



UNIVERSIDAD PERUANA
CAYETANO HEREDIA
ESCUELA DE POSGRADO

EPIDEMIOLOGÍA MOLECULAR DE
PLASMODIUM VIVAX EN CUATRO
COMUNIDADES RIBEREÑAS DEL
DISTRITO DE MAZÁN EN LA
AMAZONÍA PERUANA

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

PAULO CESAR MANRIQUE VALVERDE

LIMA - PERÚ

2018

ASESOR DE TESIS:

Dionicia Gamboa, Msc, PhD.

Departamento de Ciencias Celulares y Moleculares,

Faculta de Ciencias y Filosofía

Instituto de Medicina Tropical Alexander von Humboldt

Universidad Peruana Cayetano Heredia

DEDICATORIA

A mi familia, quienes de manera incondicional me han acompañado en las distintas aventuras que he emprendido, entre ellas mi maestría. A mi madre María quien me enseñó a ser perseverante, paciente y meticulado para cumplir mis metas. Mi padre Miguel de quien adquirí el carácter crítico que me caracteriza y que me ha sido muy útil en ciencia. A mi hermano Edgar, quien ha sido un gran compañero de debate, juegos y otras cosas, y por quien he tratado de ser un modelo a seguir. A mi fiel amigo Deimos que, aunque ya no está presente, estuvo siempre acompañándome en las largas jornadas de estudio nocturno, me despedía con su clásico ladrido de foca cada vez que yo partía al trabajo o universidad, y que siempre me pedía que jugara con él y lo alimentara a mi regreso a casa. Y a mi amor Yesenia, cuya dulce sonrisa, carisma y calidez siempre me levanto el ánimo y me motivo a continuar durante esta travesía.

AGRADECIMIENTOS

Deseo agradecer a mis tutores Oscar Nolasco, Ananias Escalante y de manera especial a Dionicia Gamboa, quienes me brindaron su apoyo, dedicación, guía y paciencia para cumplir con éxitos la meta de obtener el grado de maestro en bioquímica y biología molecular. Ha sido un privilegio contar con su asesoramiento y ayuda a lo largo de este tiempo.

También quiero agradecer a todos mis compañeros del laboratorio de Malaria 308, 309 y 310 del LID, y al personal del laboratorio satélite de Iquitos, por su compañía, amabilidad y soporte durante los años que me desempeñé como alumno de maestría y que vengo trabajando junto a ellos dentro de este grupo de investigación.

Gracias también a todas las personas que accedieron a participar en los trabajos de investigación del Laboratorio de malaria; ya que con ello nos permiten no solo obtener un grado académico, si no también y más importante nos ayudan a combatir la malaria desde un enfoque científico, multidisciplinario y con una finalidad pragmática, el desarrollo de herramientas que puedan ser trasladables al sector público.

FUENTES DE FINANCIAMIENTO

Este trabajo de investigación estuvo financiado por el proyecto “UCSD - UPCH Global Infectious Diseases Training Programme: Endemic Infectious Diseases of the Peruvian Amazon” (D43TW007120), el cual financió el programa de maestría en bioquímica y biología molecular; el proyecto “Hacia la eliminación de la malaria en el Perú” (N° 008-2014-FONDECYT), círculo de investigación financiado por el Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica (CONCYTEC), que brindó el financiamiento de mi estipendio como tesista, y de los procesos de colección, diagnóstico y caracterización genética de las muestras provenientes del distrito de Mazán; el proyecto “AMAZONIA-International Center of Excellence for Malaria Research” (5U19AI089681, NIH-USA), el cual dio apoyo financiero en los procesos de colección, diagnóstico y caracterización genética de las muestras provenientes de las áreas de estudio de Cahuide, San José de Lupuna y Santa Emilia; y finalmente el proyecto “Joint project to strengthen research skills on molecular epidemiology and to uncover malaria transmission features relevant for its control in the Peruvian Amazon” (VLIR-UOS project ZEIN2016Z195) que brindó apoyo financiero en el proceso de caracterización genética de las muestras provenientes de Cahuide, San José de Lupuna y Santa Emilia.

