-MINSA [Internet]. 2017 [citado 1 de diciembre de 2022]. Disponible en: <https://www.gob.pe/institucion/minsa/normas-legales/189846-244-2017-minsa>
27. Ministerio de Salud. Marco para la Eliminación de la Malaria [Internet]. 2017 [citado 1 de diciembre de 2022]. Disponible en: <http://bvs.minsa.gob.pe/local/MINSA/4253.pdf>
28. Resolución Ministerial N.º 034-2022-MINSA [Internet]. [citado 2 de enero de 2023]. Disponible en: <https://www.gob.pe/institucion/minsa/normas-legales/2723757-034-2022-minsa>
29. Guidelines for the Treatment of Malaria [Internet]. 3rd ed. Geneva: World Health Organization; 2015 [citado 15 de junio de 2020]. (WHO Guidelines Approved by the Guidelines Review Committee). Disponible en: <http://www.ncbi.nlm.nih.gov/books/NBK294440/>
30. Malaria diagnosis: memorandum from a WHO meeting. *Bull World Health Organ*. 1988;66(5):575-94.
31. Zheng Z, Cheng Z. Advances in Molecular Diagnosis of Malaria. *Adv Clin Chem*. 2017;80:155-92.
32. Wong YP, Othman S, Lau YL, Radu S, Chee HY. Loop-mediated isothermal amplification (LAMP): a versatile technique for detection of micro-organisms. *J Appl Microbiol*. marzo de 2018;124(3):626-43.

33. Serra-Casas E, Manrique P, Ding XC, Carrasco-Escobar G, Alava F, Gave A, et al. Loop-mediated isothermal DNA amplification for asymptomatic malaria detection in challenging field settings: Technical performance and pilot implementation in the Peruvian Amazon. *PloS One*. 2017;12(10):e0185742.
34. Vitti JJ, Grossman SR, Sabeti PC. Detecting natural selection in genomic data. *Annu Rev Genet*. 2013;47:97-120.
35. Walsh B, Lynch M. The Population Genetics of Selection [Internet]. *Evolution and Selection of Quantitative Traits*. Oxford University Press; [citado 20 de septiembre de 2020]. Disponible en: <https://oxford.universitypressscholarship.com/view/10.1093/oso/9780198830870.001.0001/oso-9780198830870-chapter-5>
36. Walsh B, Lynch M. Hitchhiking and Selective Sweeps [Internet]. *Evolution and Selection of Quantitative Traits*. Oxford University Press; [citado 20 de septiembre de 2020]. Disponible en: <https://oxford.universitypressscholarship.com/view/10.1093/oso/9780198830870.001.0001/oso-9780198830870-chapter-8>
37. Pavlidis P, Alachiotis N. A survey of methods and tools to detect recent and strong positive selection. *J Biol Res Thessalon Greece*. diciembre de 2017;24:7.
38. Pritchard JK, Pickrell JK, Coop G. The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation. *Curr Biol CB*. 23 de febrero de 2010;20(4):R208-15.
39. Walsh B, Lynch M. Using Molecular Data to Detect Selection: Signatures from Recent Single Events [Internet]. *Evolution and Selection of Quantitative Traits*. Oxford University Press; [citado 20 de septiembre de 2020]. Disponible en: <https://oxford.universitypressscholarship.com/view/10.1093/oso/9780198830870.001.0001/oso-9780198830870-chapter-9>
40. McCollum AM, Mueller K, Villegas L, Udhayakumar V, Escalante AA. Common origin and fixation of *Plasmodium falciparum dhfr* and *dhps* mutations associated with sulfadoxine-pyrimethamine resistance in a low-transmission area in South America. *Antimicrob Agents Chemother*. junio de 2007;51(6):2085-91.
41. Muiruri P, Juma DW, Ingasia LA, Chebon LJ, Opot B, Ngalah BS, et al. Selective sweeps and genetic lineages of *Plasmodium falciparum* multi-drug resistance (*pfmdr1*) gene in Kenya. *Malar J*. 30 de octubre de 2018;17(1):398.
42. Mobegi VA, Duffy CW, Amambua-Ngwa A, Loua KM, Laman E, Nwakanma DC, et al. Genome-wide analysis of selection on the malaria parasite *Plasmodium falciparum* in West African populations of differing infection endemicity. *Mol Biol Evol*. junio de 2014;31(6):1490-9.

43. Anderson TJ, Haubold B, Williams JT, Estrada-Franco JG, Richardson L, Mollinedo R, et al. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol Biol Evol.* octubre de 2000;17(10):1467-82.
44. Sullivan DJ, Gluzman IY, Goldberg DE. *Plasmodium* hemozoin formation mediated by histidine-rich proteins. *Science.* 12 de enero de 1996;271(5246):219-22.
45. Yang Y, Tang T, Feng B, Li S, Hou N, Ma X, et al. Disruption of *Plasmodium falciparum* histidine-rich protein 2 may affect haem metabolism in the blood stage. *Parasit Vectors.* 9 de diciembre de 2020;13(1):611.
46. Jani D, Nagarkatti R, Beatty W, Angel R, Slebodnick C, Andersen J, et al. HDP-a novel heme detoxification protein from the malaria parasite. *PLoS Pathog.* 25 de abril de 2008;4(4):e1000053.
47. Akinyi Okoth S, Abdallah JF, Ceron N, Adhin MR, Chandrabose J, Krishnalall K, et al. Variation in *Plasmodium falciparum* Histidine-Rich Protein 2 (*Pfhrp2*) and *Plasmodium falciparum* Histidine-Rich Protein 3 (*Pfhrp3*) Gene Deletions in Guyana and Suriname. *PloS One.* 2015;10(5):e0126805.
48. Fontecha G, Mejía RE, Banegas E, Ade MP, Mendoza L, Ortiz B, et al. Deletions of *pfhrp2* and *pfhrp3* genes of *Plasmodium falciparum* from Honduras, Guatemala and Nicaragua. *Malar J.* 31 de agosto de 2018;17(1):320.
49. Gupta H, Matambisso G, Galatas B, Cisteró P, Nhamussua L, Simone W, et al. Molecular surveillance of *pfhrp2* and *pfhrp3* deletions in *Plasmodium falciparum* isolates from Mozambique. *Malar J.* 16 de 2017;16(1):416.
50. Thomson R, Beshir KB, Cunningham J, Baiden F, Bharmal J, Bruxvoort KJ, et al. *pfhrp2* and *pfhrp3* Gene Deletions That Affect Malaria Rapid Diagnostic Tests for *Plasmodium falciparum*: Analysis of Archived Blood Samples From 3 African Countries. *J Infect Dis.* 26 de 2019;220(9):1444-52.
51. Agaba BB, Yeka A, Nsohya S, Arinaitwe E, Nankabirwa J, Opigo J, et al. Systematic review of the status of *pfhrp2* and *pfhrp3* gene deletion, approaches and methods used for its estimation and reporting in *Plasmodium falciparum* populations in Africa: review of published studies 2010-2019. *Malar J.* 6 de noviembre de 2019;18(1):355.
52. Shonhai A. The Role of Hsp70s in the Development and Pathogenicity of *Plasmodium* Species. En: Shonhai A, Blatch GL, editores. *Heat Shock Proteins of Malaria* [Internet]. Dordrecht: Springer Netherlands; 2014 [citado 26 de abril de 2021]. p. 47-69. Disponible en: https://doi.org/10.1007/978-94-007-7438-4_3

53. Cobb DW, Florentin A, Fierro MA, Krakowiak M, Moore JM, Muralidharan V. The Exported Chaperone PfHsp70x Is Dispensable for the *Plasmodium falciparum* Intraerythrocytic Life Cycle. *mSphere*. octubre de 2017;2(5).
54. Charnaud SC, Dixon MWA, Nie CQ, Chappell L, Sanders PR, Nebl T, et al. The exported chaperone Hsp70-x supports virulence functions for *Plasmodium falciparum* blood stage parasites. *PLoS ONE* [Internet]. 21 de julio de 2017 [citado 26 de abril de 2021];12(7). Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5521827/>
55. Kumar V, Behl A, Sharma R, Sharma A, Hora R. *Plasmodium* helical interspersed subtelomeric family-an enigmatic piece of the *Plasmodium* biology puzzle. *Parasitol Res*. octubre de 2019;118(10):2753-66.
56. Warncke JD, Vakonakis I, Beck HP. *Plasmodium* Helical Interspersed Subtelomeric (PHIST) Proteins, at the Center of Host Cell Remodeling. *Microbiol Mol Biol Rev MMBR*. diciembre de 2016;80(4):905-27.
57. Matesanz F, Téllez M del-Mar, Alcina A. The *Plasmodium falciparum* fatty acyl-CoA synthetase family (PfACS) and differential stage-specific expression in infected erythrocytes. *Mol Biochem Parasitol*. enero de 2003;126(1):109-12.
58. Bethke LL, Zilversmit M, Nielsen K, Daily J, Volkman SK, Ndiaye D, et al. Duplication, gene conversion, and genetic diversity in the species-specific acyl-CoA synthetase gene family of *Plasmodium falciparum*. *Mol Biochem Parasitol*. noviembre de 2006;150(1):10-24.
59. Baldeviano GC, Okoth SA, Arrospide N, Gonzalez RV, Sánchez JF, Macedo S, et al. Molecular Epidemiology of *Plasmodium falciparum* Malaria Outbreak, Tumbes, Peru, 2010-2012. *Emerg Infect Dis*. mayo de 2015;21(5):797-803.
60. Okoth SA, Chenet SM, Arrospide N, Gutierrez S, Cabezas C, Matta JA, et al. Molecular Investigation into a Malaria Outbreak in Cusco, Peru: *Plasmodium falciparum* BV1 Lineage is Linked to a Second Outbreak in Recent Times. *Am J Trop Med Hyg*. enero de 2016;94(1):128-31.
61. Nair S, Li X, Nkhoma SC, Anderson T. Fitness Costs of *pfhrp2* and *pfhrp3* Deletions Underlying Diagnostic Evasion in Malaria Parasites. *J Infect Dis*. 1 de noviembre de 2022;226(9):1637-45.
62. Awandare GA, Nyarko PB, Aniweh Y, Ayivor-Djanie R, Stoute JA. *Plasmodium falciparum* strains spontaneously switch invasion phenotype in suspension culture. *Sci Rep*. 10 de abril de 2018;8(1):5782.
63. Nyarko PB, Tarr SJ, Aniweh Y, Stewart LB, Conway DJ, Awandare GA. Investigating a *Plasmodium falciparum* erythrocyte invasion phenotype switch at the whole transcriptome level. *Sci Rep*. 14 de enero de 2020;10(1):245.
64. Centro Nacional de Epidemiología, Prevención y Control de Enfermedades [Internet]. [citado 15 de junio de 2020]. Disponible en:

https://www.dge.gob.pe/portal/index.php?option=com_content&view=article&id=664

65. Yalcindag E, Rougeron V, Elguero E, Arnathau C, Durand P, Brisse S, et al. Patterns of selection on *Plasmodium falciparum* erythrocyte-binding antigens after the colonization of the New World. *Mol Ecol.* abril de 2014;23(8):1979-93.
66. Berhane A, Anderson K, Mihreteab S, Gresty K, Rogier E, Mohamed S, et al. Major Threat to Malaria Control Programs by *Plasmodium falciparum* Lacking Histidine-Rich Protein 2, Eritrea. *Emerg Infect Dis.* marzo de 2018;24(3):462-70.
67. Mobegi VA, Duffy CW, Amambua-Ngwa A, Loua KM, Laman E, Nwakanma DC, et al. Genome-wide analysis of selection on the malaria parasite *Plasmodium falciparum* in West African populations of differing infection endemicity. *Mol Biol Evol.* junio de 2014;31(6):1490-9.
68. E.Z.N.A.® Blood DNA Mini Kit | Omega Bio-tek [Internet]. [citado 1 de diciembre de 2022]. Disponible en: <https://www.omegabiotek.com/product/e-z-n-a-blood-dna-mini-kit/>
69. Snounou G, Zhu X, Siripoon N, Jarra W, Thaithong S, Brown KN, et al. Biased distribution of *m*sp1 and *m*sp2 allelic variants in *Plasmodium falciparum* populations in Thailand. *Trans R Soc Trop Med Hyg.* 1999;93(4):369-74.
70. Mangold KA, Manson RU, Koay ESC, Stephens L, Regner M, Thomson RB, et al. Real-Time PCR for Detection and Identification of *Plasmodium* spp. *J Clin Microbiol.* mayo de 2005;43(5):2435-40.
71. Sundararaman SA, Plenderleith LJ, Liu W, Loy DE, Learn GH, Li Y, et al. Genomes of cryptic chimpanzee *Plasmodium* species reveal key evolutionary events leading to human malaria. *Nat Commun.* 22 de marzo de 2016;7:11078.
72. MAVILA SORIA A. ESTANDARIZACIÓN DE UN MÉTODO DE ENRIQUECIMIENTO DE ADN GENÓMICO DE *Plasmodium falciparum*. [Iquitos]: Universidad Nacional de la Amazonía Peruana (UNAP); 2019.
73. Nextera XT DNA Library Prep Kit | Sequence small genomes, plasmids, cDNA [Internet]. [citado 29 de noviembre de 2022]. Disponible en: <https://www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/nextera-xt-dna.html>
74. Andrews, S. FastQC A Quality Control tool for High Throughput Sequence Data [Internet]. 2010 [citado 1 de diciembre de 2022]. Disponible en: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
75. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinforma Oxf Engl.* 1 de octubre de 2016;32(19):3047-8.

76. BBTools [Internet]. DOE Joint Genome Institute. [citado 1 de diciembre de 2022]. Disponible en: <https://jgi.doe.gov/data-and-tools/software-tools/bbtools/>
77. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinforma Oxf Engl.* 1 de agosto de 2014;30(15):2114-20.
78. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinforma Oxf Engl.* 1 de marzo de 2010;26(5):589-95.
79. Picard Tools - By Broad Institute [Internet]. [citado 1 de diciembre de 2022]. Disponible en: <https://broadinstitute.github.io/picard/>
80. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinforma Oxf Engl.* 15 de marzo de 2010;26(6):841-2.
81. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* septiembre de 2010;20(9):1297-303.
82. Scaling accurate genetic variant discovery to tens of thousands of samples | bioRxiv [Internet]. [citado 1 de diciembre de 2022]. Disponible en: <https://www.biorxiv.org/content/10.1101/201178v2>
83. Hard-filtering germline short variants [Internet]. GATK. [citado 1 de diciembre de 2022]. Disponible en: <https://gatk.broadinstitute.org/hc/en-us/articles/360035890471-Hard-filtering-germline-short-variants>
84. Otto TD, Böhme U, Sanders M, Reid A, Bruske EI, Duffy CW, et al. Long read assemblies of geographically dispersed *Plasmodium falciparum* isolates reveal highly structured subtelomeres. *Wellcome Open Res.* 2018;3:52.
85. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics.* 1 de agosto de 2011;27(15):2156-8.
86. moimix: an R package for assessing clonality in high-throughput sequencing data | Zenodo [Internet]. [citado 1 de diciembre de 2022]. Disponible en: <https://zenodo.org/record/58257#.Y4ldGn3MLIU>
87. Assefa SA, Preston MD, Campino S, Ocholla H, Sutherland CJ, Clark TG. estMOI: estimating multiplicity of infection using parasite deep sequencing data. *Bioinforma Oxf Engl.* 1 de mayo de 2014;30(9):1292-4.
88. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* septiembre de 2007;81(3):559-75.

89. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* septiembre de 2009;19(9):1655-64.
90. Francis RM. pophelper: an R package and web app to analyse and visualize population structure. *Mol Ecol Resour.* enero de 2017;17(1):27-32.
91. scikit-allel [Internet]. Centre for Genomics and Global Health; 2022 [citado 1 de diciembre de 2022]. Disponible en: <https://github.com/cggh/scikit-allel>
92. Bhatia G, Patterson N, Sankararaman S, Price AL. Estimating and interpreting FST: the impact of rare variants. *Genome Res.* septiembre de 2013;23(9):1514-21.
93. Gautier M, Klassmann A, Vitalis R. rehh 2.0: a reimplementación of the R package rehh to detect positive selection from haplotype structure. *Mol Ecol Resour.* enero de 2017;17(1):78-90.
94. Gautier M, Vitalis R. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinforma Oxf Engl.* 15 de abril de 2012;28(8):1176-7.
95. Reich D, Price AL, Patterson N. Principal component analysis of genetic data. *Nat Genet.* mayo de 2008;40(5):491-2.
96. Abegaz F, Chaichoompu K, Génin E, Fardo DW, König IR, Mahachie John JM, et al. Principals about principal components in statistical genetics. *Brief Bioinform.* 27 de noviembre de 2019;20(6):2200-16.
97. Cheng Q, Gatton ML, Barnwell J, Chiodini P, McCarthy J, Bell D, et al. *Plasmodium falciparum* parasites lacking histidine-rich protein 2 and 3: a review and recommendations for accurate reporting. *Malar J.* 22 de julio de 2014;13:283.
98. Góes L, Chamma-Siqueira N, Peres JM, Nascimento JM, Valle S, Arcanjo AR, et al. Evaluation of Histidine-Rich Proteins 2 and 3 Gene Deletions in *Plasmodium falciparum* in Endemic Areas of the Brazilian Amazon. *Int J Environ Res Public Health.* 26 de diciembre de 2020;18(1).
99. Costa GL, Mascarenhas MEP, Martin TOG, Fortini LG, Louzada J, Pereira DB, et al. A Comprehensive Analysis of the Genetic Diversity of *Plasmodium falciparum* Histidine-Rich Protein 2 (PfHRP2) in the Brazilian Amazon. *Front Cell Infect Microbiol.* 2021;11:742681.
100. Kattenberg JH, Fernandez-Miño C, van Dijk NJ, Llacsahuanga Allcca L, Guetens P, Valdivia HO, et al. Malaria molecular surveillance in the Peruvian Amazon with a novel highly multiplexed *Plasmodium falciparum* Ampliseq assay. medRxiv. 1 de enero de 2022;2021.11.12.21266245.
101. Browning SR, Browning BL. Identity by descent between distant relatives: detection and applications. *Annu Rev Genet.* 2012;46:617-33.

102. ORGANISMO ANDINO DE SALUD, Convenio Hipólito Unanue. INFORME DE RESULTADOS PAMAFRO PERÚ. PROYECTO PAMAFRO – PERÚ;
103. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature. 3 de octubre de 2002;419(6906):498-511.
104. Tibúrcio M, Silvestrini F, Bertuccini L, Sander AF, Turner L, Lavstsen T, et al. Early gametocytes of the malaria parasite *Plasmodium falciparum* specifically remodel the adhesive properties of infected erythrocyte surface. Cell Microbiol. abril de 2013;15(4):647-59.
105. Biggs BA, Goozé L, Wycherley K, Wollish W, Southwell B, Leech JH, et al. Antigenic variation in *Plasmodium falciparum*. Proc Natl Acad Sci U S A. 15 de octubre de 1991;88(20):9171-4.
106. Scherf A, Hernandez-Rivas R, Buffet P, Bottius E, Benatar C, Pouvelle B, et al. Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of var genes during intra-erythrocytic development in *Plasmodium falciparum*. EMBO J. 15 de septiembre de 1998;17(18):5418-26.
107. Kumar A, Ghosh DK, Ali J, Ranjan A. Characterization of Lipid Binding Properties of *Plasmodium falciparum* Acyl-Coenzyme A Binding Proteins and Their Competitive Inhibition by Mefloquine. ACS Chem Biol. 17 de mayo de 2019;14(5):901-15.
108. Kumar A, Ghosh DK, Ranjan A. Differential Stabilities of Mefloquine-Bound Human and *Plasmodium falciparum* Acyl-CoA-Binding Proteins. ACS Omega. 11 de enero de 2021;6(3):1883-93.
109. Kafsack BFC, Rovira-Graells N, Clark TG, Bancells C, Crowley VM, Campino SG, et al. A transcriptional switch underlies commitment to sexual development in malaria parasites. Nature. 13 de marzo de 2014;507(7491):248-52.
110. Sinha A, Hughes KR, Modrzynska KK, Otto TD, Pfander C, Dickens NJ, et al. A cascade of DNA-binding proteins for sexual commitment and development in *Plasmodium*. Nature. 13 de marzo de 2014;507(7491):253-7.
111. Yuda M, Kaneko I, Murata Y, Iwanaga S, Nishi T. Mechanisms of triggering malaria gametocytogenesis by *AP2-G*. Parasitol Int. octubre de 2021;84:102403.
112. Yuda M, Kaneko I, Iwanaga S, Murata Y, Kato T. Female-specific gene regulation in malaria parasites by an AP2-family transcription factor. Mol Microbiol. enero de 2020;113(1):40-51.
113. Prajapati SK, Ayanful-Torgby R, Pava Z, Barbeau MC, Acquah FK, Cudjoe E, et al. The transcriptome of circulating sexually committed *Plasmodium*

- falciparum* ring stage parasites forecasts malaria transmission potential. Nat Commun. 2 de diciembre de 2020;11(1):6159.
114. Brancucci NMB, Bertschi NL, Zhu L, Niederwieser I, Chin WH, Wampfler R, et al. Heterochromatin protein 1 secures survival and transmission of malaria parasites. Cell Host Microbe. 13 de agosto de 2014;16(2):165-76.
 115. Filarsky M, Fraschka SA, Niederwieser I, Brancucci NMB, Carrington E, Carrió E, et al. GDV1 induces sexual commitment of malaria parasites by antagonizing HP1-dependent gene silencing. Science. 16 de marzo de 2018;359(6381):1259-63.
 116. Rosenthal PJ, McKerrow JH, Aikawa M, Nagasawa H, Leech JH. A malarial cysteine proteinase is necessary for hemoglobin degradation by *Plasmodium falciparum*. J Clin Invest. noviembre de 1988;82(5):1560-6.
 117. Sijwali PS, Koo J, Singh N, Rosenthal PJ. Gene disruptions demonstrate independent roles for the four falcipain cysteine proteases of *Plasmodium falciparum*. Mol Biochem Parasitol. noviembre de 2006;150(1):96-106.
 118. Xie SC, Dogovski C, Hanssen E, Chiu F, Yang T, Crespo MP, et al. Haemoglobin degradation underpins the sensitivity of early ring stage *Plasmodium falciparum* to artemisinins. J Cell Sci. 15 de enero de 2016;129(2):406-16.
 119. Klonis N, Crespo-Ortiz MP, Bottova I, Abu-Bakar N, Kenny S, Rosenthal PJ, et al. Artemisinin activity against *Plasmodium falciparum* requires hemoglobin uptake and digestion. Proc Natl Acad Sci U S A. 12 de julio de 2011;108(28):11405-10.
 120. Tumwebaze PK, Conrad MD, Okitwi M, Orena S, Byaruhanga O, Katairo T, et al. Decreased susceptibility of *Plasmodium falciparum* to both dihydroartemisinin and lumefantrine in northern Uganda. Nat Commun. 26 de octubre de 2022;13:6353.
 121. Siddiqui FA, Cabrera M, Wang M, Brashear A, Kemirembe K, Wang Z, et al. *Plasmodium falciparum* Falcipain-2a Polymorphisms in Southeast Asia and Their Association With Artemisinin Resistance. J Infect Dis. 2 de julio de 2018;218(3):434-42.
 122. Woodley CM, Amado PSM, Cristiano MLS, O'Neill PM. Artemisinin inspired synthetic endoperoxide drug candidates: Design, synthesis, and mechanism of action studies. Med Res Rev. noviembre de 2021;41(6):3062-95.
 123. Beck JR, Muralidharan V, Oksman A, Goldberg DE. PTEX component HSP101 mediates export of diverse malaria effectors into host erythrocytes. Nature. 31 de julio de 2014;511(7511):592-5.