TABLA DE CONTENIDOS

LISTADO DE TABLAS	IV
LISTADO DE FIGURAS	V
RESUMEN	VI
ABSTRACT	VII
I. INTRODUCCIÓN	1
II. PLANTEAMIENTO DEL PROBLEMA	5
III. MARCO TEORICO Y ANTECEDENTES	7
Malaria	7
Ciclo biológico de malaria y de <i>P. vivax</i>	8
Características de la transmisión de <i>P. vivax</i>	12
Vigilancia y control de <i>P. vivax</i>	19
Principios básicos de genética poblacional	20
Indicadores genético poblacionales en el estudio de malaria.....	25
IV. OBJETIVOS	33

Objetivo general	33
Objetivos específicos	33
V. JUSTIFICACIÓN	35
VI. MATERIALES Y MÉTODOS.....	36
Consentimiento ético.....	36
Área de estudio y población.....	36
Diseño de estudio	38
Procedimientos de laboratorio.....	45
Extracción de ADN y diagnóstico por PCR en tiempo real	45
Genotipificación por microsatélites	47
Análisis genético poblacional	49
Estructuración poblacional e identificación de infecciones importadas e introducidas	50
Análisis estadístico.....	54
VII. RESULTADOS.....	56

Complejidad de infección, diversidad genética y desequilibrio de ligamiento.....	57
Diferenciación genética, estructuración poblacional e identificación del origen de las infecciones	59
Factores demográficos y clínicos-epidemiológicos asociados a los haplogrupos genéticos.....	70
Asociación de haplogrupos genéticos, factores demográficos y clínicos-epidemiológicos con la presencia de síntomas.....	72
VIII. DISCUSIONES.....	77
IX. CONCLUSIONES.....	86
X. REFERENCIAS BIBLIOGRÁFICAS	88
XI. ANEXOS	96

LISTADO DE TABLAS

Tabla 1: Diversidad Haplotípica y Genética por comunidad y por cuenca en el distrito de Mazán.....	59
Tabla 2: Probabilidad de procedencia a una población	66
Tabla 3: Fuente u origen de las infecciones en el Distrito de Mazán	67
Tabla 4: Modelos bivariados de regresión logística para la presencia del Haplogrupo 2 en la cuenca Mazán	71
Tabla 5: Modelos bivariados y multivariado de regresión logística para la presencia de Sintomatología en el distrito de Mazán	73
Tabla 6: Modelos bivariados y multivariado de regresión logística para la presencia de Sintomatología en la cuenca Mazán	74

LISTADO DE FIGURAS

Figura 1: Distribución geográfica de las comunidades seleccionadas (Gamitanacocha, Libertad, Primero de Enero y Urco Miraño), y de las muestras genotipificadas.....	38
Figura 2: Diseño de estudio y colección de muestras.....	40
Figura 3: Criterios de inclusión y exclusión de muestras para el análisis genético poblacional.....	42
Figura 4: Flujograma de recolección, procesamiento y análisis de datos.....	44
Figura 5: Red mínima de expansión de haplotipos del total de infecciones genotipificadas en las 4 áreas de estudio (Mazán, Lupuna, Cahuide y Santa Emilia)	62
Figura 6: Estructuración poblacional de <i>P. vivax</i> en cuatro cuencas de la Amazonía Peruana.....	64
Figura 7: Distribución geográfica de los haplogrupos, casos autóctonos, e infecciones policlonales en las cuatro comunidades estudiadas	69
Figura 8: Red mínima de expansión de haplotipos del haplogrupo 6 presente en UM	76