124. Elsworth B, Matthews K, Nie CQ, Kalanon M, Charnaud SC, Sanders PR, et al. PTEX is an essential nexus for protein export in malaria parasites. *Nature*. 31 de julio de 2014;511(7511):587-91.
125. Kyabayinze DJ, Zongo I, Cunningham J, Gatton M, Angutoko P, Ategeka J, et al. HRP2 and pLDH-Based Rapid Diagnostic Tests, Expert Microscopy, and PCR for Detection of Malaria Infection during Pregnancy and at Delivery in Areas of Varied Transmission: A Prospective Cohort Study in Burkina Faso and Uganda. *PloS One*. 2016;11(7):e0156954.

XIII. ANEXOS

ANEXO 1. Secuencia de primers para caracterización de deleciones de *pfhrp2*, *pfhrp3* y flanqueantes

Gen	Secuencia 5' à '3	
<i>pfhrp2</i>	F	5' GAT TAT TAC ACG AAA CTC AAG C 3'
	R	5' AAT AAA TTT AAT GGC GTA GGC 3'
<i>pfhrp3</i>	F	5' AAA TAA GAG ATT ATT ACA CGA AAG 3'
	R	5' TGGTGTAAGTGATGCGTAGT 3'
PF3D7_0831900 (MAL7P1.230)	F	5' TAT GAA CGC AAT TTA AGT GAG GCA G 3'
	R	5' TAT CCA ATC CTT CCT TTG CAA CAC C 3'
PF3D7_0831700 (MAL7P1.228)	F	5' AGA CAA GCT ACC AAA GAT GCA GGT G 3'
	R	5' TAA ATG TGT ATC TCC TGA GGT AGC 3'
PF3D7_1372400 (MAL13P1.485)	F	5' TTG AGT GCA ATG ATG AGT GGA G 3'
	R	5' AAA TCA TTT CCT TTT ACA CTA GTG C 3'
PF3D7_1372100 (MAL13P1.475)	F	5' TTC ATG AGT AGATGTCCTAGG AG 3'
	R	5' TCG TAC AAT TCA TCA TAC TCA CC 3'
<i>GLURP</i>	F	5' TGA ATT TGA AGA TGT TCA CAC TGA AC 3'
	R	5' GTGGAATTGCTTTTTCTTCAACAACACTAA 3'
<i>MSP1</i>	F	5' AAG CTT TAG AAG ATG CAG TAT TGA C 3'
	R	5' ATT CAT TAA TTT CTT CAT ATC CAT C 3'
<i>MSP2</i>	F	5' ATG AAG GTA ATT AAA ACA TTG TCT ATT ATA 3'
	R	5' CTT TGT TAC CAT CGG TAC ATT CTT 3'

ANEXO 2. Programas de amplificación

a. Programas de amplificación de genes control

		<i>MSP1</i>	<i>MSP2</i>	<i>GLURP</i>
Paso 1	Denaturación inicial	95°C x 3min	95°C x 3min	95°C x 3min
Paso 2	Denaturación	95°C x 30sec	95°C x 30sec	94°C x 50sec
Paso 3	Hibridación	55°C x 30sec	58°C x 2min	58°C x 30sec
Paso 4	Extensión	72°C x 45sec	72°C x 2min	72°C x 2min
Paso 5	Ciclos	2-4 per 35 ciclos	2-4 per 35 ciclos	2-4 per 35 ciclos
Paso 6	Hibridación Final	55°C x 30sec	58°C x 2min	58°C x 2min
Paso 7	Extensión final	72°C x 5min	72°C x 5min	72°C x 5min

b. Programas de amplificación de genes *pfhrp2/3* y flanqueantes

		<i>Pfhrp2</i> - exon2	<i>Pfhrp3</i> - exon2	PF3D7_0831700 (MAL7P1_228)	PF3D7_0831900 (MAL7P1_230)	PF3D7_1372100 (MAL7P1_475)	PF3D7_1372400 (MAL7P1_485)
Paso 1	Denaturación inicial	95°C x 4min	95°C x 4min	95°C x 5min	95°C x 5min	95°C x 5min	95°C x 5min
Paso 2	Denaturación	95°C x 50sec	95°C x 50sec	95°C x 30sec	94°C x 30sec	94°C x 35sec	94°C x 35sec
Paso 3	Hibridación	56.3°C x 30sec	54°C x 30sec	60°C x 30sec	65°C x 30sec	61°C x 40sec	61°C x 40sec
Paso 4	Extensión	72°C x 1min	72°C x 1min	68°C x 1min	68°C x 1min	72°C x 40sec	72°C x 40sec
Paso 5	Ciclos	2-4 per 35 ciclos	2-4 per 35 ciclos	2-4 per 35 ciclos	2-4 per 35 ciclos	2-4 per 35 ciclos	2-4 per 35 ciclos
Paso 6	Extensión final	72° x 10min	72° x 10min	68°C x 5min	68°C x 6min	72°C x 7min	72°C x 7min

ANEXO 3. Protocolo detallado de Secuenciación de Genoma Completo de *Plasmodium falciparum*

13.1.1.1.1.1. Tagmentación:

Los productos de sWGA se diluyeron a la concentración de 2 ng/μL de ADN. En una cabina de bioseguridad, se adicionó en una placa de PCR 10 μL de Tagment DNA buffer (TD) y 5 μL de ADN, luego se agregó 5 μL de Amplicon Tagment Mix (ATM). Se mezcló por pipeteo x10 veces y se selló la placa para ser centrifugada por 1 minuto a un máximo de 280 g. La placa se colocó en un termociclador Mastercycler Nexus (Eppendorf), y se incubó a 55°C por 5 minutos y luego un paso a 10°C.

Una vez cumplido el tiempo de incubación y cuando la temperatura alcanzó los 10°C, se procedió a retirar la placa inmediatamente y se agregó 5 μL de Neutralize Tagment Buffer (NT). Se homogenizó por pipeteo, se selló la placa para centrifugar y luego se dejó incubando a temperatura ambiente por 5 minutos.

13.1.1.1.1.2. Amplificación de bibliotecas

Se adicionó 5 μL de los adaptadores de *index i5* e *i7* a cada muestra. Se elaboraron combinaciones diferentes para cada muestra (Figura 6). Paso seguido, se agregó 15 μL de Nextera PCR Master Mix (NPM), la mezcla se homogenizó por pipeteo, se selló la placa y se centrifugó.

El protocolo de amplificación de bibliotecas consistió en 95°C por 30 segundos, seguido de 12 ciclos de 95°C por 10 segundos, 55°C por 30 segundos y 72°C por 30 segundos, y un paso final a 72°C por 5 minutos, para luego permanecer a 10°C.

13.1.1.1.3. Lavado de las bibliotecas

Para el lavado de las bibliotecas se empleó el reactivo Ampure XP. La placa se centrifugó por 1 minuto a 280 x g para coleccionar el contenido que pueda haberse adherido a las paredes o la cubierta de la placa.

Se agregaron 30 μL del reactivo a cada pocillo con bibliotecas. La placa se selló y se colocó en un agitador a 1800 rpm por 2 minutos. Adicionalmente, se homogenizó por pipeteo suave. La mezcla se dejó incubar por 5 minutos a temperatura ambiente. Luego la placa fue colocada en la base magnética hasta que la solución se aclaró. Se retiró el sobrenadante y se agregó 200 μL de etanol al 80% preparado al momento, se dejó incubar por 30 segundos, se eliminó el sobrenadante y se repitió este paso. Los residuos de etanol se eliminaron con la ayuda de una pipeta de 10 μL , para luego dejar secar la placa por 7-8 minutos. Se retiró la placa del soporte magnético y se agregó 52.5 μL de Buffer de re-suspensión (RSB). Se homogenizó por agitación a 1200 rpm por 2 minutos. La mezcla se colocó nuevamente en el soporte magnético hasta que la solución se aclaró, para finalmente recuperar 50 μL del sobrenadante en un *strip* de 8 pocillos.

Las bibliotecas se cuantificaron empleando el kit Qubit™ dsDNA HS Assay (Invitrogen).

13.1.1.1.4. Dilución y *pooling* de las bibliotecas

Se realizó el cálculo de la cantidad de bibliotecas y buffer RSB necesario para diluir las bibliotecas a la concentración de 2 nM. Se empleó la fórmula proporcionada por el fabricante para obtener la concentración molar inicial, reemplazando el valor obtenido de la cuantificación y como tamaño promedio se usó 1000 pb:

$$\frac{ng/\mu L \times 10^6}{660 \frac{g}{mol} \times \text{tamaño promedio de librería (bp)}} = \text{Molaridad (nM)}$$

El *pool* se preparó mezclando las bibliotecas en un microtubo de 1.5 mL a la concentración de 2 nM cada una.

13.1.1.1.5. Denaturación de bibliotecas y dilución final

La denaturación de las bibliotecas es a la vez una segunda dilución de las bibliotecas de 2 nM a 10 pM.

Para esto, en un microtubo de 1.5 mL se colocó 5 µL de bibliotecas y se le agregó 5 µL de NaOH 0.1 N, preparado en el momento. La mezcla se homogenizó por pipeteo y luego de 5 minutos se neutralizó la reacción adicionando 990 µL del buffer HT1.

La concentración final del *pool* denaturado de bibliotecas que se cargará en el equipo de secuenciación depende del kit de secuenciación usado. En el caso de la corrida realizada con el MiSeq reagent kit v2 la concentración final fue de 8 pM, y para las corridas usando el MiSeq reagent kit v3 la concentración final fue de 10 pM. Para dilución se preparó un volumen final de 660 µL agregando buffer HT1, además se incorporó como control el genoma del fago PhiX al 1%.

13.1.1.1.6. Secuenciación

La secuenciación de las bibliotecas se realizó empleando el equipo MiSeq de Illumina y se emplearon los componentes de los MiSeq reagent kit v2 o v3. Se prepararon los archivos necesarios (*Sample sheet* y *sample map*) con la información de las muestras procesadas y los reactivos que se utilizaron, se empleó el programa Illumina Experiment Manager 1.19.1.

Previo a la carga en el equipo, el cartucho del kit se dejó a descongelar dentro de un recipiente con agua por ~2 horas. Una vez descongelado se cargó el volumen de 660 μ L del pool de bibliotecas y se cargó en el equipo, cuando el pool no se cargó inmediatamente el cartucho fue colocado a 4°C protegido de la luz hasta el momento de la carga del pool de bibliotecas. De igual forma, la *flowcell* del kit fue retirada de su envase para su lavado con agua miliQ, con el fin de eliminar los residuos de la solución de almacenamiento.

ANEXO 4. Modificaciones al protocolo de Secuenciación de Genoma

Completo

a. Primera corrida

El primer grupo de muestras secuenciado fue enriquecido mediante el protocolo de sWGA de una sola ronda. Se siguieron los pasos de WGS descritos previamente. Los resultados de la secuenciación fueron evaluados para las decisiones respecto a las siguientes corridas. Se evaluaron las métricas de cobertura y porcentaje de reads alineados con el genoma de referencia.

b. Segunda corrida

El segundo grupo de muestras fue ejecutado en un número de 12 debido a la mayor capacidad del kit MiSeq v3. Las muestras empleadas fueron amplificadas utilizando el protocolo de sWGA de dos rondas. Igualmente, se procedió al análisis de los datos para la decisión de las corridas posteriores.

c. Tercera corrida

Con el fin de mejorar el rendimiento de la data generada, se redujo el número de muestras a 10. Asimismo, se trabajó con muestras de ADN recién extraídas para lidiar con el problema de la calidad de la corrida. En esta corrida, seis muestras fueron enriquecidas por dos rondas de amplificación genómica y cuatro por solo una. Esta modificación se hizo con la finalidad de evaluar la capacidad del protocolo de sWGA de enriquecer el genoma parasitario usando solo una ronda. Para ello, en base a los resultados de la corrida 1, se seleccionó un valor máximo de Cq de 24, a partir de los resultados de qPCR de las muestras de dicha corrida.

d. Corridas cuatro y cinco

Para sobrellevar el problema de calidad se procedió a evaluar las muestras de ADN por espectrofotometría y la integridad mediante electroforesis en agarosa al 0.7%. Se adicionó un paso de lavado con beads de Ampure XP al producto final de sWGA a fin de eliminar residuos orgánicos.

Los productos de sWGA lavados, fueron utilizados siguiendo los pasos de secuenciación descritos. Se procuró el manejo rápido de las muestras durante los pasos de digestión enzimática mediante el uso de la pipeta multicanal en la quinta corrida. Se ajustó la carga final del pool de bibliotecas a 10 pM para mejorar el rendimiento de la secuenciación.

ANEXO 5. Perfil de las deleciones de los genes *pfhrp2*, *pfhrp3* y genes flaqueantes

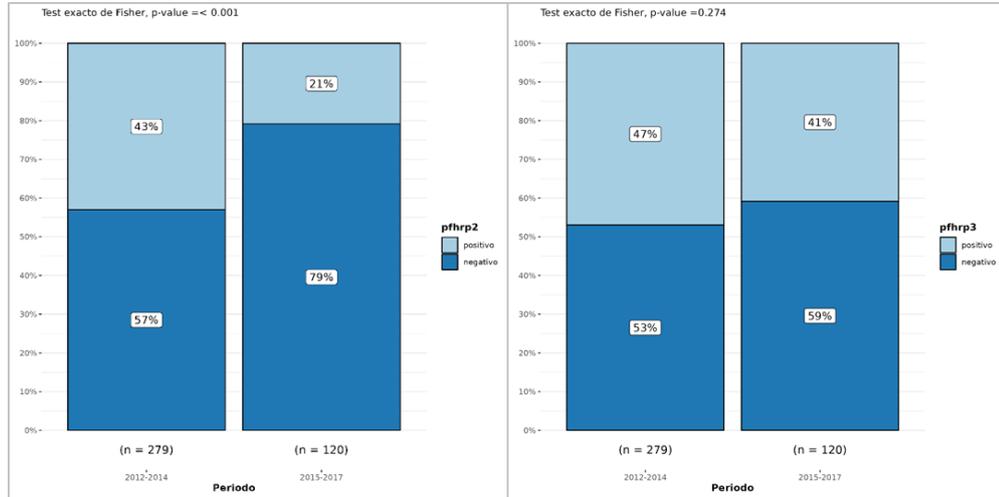
a. Proporción de la deleción de los genes *pfhrp2* y *pfhrp3*

Perfil	n	Proporción	IC 95
<i>pfhrp2</i> +	40	23.3	[17.3 - 30.4]
<i>pfhrp2</i> -	132	76.7	[69.6 - 82.7]
<i>pfhrp3</i> +	63	36.6	[29.5 - 44.3]
<i>pfhrp3</i> -	109	63.4	[55.6 - 70.5]
<i>pfhrp2/3</i> + / +	37	21.5	[15.8 - 28.5]
<i>pfhrp2/3</i> - / -	106	61.6	[53.9 - 68.8]
<i>pfhrp2/3</i> + / -	3	1.7	[0.4 - 5.4]
<i>pfhrp2/3</i> - / +	26	15.1	[10.3 - 21.5]

b. Tabla resumen con los perfiles de las deleciones de los genes *pfhrp2*,
pfhrp3 y los genes flanqueantes

Perfil	n	Proporción	IC 95
PF3D7_0831700 / pfhrp2 / PF3D7_0831900			
- / - / -	3	1.8	[0 - 8.6]
- / - / +	1	0.6	[0.000 - 7.4]
+ / - / -	122	71.3	[64.9 - 78.1]
+ / - / +	5	2.9	[0 - 9.7]
+ / + / -	1	0.6	[0 - 7.4]
+ / + / +	39	22.8	[16.4 - 29.6]
PF3D7_1372100 / pfhrp3 / PF3D7_1372400			
- / - / -	6	3.5	[0 - 1.7]
- / - / +	95	55.6	[48.5 - 63.7]
- / + / -	1	0.6	[0 - 8.8]
- / + / +	1	0.6	[0 - 8.8]
+ / - / -	2	1.2	[0 - 9.3]
+ / - / +	5	2.9	[0 - 11.1]
+ / + / +	61	35.7	[28.7 - 43.8]
PF3D7_0831700 / pfhrp2 / PF3D7_0831900 / PF3D7_1372100 / pfhrp3 / PF3D7_1372400			
- / - / - - / - / -	2	1.2	[0 - 9.1]
- / - / - + / - / +	1	0.6	[0 - 8.5]
- / - / + + / - / +	1	0.6	[0 - 8.5]
+ / - / - - / - / -	3	1.8	[0 - 9.6]
+ / - / - - / - / +	93	54.4	[47.4 - 62.3]
+ / - / - - / + / +	1	0.6	[0 - 8.5]
+ / - / - + / + / +	25	14.6	[7.6 - 22.5]
+ / - / + - / - / -	1	0.6	[0 - 8.5]
+ / - / + + / - / -	1	0.6	[0 - 8.5]
+ / - / + + / - / +	3	1.8	[0 - 9.6]
+ / + / - + / - / -	1	0.6	[0 - 8.5]
+ / + / + - / - / +	2	1.2	[0 - 9.1]
+ / + / + - / + / -	1	0.6	[0 - 8.5]
+ / + / + + / + / +	36	21.1	[14.0 - 28.9]

ANEXO 6. Comparación de la proporción de la delección de los genes *pfhrp2* y *pfhrp3* en muestras colectadas en el periodo 2012 – 2014 y 2015 - 2017



Para este análisis se emplearon los datos obtenidos en la tesis de Mariela Quispe (6), debido a que empleó el mismo protocolo para la genotipificación de las delecciones. Se combinaron los datos de las muestras que correspondieran al periodo 2012 – 2014 del presente estudio, con los datos del estudio previo de Quispe; y se realizó la comparación con los datos de las muestras restantes (2015 – 2017).

ANEXO 7. Año de colecta y perfil de la delección del gen *pfhrp2* de los genomas incluidos

Año	<i>pfhrp2</i> positivo	<i>pfhrp2</i> negativo
2006	0	1
2007	1	0
2008	5	5
2009	2	1
2010	3	2
2011	8	5
2012	0	1
2013	3	4
2014	1	1
2015	0	9
2016	17	21
2018	1	9

ANEXO 8. Principales SNPs contribuyentes a cada componente principal del PCA

a. Primeros 20 SNPs de mayor peso para el PC1

Cromosoma	SNP	PC1
Pf3D7_11_v3	Pf3D7_11_v3_1845877_A_G	2.88362
Pf3D7_11_v3	Pf3D7_11_v3_1858572_A_T	2.88362
Pf3D7_13_v3	Pf3D7_13_v3_1125789_C_T	2.88362
Pf3D7_14_v3	Pf3D7_14_v3_2609498_A_T	2.86083
Pf3D7_02_v3	Pf3D7_02_v3_667742_T_C	2.85506
Pf3D7_11_v3	Pf3D7_11_v3_1815802_G_A	2.85506
Pf3D7_13_v3	Pf3D7_13_v3_1265623_T_C	2.85205
Pf3D7_12_v3	Pf3D7_12_v3_712024_A_C	2.85049
Pf3D7_14_v3	Pf3D7_14_v3_303959_T_C	2.84407
Pf3D7_14_v3	Pf3D7_14_v3_2476165_A_G	2.84122
Pf3D7_14_v3	Pf3D7_14_v3_508547_G_A	2.82825
Pf3D7_05_v3	Pf3D7_05_v3_503740_T_C	2.8271
Pf3D7_11_v3	Pf3D7_11_v3_1245173_T_C	2.81166
Pf3D7_10_v3	Pf3D7_10_v3_433422_G_A	2.80624
Pf3D7_14_v3	Pf3D7_14_v3_188509_A_G	2.79877
Pf3D7_14_v3	Pf3D7_14_v3_564690_G_A	2.79758
Pf3D7_14_v3	Pf3D7_14_v3_852041_C_A	2.79204
Pf3D7_08_v3	Pf3D7_08_v3_264697_A_T	2.78683
Pf3D7_02_v3	Pf3D7_02_v3_257473_G_A	2.78112
Pf3D7_13_v3	Pf3D7_13_v3_939288_C_A	2.77221

b. Primeros 20 SNPs de mayor peso para el PC2

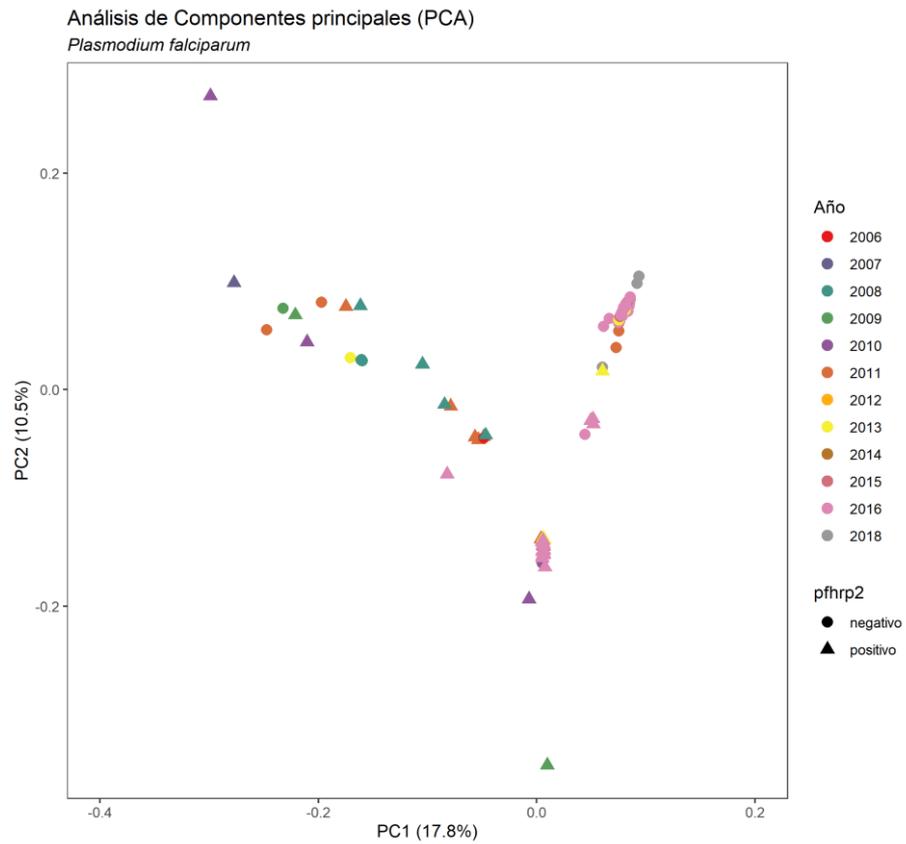
Cromosoma	SNP	PC2
Pf3D7_09_v3	Pf3D7_09_v3_714147_C_G	3.50953
Pf3D7_04_v3	Pf3D7_04_v3_895751_C_A	3.43446
Pf3D7_13_v3	Pf3D7_13_v3_2759719_A_G	3.40699
Pf3D7_13_v3	Pf3D7_13_v3_2770580_A_G	3.3519
Pf3D7_13_v3	Pf3D7_13_v3_422254_T_A	3.29779
Pf3D7_11_v3	Pf3D7_11_v3_750688_C_G	3.18658
Pf3D7_14_v3	Pf3D7_14_v3_365809_T_A	3.16733
Pf3D7_02_v3	Pf3D7_02_v3_86870_C_T	3.15962
Pf3D7_14_v3	Pf3D7_14_v3_1974547_G_C	3.15066
Pf3D7_13_v3	Pf3D7_13_v3_756010_A_G	3.14781
Pf3D7_04_v3	Pf3D7_04_v3_821816_A_C	3.13448
Pf3D7_06_v3	Pf3D7_06_v3_251219_T_A	3.1336
Pf3D7_04_v3	Pf3D7_04_v3_798595_G_T	3.10956
Pf3D7_13_v3	Pf3D7_13_v3_434220_A_T	3.10495
Pf3D7_13_v3	Pf3D7_13_v3_939288_C_A	3.09602
Pf3D7_06_v3	Pf3D7_06_v3_389832_G_A	3.09359
Pf3D7_03_v3	Pf3D7_03_v3_142239_T_C	3.0906
Pf3D7_05_v3	Pf3D7_05_v3_1124184_G_A	3.08916
Pf3D7_13_v3	Pf3D7_13_v3_2232592_C_A	3.08067
Pf3D7_05_v3	Pf3D7_05_v3_143047_C_A	3.07678