RESUMEN

Plasmodium vivax es la principal causa de malaria en la Amazonía Peruana, y ante el incremento de la incidencia de casos, las herramientas genético poblacionales podrían ser componentes claves para el diseño y focalización de estrategias de control. Se genotipificaron 173 Infecciones causadas por *P. vivax* en 2 cuencas, Mazan (MZ) y Napo (NP). Se evaluó la diversidad genética, se identificó el número de grupos genéticos y las infecciones foráneas. MZ presentó 35.2% de infecciones policlonales, alta diversidad ($He = 0.71$), y menor desequilibrio de ligamiento ($I_A^S = 0.07$) con respecto a NP (18.8%, 0.39, 0.21). Se identificaron cinco grupos genéticos y 20 infecciones foráneas en MZ; en NP se observaron 2 grupos y se importaron 3 infecciones. En MZ, el grupo 2 se asoció con personas que salieron de su comunidad (OR = 4.33); y la parasitemia de la infección (OR = 1.10), infecciones causadas por el grupo 1 (OR = 4.83), y vivir con alguien que salió de la comunidad (OR = 4.48) aumentó el riesgo de síntomas. Esto sugiere que la movilización humana fomenta el ingreso de nuevos grupos genéticos en MZ, por lo que se requieren estrategias que consideren este factor para el control de Malaria.

Palabras clave: Malaria, *Plasmodium vivax*, Diversidad Genética, Estructuración Poblacional, Migración, Movilización Humana.

ABSTRACT

Plasmodium vivax is the main cause of malaria in the Peruvian Amazon, and with its rising incidences population genetic tools could be key components for the design and targeting of control strategies. 173 infections caused by *P. vivax* were genotyped in 2 basins, Mazan (MZ) and Napo (NP). Genetic diversity, the number of genetic clusters, and foreign infections were determined. MZ presented 35.2% of polyclonal infections, high genetic diversity ($He = 0.71$), and lower linkage disequilibrium (I_A^S 0.07) with respect to NP (18.8%, 0.39, 0.21). Five genetic clusters and 20 foreign infections were identified in MZ; in NP only 2 clusters were observed and just 3 infections were imported. In MZ cluster 2 was associated with people who traveled outside their community (OR = 4.33, $p = 0.04$); and the parasitemia of the infection (OR = 1.10, $p = 0.04$), being infected by a parasite from cluster 1 (OR = 4.83, $p = 0.03$), and living with someone who traveled outside (OR = 4.48, $p = 0.01$) increased the probability of presenting symptoms. Our results suggest that human mobilization encourages the entry of new genetic clusters in MZ, so malaria control strategies that consider this factor are required.

Keywords: Malaria, *Plasmodium vivax*, Genetic diversity, Population structure, Migration, Human mobilization.

I. INTRODUCCIÓN

Malaria es una de las enfermedades tropicales que causan mayor morbilidad y es un serio problema de salud pública en el Perú y a nivel mundial. A diferencia de la tendencia de otros países donde se observa una disminución global de casos, en el Perú hay un aumento progresivo de la malaria desde el año 2011 (1, 2), llegando en el 2014, a más de ~65 mil casos, localizados en su mayoría (94%) en la Región Amazónica. De las cinco especies de *Plasmodium spp.* que causan esta enfermedad, en el Perú circulan principalmente dos, *P. falciparum* y *P. vivax*, siendo este último el responsable de más del 80% de los casos en el país.

Este parásito, *P. vivax*, posee ciertas características biológicas que favorecen su transmisión (3), entre las cuales resaltan su bajo rango de temperatura de desarrollo en el mosquito (4, 5), las bajas parasitemias, el rápido desarrollo de inmunidad, la producción temprana y continua de gametocitos (4, 6, 7), y quizás la más resaltante, la presencia de hipnozoitos en el hígado (8, 9). Estas características permiten que este parásito persista pese a los esfuerzos de control realizados, representando un desafío para las estrategias actuales.