c. Primeros 20 SNPs de mayor peso para el PC3

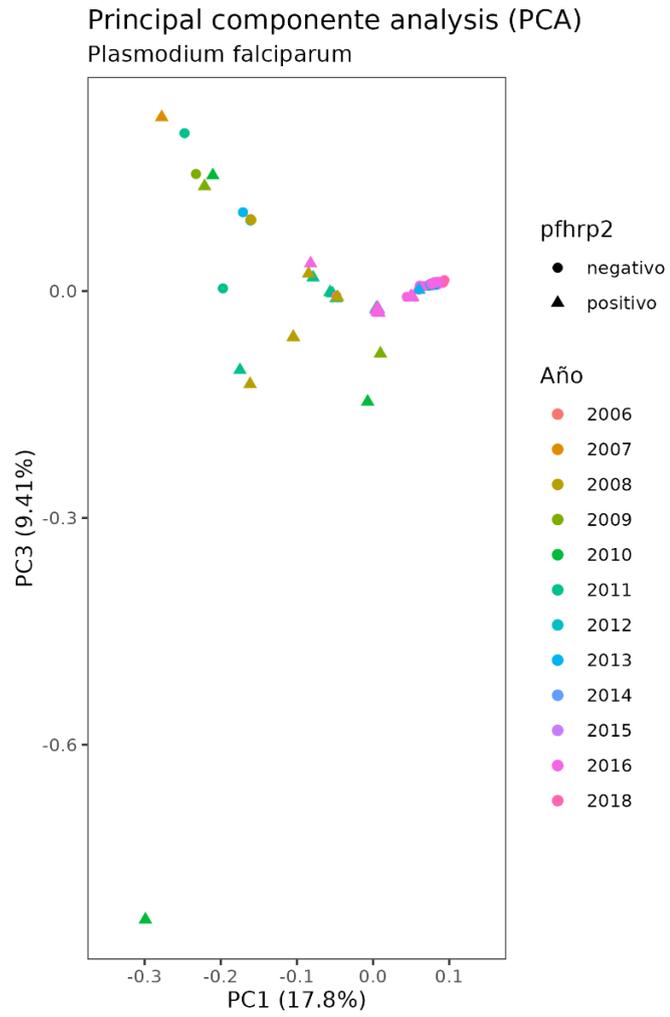
Cromosoma	SNP	PC3
Pf3D7_14_v3	Pf3D7_14_v3_480428_T_C	2.51647
Pf3D7_12_v3	Pf3D7_12_v3_1758089_G_T	2.4803
Pf3D7_14_v3	Pf3D7_14_v3_2261274_A_G	2.47896
Pf3D7_07_v3	Pf3D7_07_v3_1095414_G_C	2.47603
Pf3D7_13_v3	Pf3D7_13_v3_1155190_A_G	2.47514
Pf3D7_10_v3	Pf3D7_10_v3_544490_A_T	2.46173
Pf3D7_09_v3	Pf3D7_09_v3_453627_G_A	2.45265
Pf3D7_14_v3	Pf3D7_14_v3_2643430_G_C	2.45235
Pf3D7_13_v3	Pf3D7_13_v3_2074256_A_G	2.44675
Pf3D7_13_v3	Pf3D7_13_v3_2102174_T_A	2.4382
Pf3D7_13_v3	Pf3D7_13_v3_208878_A_C	2.43647
Pf3D7_08_v3	Pf3D7_08_v3_373052_A_C	2.4363
Pf3D7_09_v3	Pf3D7_09_v3_127688_A_T	2.4363
Pf3D7_08_v3	Pf3D7_08_v3_526021_A_G	2.42789
Pf3D7_09_v3	Pf3D7_09_v3_311366_G_T	2.41617
Pf3D7_07_v3	Pf3D7_07_v3_180921_A_C	2.40829
Pf3D7_07_v3	Pf3D7_07_v3_365329_A_G	2.40829
Pf3D7_08_v3	Pf3D7_08_v3_505950_A_G	2.39489
Pf3D7_14_v3	Pf3D7_14_v3_2967750_C_T	2.39474
Pf3D7_07_v3	Pf3D7_07_v3_1044052_T_C	2.39251

ANEXO 9. Análisis de componentes principales. Se identifican de acuerdo al año de colecta y a la presencia o ausencia del gen *pfhrp2*.

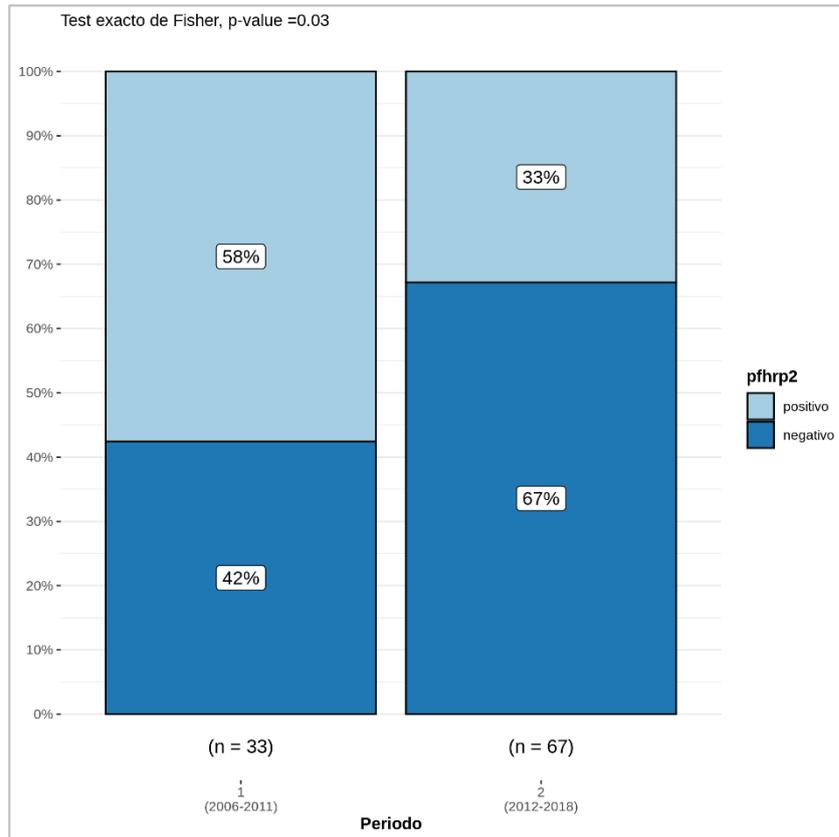
a. PC1 vs PC2



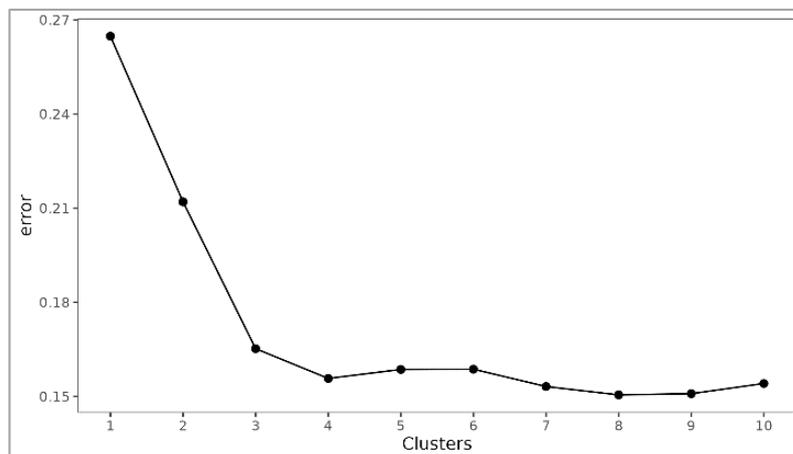
b. PC1 vs PC3



ANEXO 10. Comparación de la proporción de la delección de los genes *pfhrp2* en muestras de los Periodo 1 y Periodo 2

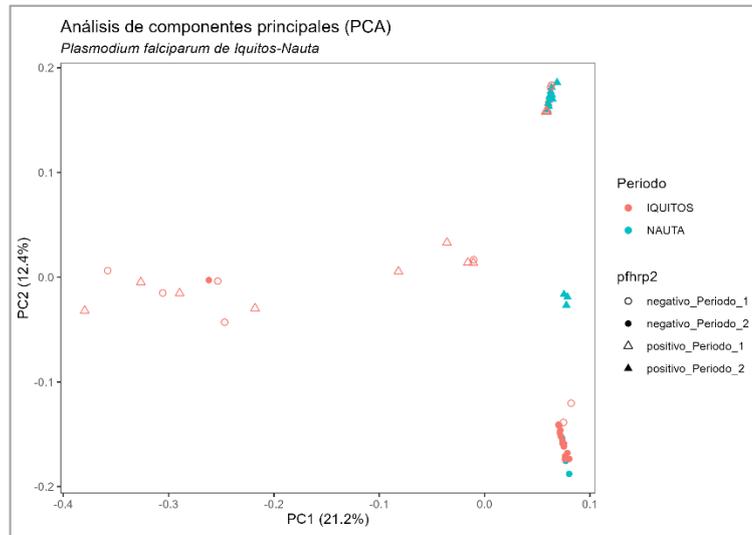


ANEXO 11. Error *cross-validation*, resultado del análisis con el programa Admixture.



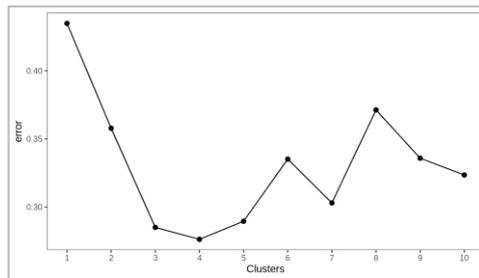
ANEXO 12. Análisis de estructuración poblacional para muestras de Iquitos y Nauta

a. PCA

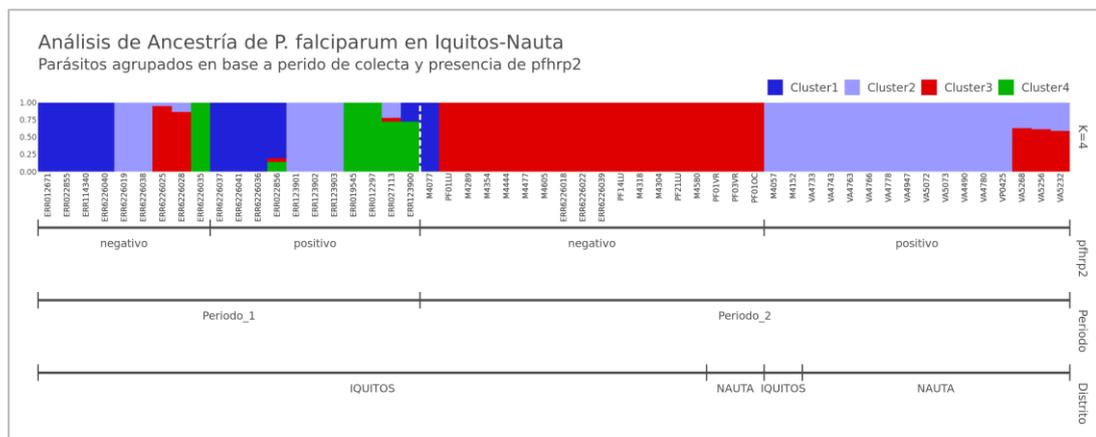


PCA de las muestras de Iquitos y Nauta. Se indica también la presencia de las delecciones de los genes *pfhrp2* y el periodo de colecta.

b. Análisis de ancestría



1. Error *cross-validation*, resultado del análisis con el programa Admixture.



2. Gráfico de ancestría. Se observan los clusters presentes entre Iquitos y Nauta.

ANEXO 13. Loci candidatos a estar seleccionados por *balancing selection*, a partir del análisis D de Tajima para los parásitos *pfhrp2/3* negativos del Periodo 2

Crom.	TajD	Inicio	Fin	N SNPs	Gen	
4	2.93	156360	156859	38	PF3D7_0402500	rifin pseudogene
6	2.61	723467	723966	29	PF3D7_0617400	erythrocyte membrane protein 1, PfEMP1
7	2.87	563131	563630	22		
7	3.73	563631	564130	10	PF3D7_0712500	RIF - rifin, pseudogene
7	3.17	567131	567630	10	PF3D7_0712600	VAR - erythrocyte membrane protein 1, PfEMP1
10	4.06	68325	68824	11	PF3D7_1001100	ACBP1 - Acyl-CoA binding protein, isoform 1, ACBP1
11	2.71	592872	593371	10	PF3D7_1115700	FP2A - cysteine proteinase falcipain 2a
12	2.86	767291	767790	10	PF3D7_1219300	VAR - erythrocyte membrane protein 1, PfEMP1
12	2.64	1711791	1712290	8	PF3D7_1240400	VAR - erythrocyte membrane protein 1, PfEMP1

ANEXO 14. Loci candidatos a estar seleccionados por *balancing selection*, a partir del análisis D de Tajima para los parásitos *pfhrp2/3* positivos del Periodo 2

Crom.	TajD	Inicio	Fin	N SNPs	Gen	
4	2.45	1133360	1133859	15	PF3D7_0425050	<i>Plasmodium</i> exported protein, unknown function, pseudogene
6	2.08	723467	723966	29	PF3D7_0617400	VAR - erythrocyte membrane protein 1, PfEMP1
7	2.89	525631	526130	8	PF3D7_0711900	RIF - rifin, pseudogene
10	3.59	68325	68824	11	PF3D7_1001100	acyl-CoA binding protein, isoform 1, ACBP1
12	2.23	407791	408290	12	PF3D7_1208800	zinc finger protein, putative
12	2.68	767291	767790	10	PF3D7_1219300	VAR - erythrocyte membrane protein 1, PfEMP1
12	2.22	1742791	1743290	21	PF3D7_1240900	VAR - erythrocyte membrane protein 1, PfEMP1
14	2.16	620977	621476	11	PF3D7_1415400	Btz domain-containing protein, putative

ANEXO 15. Resultados significativos del análisis iHS para los parásitos *pfhrp2/3* negativos del Periodo 2

SNP	Crom.	Posición	iHS	LogP-value	Gen	
4_354004_C_T	4	354004	5.09	6.45	PF3D7_0406900	conserved <i>Plasmodium</i> protein, unknown function
4_354006_C_T	4	354006	5.09	6.45		
4_354219_C_T	4	354219	4.97	6.18		
4_354231_C_T	4	354231	5.40	7.19		
4_536732_T_C	4	536732	5.31	6.97	PF3D7_0412000	LITAF-like zinc finger protein, putative
4_603734_G_A	4	603734	4.46	5.09	PF3D7_0413200	rifin
4_603738_C_G	4	603738	4.46	5.09		
5_602909_T_A	5	602909	4.50	5.16	PF3D7_0514300	aspartate--tRNA ligase, putative
5_798716_C_T	5	798716	4.38	4.92	PF3D7_0519400	40S ribosomal protein S24
5_798717_C_T	5	798717	4.38	4.92		
7_527827_G_C	7	527827	4.20	4.57	PF3D7_0712000	erythrocyte membrane protein 1, PfEMP1
11_1331667_C_A	11	1331667	4.74787302	5.69	PF3D7_1134200	conserved <i>Plasmodium</i> protein, unknown function
12_767824_G_A	12	767824	6.53662895	10.20	PF3D7_1219300	erythrocyte membrane protein 1, PfEMP1
12_909914_C_T	12	909914	4.14928018	4.48	PF3D7_1222600	AP2 domain transcription factor AP2-G
12_1176505_G_T	12	1176505	7.86868169	14.44	PF3D7_1228800	WD repeat-containing protein, putative
13_1120286_G_A	13	1120286	7.44168931	13.00	PF3D7_1326600	conserved <i>Plasmodium</i> protein, unknown function
14_1185608_G_A	14	1185608	5.25101496	6.82	PF3D7_1430100	serine/threonine protein phosphatase 2A activator
14_2018850_G_A	14	2018850	7.24677539	12.37	PF3D7_1449300	transcription factor IIIb subunit, putative
14_2303623_C_T	14	2303623	4.39431343	4.95	PF3D7_1456200	conserved <i>Plasmodium</i> protein, unknown function

ANEXO 16. Resultados significativos del análisis iHS para los parásitos *pfhrp2/3* positivos del Periodo 2

SNP	Crom.	Posición	iHS	LogP-value	Gen	
4_617475_A_T	4	617475	6.41	9.84	PF3D7_0413600	26S protease regulatory subunit 6B, putative
4_970199_C_G	4	970199	7.33	12.64	PF3D7_0421300	erythrocyte membrane protein 1, PfEMP1
5_751695_A_C	5	751695	-5.48	7.37	PF3D7_0518000	NOSIP domain-containing protein, putative
11_618602_A_T	11	618602	7.47	13.08	PF3D7_1116300	peptidyl-prolyl cis-trans isomerase
11_636826_A_G	11	636826	6.32	9.58	PF3D7_1116800	heat shock protein 101
12_767824_G_A	12	767824	6.35	9.66	PF3D7_1219300	erythrocyte membrane protein 1, PfEMP1
13_1041516_T_A	13	1041516	4.35	4.87	PF3D7_1324900	L-lactate dehydrogenase
13_1184534_A_G	13	1184534	4.13	4.45	PF3D7_1328100	proteasome subunit beta type-7, putative
14_1093398_T_G	14	1093398	7.42	12.92	PF3D7_1427900	leucine-rich repeat protein

ANEXO 17. Resultados significativos del análisis XP-EHH para los parásitos *pfhrp2/3* negativos vs positivos del Periodo 2

SNP	C	Posición	XPEHH_hrp2_positivo_ hrp2_negativo	LogP-value	Pop_sel	Gen		
4_603706_G_T	4	603706	4.66	5.51	hrp2_positive	PF3D7_0413200	RIF - rifin	
4_603708_C_T	4	603708	4.92	6.06	hrp2_positive			
4_603716_C_G	4	603716	5.12	6.52	hrp2_positive			
4_603749_T_A	4	603749	4.95	6.14	hrp2_positive			
4_603763_T_C	4	603763	5.05	6.36	hrp2_positive			
4_603764_T_G	4	603764	5.04	6.33	hrp2_positive			
4_603820_C_T	4	603820	4.95	6.13	hrp2_positive			
4_603827_A_C	4	603827	5.02	6.30	hrp2_positive			
11_593409_G_A	11	593409	-7.08	11.86	hrp2_negative			PF3D7_1115700
11_1054078_C_T	11	1054078	7.69	13.85	hrp2_positive	PF3D7_1126900	small nuclear ribonucleoprotein F, putative	
12_1712088_A_T	12	1712088	4.57	5.31	hrp2_positive	PF3D7_1240400	VAR - erythrocyte membrane protein 1, PfEMP1	
12_1712095_C_G	12	1712095	4.62	5.41	hrp2_positive			
12_1712113_A_C	12	1712113	10.46	24.91	hrp2_positive			
12_1712115_A_T	12	1712115	9.96	22.65	hrp2_positive			
12_1712118_G_A	2	1712118	9.96	22.65	hrp2_positive			
12_1712125_T_A	12	1712125	10.93	27.09	hrp2_positive			
12_1712129_A_C	12	1712129	11.06	27.73	hrp2_positive			
12_1715843_A_T	12	1715843	9.64	21.26	hrp2_positive			intergenic region
12_1715884_T_C	12	1715884	9.48	20.59	hrp2_positive			PF3D7_1240500

ANEXO 18. Resultados significativos del análisis rsb para los parásitos *pfhrp2/3* negativos vs positivos del Periodo 2

SNP	Crom.	Posición	RSB_hrp2_positive_hrp2_negative	LogP-value	Pop_sel	Gen	
4_491504_A_T	4	491504	-5.60	7.68	hrp2_negative	PF3D7_0410900	conserved <i>Plasmodium</i> protein, unknown function
4_491520_T_A	4	491520	-4.72	5.634	hrp2_negative		
4_603734_G_A	4	603734	4.43	5.02	hrp2_positive	PF3D7_0413200	RIF - rifin
4_603738_C_G	4	603738	4.43	5.02	hrp2_positive		
4_603744_C_T	4	603744	4.71	5.60	hrp2_positive		
4_603763_T_C	4	603763	5.11	6.49	hrp2_positive		
4_603794_T_C	4	603794	5.65	7.80	hrp2_positive		
4_603820_C_T	4	603820	5.33	7.02	hrp2_positive		
11_1191860_C_G	11	1191860	5.04	6.34	hrp2_positive	PF3D7_1130900	conserved <i>Plasmodium</i> protein, unknown function
11_1199783_T_A	11	1199783	5.84	8.30	hrp2_positive	PF3D7_1131200	conserved <i>Plasmodium</i> protein, unknown function
11_1346267_T_C	11	1346267	5.20	6.70	hrp2_positive	PF3D7_1134500	alpha/beta hydrolase, putative
11_1348187_C_T	11	1348187	5.06	6.38	hrp2_positive		
11_1352911_C_T	11	1352911	5.05	6.36	hrp2_positive	PF3D7_1134600	zinc finger protein, putative
11_1414225_A_T	11	1414225	5.64	7.76	hrp2_positive	PF3D7_1136000	conserved <i>Plasmodium</i> protein, unknown function
11_1415476_C_A	11	1415476	5.60	7.67	hrp2_positive	PF3D7_1136100	conserved <i>Plasmodium</i> protein, unknown function
11_1417280_T_C	11	1417280	5.60	7.66	hrp2_positive	PF3D7_1136200	conserved <i>Plasmodium</i> protein, unknown function
12_1712113_A_C	12	1712113	5.71	7.94	hrp2_positive	PF3D7_1240400	erythrocyte membrane protein 1, PfEMP1
12_1712115_A_T	12	1712115	11.16	28.18	hrp2_positive		
12_1712118_G_A	12	1712118	11.16	28.18	hrp2_positive		
12_1712125_T_A	12	1712125	6.98	11.53	hrp2_positive		
12_1712129_A_C	12	1712129	6.73	10.76	hrp2_positive		
12_1715843_A_T	12	1715843	10.80	26.43	hrp2_positive	intergenic	

ANEXO 19. Resultados de ontología genética para grupo *pfhrp2/3* negativos del Periodo 2