El Perú es considerado un país hipoendémico (10), y operacionalmente el control de esta enfermedad se lleva a cabo a nivel sub-nacional por regiones o distritos (2, 11). Sin embargo, la intensidad de transmisión no está distribuida de manera homogénea a nivel geográfico (11-15), por lo que medidas basadas en métricas globales o muy generalizadas, resultan no ser efectivas y económicamente poco rentables. Esta

heterogeneidad espacial en la micro-epidemiología de malaria se debe a que la transmisión ocurre en pequeñas unidades geográficas denominadas “puntos calientes” o “*hot-spots*” (12, 13), dentro de los cuales el riesgo de exposición se ve influenciado por factores como la proporción de niños en el hogar, estatus socio-económico, hábitos o actividades de riesgo del huésped, la falta de uso de medidas de control vectorial, distribución y comportamiento del vector, el tipo de vivienda y factores micro-ambientales. Es así que para que el control y eliminación de esta enfermedad sea efectivo, las herramientas de monitoreo no sólo deben coleccionar información sobre la presencia de focos de transmisión, sino también deben identificar los patrones de transmisión, considerar la heterogeneidad temporal y espacial de la misma, y elucidar los factores de riesgo que favorecen la propagación de las poblaciones parasitarias (16, 17).

El compromiso del gobierno peruano por eliminar la malaria se ha incrementado desde el año 2015 (2), y es así que actualmente la vigilancia de la malaria en el país se realiza de manera rutinaria a través de la detección pasiva de casos (DPC), y la implementación excepcional de detección activa de casos (DAC) en comunidades de alto riesgo basado en el índice de positividad de lámina (IPL) y la prevalencia de la enfermedad (2). Si bien los cambios en la incidencia y prevalencia de la malaria son indicadores clave en cualquier programa de control por que permiten detectar focos de transmisión residual (15), estas métricas por si solas no otorgan la información suficiente para revelar los factores involucrados en la transmisión, y así, focalizar y optimizar los esfuerzos (3).

La epidemiología molecular y el empleo de herramientas genético poblacionales no solo permiten realizar estimaciones más sensibles de la prevalencia e incidencia de la enfermedad, o evaluar la efectividad de las estrategias de intervención; sino también, brindan información sobre el aporte individual de los factores que favorecen la transmisión, y evalúan los patrones demográficos de las poblaciones parasitarias los cuales son influenciados por la movilización humana (18). Esto otorga información valiosa para adecuar las estrategias de control de manera focalizada y pertinente de acuerdo al contexto micro-epidemiológico.

De esta manera los estudios genético-poblacionales en el Perú han mostrado que la transmisión de *P. vivax* en nuestra Amazonía tiene un patrón espacial y temporal heterogéneo en donde confluyen escenarios ribereños aislados, periurbanos o de carretera, con alta o baja movilidad humana (2, 19-22). Esta alta variabilidad ha provocado que las poblaciones de *P. vivax* en áreas periurbanas y altamente móviles, sean muy diversas y estructuradas a nivel genético, y que la migración tenga un fuerte impacto en la diversidad y la intensidad de transmisión (19). Mientras que zonas geográficamente aisladas o poco móviles, presentan una fuerte diferenciación genética frente a otras poblaciones geográficas de la misma región, y su patrón de estructuración se mantiene constante en el tiempo (20). Si bien los migrantes pueden incrementar la diversidad, hay poca evidencia que indique que estos se lleguen a fijar en las poblaciones donde ingresan, y además, este movimiento migratorio micro-geográfico podría tener consecuencias en la efectividad de las estrategias de control; por lo que es necesario el desarrollo y uso de herramientas que permitan un

monitoreo adecuado y dilucidan los distintos patrones de transmisión presentes en nuestra Amazonía.