ID	Nombre	Conteo de Background	Conteo de Resultado	Lista de genes resultante	% de Background	Enriquecimiento	Odds ratio	P-value	Benjamini	Bonferroni
GO:0044257	cellular protein catabolic process	138	3	PF3D7_0413600,PF3D7_1115700,PF3D7_1328100,	2.2	9.62	13.12	0.003124	0.06584813	0.29567801
GO:0051603	proteolysis involved in cellular protein catabolic process	138	3	PF3D7_0413600,PF3D7_1115700,PF3D7_1328100,	2.2	9.62	13.12	0.003124	0.06584813	0.29567801
GO:0030163	protein catabolic process	146	3	PF3D7_0413600,PF3D7_1115700,PF3D7_1328100,	2.1	9.09	12.36	0.00365347	0.06584813	0.34707951
GO:0043335	protein unfolding	2	1	PF3D7_1116800,	50	221.27	485.6	0.00451467	0.06584813	0.42889336
GO:1901565	organonitrogen compound catabolic process	174	3	PF3D7_0413600,PF3D7_1115700,PF3D7_1328100,	1.7	7.63	10.28	0.00599174	0.06584813	0.56921568
GO:0098609	cell-cell adhesion	62	2	PF3D7_0421300,PF3D7_1219300,	3.2	14.28	17.77	0.00815313	0.06584813	0.77454724
GO:2000144	positive regulation of DNA-templated transcription, initiation	4	1	PF3D7_0413600,	25	110.64	161.8	0.0090108	0.06584813	0.85602572
GO:0045898	regulation of RNA polymerase II transcription preinitiation complex assembly	4	1	PF3D7_0413600,	25	110.64	161.8	0.0090108	0.06584813	0.85602572
GO:0044053	translocation of peptides or proteins into host cell cytoplasm	4	1	PF3D7_1116800,	25	110.64	161.8	0.0090108	0.06584813	0.85602572
GO:0060260	regulation of transcription initiation from RNA polymerase II promoter	4	1	PF3D7_0413600,	25	110.64	161.8	0.0090108	0.06584813	0.85602572
GO:0045899	positive regulation of RNA polymerase II transcription preinitiation complex assembly	4	1	PF3D7_0413600,	25	110.64	161.8	0.0090108	0.06584813	0.85602572
GO:2000142	regulation of DNA-templated transcription, initiation	4	1	PF3D7_0413600,	25	110.64	161.8	0.0090108	0.06584813	0.85602572
GO:0060261	positive regulation of transcription initiation from RNA polymerase II promoter	4	1	PF3D7_0413600,	25	110.64	161.8	0.0090108	0.06584813	0.85602572
GO:0007155	cell adhesion	70	2	PF3D7_0421300,PF3D7_1219300,	2.9	12.64	15.65	0.01031013	0.06985174	0.97946244
GO:0019538	protein metabolic process	5	1	PF3D7_1116800,	20	88.51	121.33	0.01125193	0.06985174	1
GO:0009405	obsolete pathogenesis	75	2	PF3D7_0421300,PF3D7_1219300,	2.7	11.8	14.56	0.0117742	0.06985174	1
GO:0006508	proteolysis	227	3	PF3D7_0413600,PF3D7_1115700,PF3D7_1328100,	1.3	5.85	7.76	0.01249979	0.06985174	1
GO:0034605	cellular response to heat	6	1	PF3D7_1116800,	16.7	73.76	97.04	0.01348846	0.07118909	1
GO:0070682	proteasome regulatory particle assembly	7	1	PF3D7_0413600,	14.3	63.22	80.85	0.01572039	0.07467183	1

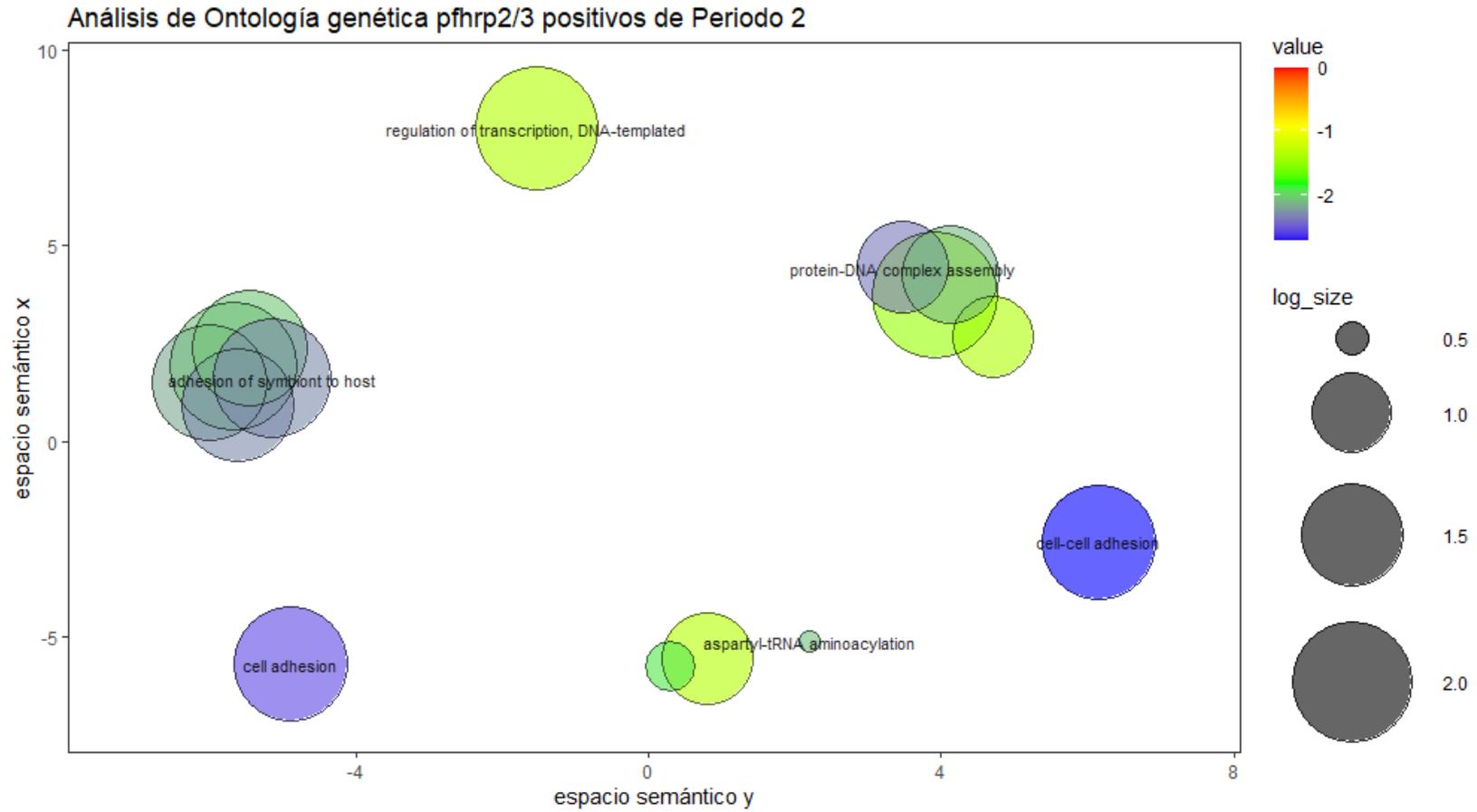
GO:0031334	positive regulation of protein-containing complex assembly	7	1	PF3D7_0413600,	14.3	63.22	80.85	0.01572039	0.07467183	1
GO:0006511	ubiquitin-dependent protein catabolic process	113	2	PF3D7_0413600,PF3D7_1328100,	1.8	7.83	9.5	0.02562021	0.09884998	1
GO:0000413	protein peptidyl-prolyl isomerization	12	1	PF3D7_1116300,	8.3	36.88	44.05	0.02681132	0.09884998	1
GO:0019941	modification-dependent protein catabolic process	116	2	PF3D7_0413600,PF3D7_1328100,	1.7	7.63	9.25	0.02690516	0.09884998	1
GO:0043632	modification-dependent macromolecule catabolic process	120	2	PF3D7_0413600,PF3D7_1328100,	1.7	7.38	8.92	0.02865934	0.09884998	1
GO:0043248	proteasome assembly	13	1	PF3D7_0413600,	7.7	34.04	40.38	0.02901582	0.09884998	1
GO:0018208	peptidyl-proline modification	13	1	PF3D7_1116300,	7.7	34.04	40.38	0.02901582	0.09884998	1
GO:0010499	proteasomal ubiquitin-independent protein catabolic process	14	1	PF3D7_1328100,	7.1	31.61	37.26	0.03121578	0.09884998	1
GO:0045944	positive regulation of transcription by RNA polymerase II	14	1	PF3D7_0413600,	7.1	31.61	37.26	0.03121578	0.09884998	1
GO:0042000	translocation of peptides or proteins into host	14	1	PF3D7_1116800,	7.1	31.61	37.26	0.03121578	0.09884998	1
GO:0044417	translocation of molecules into host	14	1	PF3D7_1116800,	7.1	31.61	37.26	0.03121578	0.09884998	1
GO:0043254	regulation of protein-containing complex assembly	16	1	PF3D7_0413600,	6.3	27.66	32.28	0.03560212	0.10256883	1
GO:0006096	glycolytic process	17	1	PF3D7_1324900,	5.9	26.03	30.26	0.03778852	0.10256883	1
GO:0006757	ATP generation from ADP	17	1	PF3D7_1324900,	5.9	26.03	30.26	0.03778852	0.10256883	1
GO:0009179	purine ribonucleoside diphosphate metabolic process	17	1	PF3D7_1324900,	5.9	26.03	30.26	0.03778852	0.10256883	1
GO:0046031	ADP metabolic process	17	1	PF3D7_1324900,	5.9	26.03	30.26	0.03778852	0.10256883	1
GO:0009185	ribonucleoside diphosphate metabolic process	18	1	PF3D7_1324900,	5.6	24.59	28.47	0.0399704	0.10262671	1
GO:0009135	purine nucleoside diphosphate metabolic process	18	1	PF3D7_1324900,	5.6	24.59	28.47	0.0399704	0.10262671	1
GO:0016052	carbohydrate catabolic process	21	1	PF3D7_1324900,	4.8	21.07	24.18	0.0464891	0.11324268	1
GO:0006165	nucleoside diphosphate phosphorylation	21	1	PF3D7_1324900,	4.8	21.07	24.18	0.0464891	0.11324268	1
GO:0006986	response to unfolded protein	22	1	PF3D7_1116800,	4.5	20.12	23.03	0.04865304	0.11411106	1
GO:0051701	biological process involved in interaction with host	382	3	PF3D7_0421300,PF3D7_1116800,PF3D7_1219300,	0.8	3.48	4.43	0.04924793	0.11411106	1

ANEXO 20. Resultados de ontología genética para grupo *pfhrp2/3* positivos del Periodo 2