Mazán es considerado un distrito de muy alto riesgo de transmisión, habiendo reportado 1,954 casos en el 2014 (11). El centro poblado Mazán es la comunidad más grande del distrito y está ubicada en la cuenca del río Napo; sin embargo, no es el área con mayor intensidad de transmisión del distrito. De hecho, las comunidades que presentan mayor transmisión en este distrito están ubicadas a lo largo de las cuencas de los ríos Mazán y Napo; y entre los meses de Marzo y Abril del 2015, en las comunidades de Liberta, Primero de Enero, Gamitanacocha y Urco Miraño, más del 20% de la población de estas comunidades acudió al centro de salud con episodios febriles o síntomas de malaria, y el IPL estuvo por encima del 5% por más de 8 semanas consecutivas. Es por ello que en este trabajo se evaluó la distribución espacial a nivel micro-geográfico de distintos parámetros genético poblacionales en estas cuatro comunidades ribereñas del distrito de Mazán. También se describen los factores socio-demográficos, y clínico-epidemiológicos asociados a la expansión de sub-poblaciones genéticas específicas en cada comunidad, y los efectos de la introducción de nuevas poblaciones genéticas sobre las manifestaciones clínicas de la población susceptible.

II. PLANTEAMIENTO DEL PROBLEMA

La transmisión de la malaria en el país ha seguido un patrón intermitente a lo largo de los años, el cual se ha visto fuertemente influenciado por factores climatológicos como el Fenómeno el Niño (ENSO), el cambio de políticas públicas e implementación de nuevas estrategias (cambio de fármacos terapéuticos, uso de nuevas herramientas de monitoreo, o cambio en las estrategias de búsquedas de casos), y cambios en la distribución, comportamiento y características biológicas y demográficas del hospedero humano y el vector (1, 2).

Si bien algunas de las estrategias implementadas mostraron tener un efecto positivo aparente (solo evidenciado por la disminución de casos clínicos) en el corto plazo; la falta de continuidad por decisiones políticas; el desconocimiento del efecto real de estas intervenciones (efecto sobre aquellas infecciones que escapan la vigilancia tradicional); y el desconocimiento de las características del patógeno, el vector y el humano que favorecen la transmisión de la enfermedad, han hecho que las medidas tomadas anteriormente fracasasen (2). Todo el terreno ganado durante el periodo en el que el Proyecto, financiado por el Fondo Global, Malaria en Zonas de Frontera “PAMAFRO” se llevó a cabo, el cual generó que el número de casos disminuyera hasta 23, 060 casos en el 2011 (23), rápidamente se perdió en cuanto disminuyó el control intenso realizado, llegando a 64, 676 casos en el 2014, y a 63, 273 en el 2015 (año en que se llevó a cabo la recolección de muestras para este trabajo).

Las principales herramientas de monitoreo solo permiten medir la intensidad de transmisión e identificar posibles focos transmisión residual en el tiempo y espacio. Sin embargo, para que el control de una enfermedad sea efectivo no solo es necesario coleccionar información del el número de infecciones o personas expuestas, sino también el uso de herramientas que permitan identificar los patrones de transmisión y que consideren la heterogeneidad temporal y espacial de la misma, para así caracterizar y determinar las subpoblaciones o linajes que están adaptados a microambientes específicos y si estos se mantienen en el tiempo (16, 17).

En este sentido los marcadores moleculares son capaces de mapear la ruta de dispersión de los parásitos, y discriminar entre infecciones importados y rastrear su origen, o si son infecciones locales, definir si han sido recientemente introducidos (linajes de primera generación) o si son autóctonas de la región (segunda o más generaciones). Esto permitiría tener un conocimiento más profundo de los factores involucrados en la transmisión de esta enfermedad, y desarrollar estrategias de control específicas y más efectivas para combatir la Malaria en la Amazonía Peruana.

III. MARCO TEORICO Y ANTECEDENTES

Malaria

Malaria es una enfermedad causada por parásitos del género *Plasmodium* y transmitida por mosquitos hembra del género *Anopheles* (24, 25). Las 5 especies que causan la enfermedad en el humano son *P. falciparum*, *P. vivax*, *P. malariae*, *P. ovale* y *P. knowlesi*, y su distribución geográfica varía dependiendo de la especie. Es así que *P. falciparum* es la especie predominante en el continente africano, mientras que fuera de África *P. vivax* es la especie que genera mayor morbilidad.

Los primeros registros históricos de esta enfermedad datan quizás de hace más de 3000 años, y pese a que desde entonces se habían descrito muchos casos de personas con esplenomegalia, episodios periódicos de fiebre, dolor de cabeza, escalofríos y debilitamiento (comúnmente conocido como “mal aire”), el agente causal no fue identificado sino hasta el año 1880 por Alphonse Laveran, quien describió unos cuerpos de forma curvada (Gametocitos) en los glóbulos rojos, y células ovaladas con flagelos móviles (microgametos) en sangre fresca derramada de soldados con síntomas de malaria (26-28). En 1886 Camillo Golgi demostró que el parásito se reproduce asexualmente en la sangre, y asoció los episodios de fiebre con la ruptura de los glóbulos rojos (26). Años más tarde, en 1896 William MacCallum y Eugene Opie encontraron las mismas morfologías en sangre de loros y cuervos infectados con *Haemoproteus* (parasito relacionado a malaria en aves), y postularon que la formación de los gametocitos ocurre en la sangre, mientras que cuando la sangre es

tomada por un mosquito se da la liberación de los gametos y la fertilización (26-28). Sin embargo, fueron Ronald Ross y Giovanni Battista Grassi quienes resolvieron el ciclo de transmisión de la enfermedad al demostrar la transmisión desde el hospedero vertebrado al invertebrado y viceversa, e incriminaron a los mosquitos hembras del género *Anopheles* como los vectores que transmiten malaria al humano (29-31).

Los estudios sistemáticos y filogenéticos ubican al género *Plasmodium* dentro del infrareino Alveolata, filo Apicomplexa, clase Aconoidasida, y orden Haemosporidia. Además, estos estudios muestran que las especies causantes de malaria en el humano no están estrechamente relacionadas unas con otras. De hecho *P. falciparum* (junto a *P. reichenowi*) no forma siquiera un grupo monofilético junto con las otras especies de *Plasmodium* que afectan a primates (32). El origen de *P. vivax* es también controversial, ya que, si bien inicialmente se postuló que su origen se dio en África, debido a la ausencia de expresión del antígeno Duffy en poblaciones africanas (esta ausencia confiere resistencia a la invasión por este parásito), estudios recientes muestran que su origen se dio hace 2 a 3 millones en el continente asiático, lo cual abre el debate respecto al porqué existe una alta frecuencia de la ausencia de expresión del antígeno Duffy en África (33, 34).

Ciclo biológico de malaria y de *P. vivax*

Todas las especies de *Plasmodium* que causan malaria en humanos son transmitidos por la picadura de un mosquito hembra grávido del género *Anopheles* infectado. Este inocula alrededor de 25 esporozoítos haploides presentes en sus glándulas salivales al

momento de su alimentación, los cuales viajan a través del torrente sanguíneo de la persona hasta las células hepáticas en un tiempo promedio de una hora (27, 28). Al llegar a las células del hígado empieza su multiplicación asexual exoeritrocitaria, la cual genera más de 10 mil células infectivas denominadas merozoítos, y esta fase dura alrededor de 12 días. En el caso de *P. vivax* estadios durmientes llamados hipnozoítos pueden persistir en el hígado por meses o incluso años (35, 36).