ID	Nombre	Conteo de Background	Conteo de Resultado	Lista de genes resultante	% de Background	Enriquecimiento	Odds ratio	P-value	Benjamini	Bonferroni
GO:0098609	cell-cell adhesion	62	3	PF3D7_0712000,PF3D7_1219300,PF3D7_1240400,	4.8	11.78	14.32	0.00192212	0.03310842	0.14608132
GO:0007155	cell adhesion	70	3	PF3D7_0712000,PF3D7_1219300,PF3D7_1240400,	4.3	10.43	12.59	0.00272431	0.03310842	0.20704757
GO:0009405	obsolete pathogenesis	75	3	PF3D7_0712000,PF3D7_1219300,PF3D7_1240400,	4	9.74	11.71	0.00331684	0.03310842	0.25208004
GO:0070898	RNA polymerase III preinitiation complex assembly	1	1	PF3D7_1449300,	100	243.4	inf	0.00410846	0.03310842	0.31224322
GO:0065004	protein-DNA complex assembly	24	2	PF3D7_1228800,PF3D7_1449300,	8.3	20.28	24.37	0.00419303	0.03310842	0.31867063
GO:0044406	adhesion of symbiont to host	176	4	PF3D7_0413200,PF3D7_0712000,PF3D7_1219300,PF3D7_1240400,	2.3	5.53	6.8	0.00507933	0.03310842	0.38602934
GO:0020035	cytoadherence to microvasculature, mediated by symbiont protein	176	4	PF3D7_0413200,PF3D7_0712000,PF3D7_1219300,PF3D7_1240400,	2.3	5.53	6.8	0.00507933	0.03310842	0.38602934
GO:0020013	modulation by symbiont of host erythrocyte aggregation	190	4	PF3D7_0413200,PF3D7_0712000,PF3D7_1219300,PF3D7_1240400,	2.1	5.12	6.27	0.00666356	0.03310842	0.50643056
GO:0044068	modulation by symbiont of host cellular process	192	4	PF3D7_0413200,PF3D7_0712000,PF3D7_1219300,PF3D7_1240400,	2.1	5.07	6.2	0.00691402	0.03310842	0.52546575
GO:0022610	biological adhesion	195	4	PF3D7_0413200,PF3D7_0712000,PF3D7_1219300,PF3D7_1240400,	2.1	4.99	6.1	0.00730145	0.03310842	0.55491058
GO:0071824	protein-DNA complex subunit organization	33	2	PF3D7_1228800,PF3D7_1449300,	6.1	14.75	17.27	0.00784589	0.03310842	0.59628776
GO:0006422	aspartyl-tRNA aminoacylation	2	1	PF3D7_0514300,	50	121.7	255.11	0.00820089	0.03310842	0.62326749
GO:0020033	antigenic variation	202	4	PF3D7_0413200,PF3D7_0712000,PF3D7_1219300,PF3D7_1240400,	2	4.82	5.87	0.00826143	0.03310842	0.62786902
GO:0042783	evasion of host immune response	205	4	PF3D7_0413200,PF3D7_0712000,PF3D7_1219300,PF3D7_1240400,	2	4.75	5.78	0.00869739	0.03310842	0.66100139
GO:0044003	modulation by symbiont of host process	205	4	PF3D7_0413200,PF3D7_0712000,PF3D7_1219300,PF3D7_1240400,	2	4.75	5.78	0.00869739	0.03310842	0.66100139
GO:0052572	response to host immune response	208	4	PF3D7_0413200,PF3D7_0712000,PF3D7_1219300,PF3D7_1240400,	1.9	4.68	5.69	0.00914838	0.03310842	0.69527672
GO:0075136	response to host	208	4	PF3D7_0413200,PF3D7_0712000,PF3D7_1219300,PF3D7_1240400,	1.9	4.68	5.69	0.00914838	0.03310842	0.69527672
GO:0052200	response to host defenses	208	4	PF3D7_0413200,PF3D7_0712000,PF3D7_1219300,PF3D7_1240400,	1.9	4.68	5.69	0.00914838	0.03310842	0.69527672
GO:0051707	response to other organism	208	4	PF3D7_0413200,PF3D7_0712000,PF3D7_1219300,PF3D7_1240400,	1.9	4.68	5.69	0.00914838	0.03310842	0.69527672

GO:0052173	response to defenses of other organism	208	4	PF3D7_0413200,PF3D7_0712000,PF3D7_1219300,PF3D7_1240400,	1.9	4.68	5.69	0.00914838	0.03310842	0.69527672
GO:0043207	response to external biotic stimulus	208	4	PF3D7_0413200,PF3D7_0712000,PF3D7_1219300,PF3D7_1240400,	1.9	4.68	5.69	0.00914838	0.03310842	0.69527672
GO:0006384	transcription initiation from RNA polymerase III promoter	3	1	PF3D7_1449300,	33.3	81.13	127.53	0.01227733	0.04056858	0.93307733
GO:0070897	transcription preinitiation complex assembly	3	1	PF3D7_1449300,	33.3	81.13	127.53	0.01227733	0.04056858	0.93307733
GO:0034622	cellular protein-containing complex assembly	177	3	PF3D7_1126900,PF3D7_1228800,PF3D7_1449300,	1.7	4.13	4.74	0.03417626	0.10822483	1
GO:0006334	nucleosome assembly	10	1	PF3D7_1228800,	10	24.34	28.3	0.04036996	0.11983066	1
GO:0006503	protein-containing complex assembly	193	3	PF3D7_1126900,PF3D7_1228800,PF3D7_1449300,	1.6	3.78	4.33	0.042554	0.11983066	1
GO:1902850	microtubule cytoskeleton organization involved in mitosis	11	1	PF3D7_1430100,	9.1	22.13	25.46	0.04432069	0.11983066	1
GO:0007052	mitotic spindle organization	11	1	PF3D7_1430100,	9.1	22.13	25.46	0.04432069	0.11983066	1
GO:0006355	regulation of transcription, DNA-templated	84	2	PF3D7_1222600,PF3D7_1449300,	2.4	5.8	6.46	0.04572486	0.11983066	1
GO:0006383	transcription by RNA polymerase III	12	1	PF3D7_1449300,	8.3	20.28	23.14	0.04825595	0.12224841	1

ANEXO 21. Gráfico de resultados de análisis de ontología genética para los parásitos *pfhrp2/3* positivos.



ANEXO 22. SNPs identificados en el gen *FP2A*, en los 100 genomas analizados.

SNP	Alelo	Periodo_1_positivo (19)	Periodo_1_negativo (14)	Periodo_2_positivo (22)	Periodo_2_negativo (45)	ANN[*].HGVS_P	Región en proteína
11_592551_T_C	C	1	1	0.977	0.977	p.Asp345Gly	Peptidase_C1 Peptidase C1A, papaína C-terminal
	T	0	0	0.023	0.023		
11_592558_T_G	G	1	1	0.977	0.977	p.Thr343Pro	Peptidase_C1 Peptidase C1A, papaína C-terminal
	T	0	0	0.023	0.023		
11_592814_A_T	A	0.5	0.357	0.955	0.977	p.Asn257Lys	Región de baja complejidad
	T	0.5	0.643	0.045	0.023		
11_592816_T_C	T	0.5	0.357	0.955	0.977	p.Asn257Asp	Región de baja complejidad
	C	0.5	0.643	0.045	0.023		
11_592821_T_C	C	1	1	1	0.978	p.Lys255Arg	Región de baja complejidad
	T	0	0	0	0.022		
11_592902_C_G	C	0.688	0.857	1	1	p.Ser228Thr	Región de baja complejidad
	G	0.312	0.143	0	0		
11_592994_A_T	A	0.938	0.964	0.909	0.956	p.Asn197Lys	Inhibitor_I29 Dominio inhibidor del propéptido de catepsina (I29)
	T	0.062	0.036	0.091	0.044		
11_593085_A_G	A	0.733	0.821	0.75	0.611	p.Met167Thr	Inhibitor_I29 Dominio inhibidor del propéptido de catepsina (I29)
	G	0.267	0.179	0.25	0.389		
11_593261_A_T	A	0.794	0.821	0.705	0.544	p.Asn108Lys	Región de baja complejidad
	T	0.206	0.179	0.295	0.456		
11_593265_T_A	T	0.794	0.821	0.705	0.544	p.Lys107Met	Región de baja complejidad
	A	0.206	0.179	0.295	0.456		
11_593271_C_T	C	0.794	0.821	0.705	0.544	p.Ser105Asn	Región de baja complejidad
	T	0.206	0.179	0.295	0.456		
11_593409_G_A	A	0.406	0.286	0.864	0.978	p.Ser59Phe	Región de baja complejidad

	G	0.594	0.714	0.136	0.022		
11_593540_T_A	T	0.812	0.5	0.95	0.977	p.Gln15His	Región de baja complejidad
	A	0.188	0.5	0.05	0.023		

Se indican las frecuencias respectivas de cada SNP y la región de la proteína donde se localizaría. En **negrita** se resalta el SNPs que forma el barrido selectivo identificado.

ANEXO 23. SNPs identificados en el gen *AP2-G*, en los 100 genomas analizados.

SNP	Alelo	Periodo_1_positivo (19)	Periodo_1_negativo (14)	Periodo_2_positivo (22)	Periodo_2_negativo (45)	ANN[*].HGVS_P	Región en proteína
12_907264_A_G	A	0.737	0.786	1	1	p.Lys21Arg	Región de baja complejidad
	G	0.263	0.214	0	0		
12_907403_G_A	G	1	1	1	1	p.Met67Ile	Región de baja complejidad
12_907407_G_T	G	1	1	1	1	p.Glu69*	Región de baja complejidad
12_907426_G_A	G	1	1	1	0.989	p.Cys75Tyr	Región de baja complejidad
	A	0	0	0	0.011		
12_907556_C_T	C	0.974	1	1	1	p.Asn118Asn	Región de baja complejidad
	T	0.026	0	0	0		
12_908019_T_C	C	0.947	0.571	1	0.978	p.Tyr273His	Región de baja complejidad
	T	0.053	0.429	0	0.022		
12_908962_T_C	T	0.947	1	1	1	p.Ile587Thr	Región de baja complejidad
	C	0.053	0	0	0		
12_909162_G_A	G	0.947	0.571	1	0.978	p.Asp654Asn	Región de baja complejidad
	A	0.053	0.429	0	0.022		
12_909914_C_T	C	0.711	0.462	1	0.978	p.Asn904Asn	Región de baja complejidad
	T	0.289	0.538	0	0.022		
12_910153_A_C	C	0.895	0.571	1	0.978	p.Lys984Thr	Región de baja complejidad
	A	0.105	0.429	0	0.022		
12_910163_C_T	C	1	1	1	1	p.Asn987Asn	Región de baja complejidad
12_910247_T_A	T	1	1	1	1	p.Ile1015Ile	Región de baja complejidad
12_910248_G_T	G	1	1	1	1	p.Val1016Phe	Región de baja complejidad
12_910257_A_T	A	1	1	1	1	p.Ile1019Leu	Región de baja complejidad
12_910324_T_A	A	1	1	1	1	p.Ile1041Lys	Región de baja complejidad

12_910732_C_T	C	1	1	1	1	p.Thr1177Ile	Región de baja complejidad
12_910741_G_A	G	0.684	0.357	1	0.978	p.Gly1180Glu	Región de baja complejidad
	A	0.316	0.643	0	0.022		
12_911273_T_C	T	0.944	1	1	1	p.Asn1357Asn	Región de baja complejidad
	C	0.056	0	0	0		
12_911300_C_T	C	1	1	1	1	p.Asn1366Asn	Región de baja complejidad
12_911306_T_C	T	0.722	0.385	1	0.976	p.Asn1368Asn	Región de baja complejidad
	C	0.278	0.615	0	0.024		
12_911532_A_G	G	1	1	1	1	p.Ser1444Gly	Región de baja complejidad
12_912091_G_A	G	1	1	1	1	p.Cys1630Tyr	Región de baja complejidad
12_912233_G_A	G	1	1	1	1	p.Gly1677Gly	Región de baja complejidad
12_912840_G_A	A	0.684	0.357	1	0.978	p.Asp1880Asn	Región de baja complejidad
	G	0.316	0.643	0	0.022		
12_913124_A_T	T	1	1	1	1	p.Lys1974Asn	Región de baja complejidad
12_913144_A_G	A	1	1	1	1	p.Asn1981Ser	Región de baja complejidad
12_913148_G_T	G	1	1	1	1	p.Lys1982Asn	Región de baja complejidad
12_913157_C_A	A	1	1	1	1	p.Asn1985Lys	Región de baja complejidad
12_913895_C_T	C	1	1	1	0.989	p.Tyr2231Tyr	Región de baja complejidad
	T	0	0	0	0.011		
12_914436_T_A	T	1	1	1	0.989	p.Tyr2412Asn	Región de baja complejidad
	A	0	0	0	0.011		

Se indican las frecuencias respectivas de cada SNP y la región de la proteína donde se localizaría. En **negrita** se resalta el SNPs que forma el barrido selectivo identificado