Una vez concluido el ciclo exoeritrocítico, los merozoítos generados son liberados en el torrente sanguíneo e invaden a los glóbulos rojos (*P. vivax* solo invade a los reticulocitos, glóbulos rojos inmaduros nucleados), en donde se diferencian en trofozoítos y entran en división mitótica para formar los esquizontes sanguíneos. Finalmente, los nuevos esquizontes lisan la célula huésped liberando nuevos merozoítos al torrente sanguíneo formando así el ciclo eritrocitario. Los merozoítos liberados pueden invadir nuevos glóbulos rojos no infectados, y en cada ciclo la densidad parasitaria incrementa hasta unas diez veces su concentración (7, 37). En esta etapa la infección se mantiene aún asintomática hasta que se llega al umbral pirogénico del individuo infectado (densidad parasitaria mínima requerida para generar fiebre); donde las toxinas liberadas por la ruptura de los esquizontes (por ejemplo: el glicosil-fosfatidil-inositol o GPI del parásito, motivos de ADN, y la hemozoína, subproducto de la degradación del grupo hemo de la hemoglobina) son reconocidas por las células fagocíticas del huésped, generando una respuesta inmune proinflamatoria y el desencadenamiento de los síntomas (37-39).

Uno de cada 156 merozoítos generados optan por comenzar la gametocitogénesis y no continuar con el ciclo eritrocitario (4). Para ello se requiere un compromiso de todos los merozoítos de un esquizonte para diferenciarse en gametocitos machos o hembras. La información necesaria para el desarrollo de los gametocitos se encuentra en el cromosoma 9 y se han identificado entre 250 a 300 genes específicos de este estadio (4). El desarrollo de los gametocitos se subdivide en 5 estadios morfológicamente diferenciados. Las principales características morfológicas son la presencia del citoesqueleto subpelicular de microtúbulos y la doble membrana, y las diferencias más aparentes entre el micro y macrogametocito se observan en el estadio 5. En el caso de *P. falciparum*, solo el estadio 5 se puede visualizar en sangre periférica, mientras que en *P. vivax* todos los estadios pueden ser visualizados debido a la ausencia de proteínas de citoadherencia. A partir del día 6 desde el inicio de la formación de los gametocitos, estos quedan arrestados en G0 y solo se da la síntesis de ARN. Una vez llegado al estadio 5, toma entre 2 a 3 días para que los gametocitos se vuelvan infectivos, y la replicación se reactiva sólo cuando el gametocito entra al intestino medio del mosquito, y se concluye con la formación de los macro y microgametos.

La gametocitogénesis puede ser inducida por múltiples factores entre los cuales se encuentran la inmunidad del hospedero (6), la densidad de parásitos asexuados, concentración de hemoglobina en sangre, y la presencia de reticulocitos (*P. falciparum*). En el caso de *P. vivax* la gametocitogénesis puede empezar desde la primera generación de merozoítos (en el hígado), y estos son detectados por

microscopia dentro de los 3 días desde que el primer parasito asexuado ha sido visto; sin embargo, el tiempo de vida de sus gametocitos es menor que el de *P. falciparum*. Estudios con pacientes con neurosífilis mostraron que la producción de gametocitos puede ser continua a lo largo del primer mes de infección y su transmisión hacia los mosquitos se puede dar incluso a concentraciones submicroscópicas (40).

Una vez el mosquito ingiere los gametocitos las oportunidades de infección dependen de la madurez de los mismos, la concentración de ambos sexos, y la respuesta inmune del humano y del vector (24). Al cabo de 15 – 20 minutos de la ingesta de los gametocitos por el mosquito hembra, los gametocitos machos reinician su replicación y se dividen hasta formar 8 microgametos flagelados. Este proceso se denomina exoflagelación y dura alrededor de 10 – 12 minutos, y es así como empieza el ciclo esporogónico (41). Los microgametos se desplazan en busca del macrogameto para llevar a cabo la fecundación y formar el cigoto. Esta nueva célula diploide se diferencia en un ookinete móvil en un periodo de 15 - 30 horas, y luego el ookinete se desplaza desde la pared del intestino medio del mosquito hasta la membrana basal, en donde secreta una fina membrana quística y forma el oociste (4, 42). Cuatro a quince días después de la ingesta de sangre por parte del mosquito empieza la formación de los esporozoítos. Cada oociste puede liberar hasta 5 mil esporozoítos en el hemocele del mosquito, sin embargo, solo un 20% llegará a las glándulas salivales. Tanto para *P. falciparum* como para *P. vivax* el ciclo esporogónico dura de 8 a 10 días a 28°C, y 16 días a 20°C. Sin embargo, *P. vivax* puede desarrollarse incluso a

14.5°C, mientras que *P. falciparum* solo por encima de 16°C, lo cual ha favorecido la amplia dispersión geográfica de *P. vivax* incluso en zonas templadas (4, 42).

Características de la transmisión de *P. vivax*

El intrincado ciclo biológico de *P. vivax*, revela ciertas características biológicas favorables que este parásito posee en comparación a *P. falciparum* (3), y que le han permitido sostener su transmisión, incluso en escenarios de control intensivo, generando grandes dificultades y retos para su control y eliminación. Quizás la característica más resaltante es la presencia de hipnozoítos (36, 43), los cuales permanecen latentes en el hígado del huésped humano, y aún no se conocen los factores que desencadenan su reactivación. Los episodios ocasionados por la reactivación de los hipnozoítos se denominan relapsos. En escenarios de alta transmisión, y en individuos con un largo y constante historial de exposición se ha observado que los hipnozoítos se reactivan en periodos cortos de tiempo (~3 meses desde la infección primaria) y que los relapsos son mayoritariamente heterólogos (9, 43, 44), es decir que el genotipo causante del episodio de relapso es genéticamente distinto al genotipo de la primera infección. En cambio, en baja transmisión, el tiempo de reactivación tiende a ser mayor (más de 6 meses o incluso años), y los relapsos suelen ser homólogos (8, 45-47), los genotipos de la infección primaria y el relapso son iguales. Esta capacidad de mantenerse oculto por periodos largos, permite al parásito poder evadir las estrategias de control implementadas, y además otorga una mayor capacidad de dispersión geográfica.

De esta manera la presencia de hipnozoítos tiene un fuerte efecto sobre el intercambio génico de las poblaciones parasitarias, ya que se ve favorecido el desplazamiento geográfico de los parásitos (3, 16). También incrementa la diversidad genética debido a que se incrementan las posibilidades de que ocurran infecciones policlonales en el hospedero humano y posteriormente la recombinación en el mosquito (48, 49). Por último, el patrón de estructuración de la población parasitaria se puede ver oscurecido debido a que los relapsos otorgan una fuente adicional de parásitos que pueden o no ser similares a las poblaciones parasitarias que se están transmitiendo de forma activa.

Otra característica biológica importante en *P. vivax* es la baja densidad parasitaria asociada con su infección, la cual en parte se explica por su invasión restringida solo a reticulocitos (7). Sin embargo, esta baja carga parasitaria también se ve favorecida por la rápida adquisición de inmunidad no esterilizante que se ha observado en infecciones causadas por *P. vivax* respecto a las infecciones causada por las otras especies de *Plasmodium*. De hecho, *P. vivax* muestra una mayor respuesta anti-inflamatoria que *P. falciparum* la cual esta mediada por la citoquina IL-10 (38). Estos dos factores hacen que las poblaciones humanas expuestas a este parásito rápidamente adquieran inmunidad, se conviertan en portadores asintomáticos, y posteriormente controlen la densidad parasitaria generando una alta proporción de infecciones submicroscópicas (13). Esta característica confiere a las poblaciones parasitarias la capacidad de escapar de las estrategias de búsqueda y manejo de casos (principalmente focalizadas en el portador sintomático), y a las herramientas de

detección (basadas en la microscopía como prueba referencial) favoreciendo así su transmisión y dispersión.

Otra característica adicional que favorece la transmisión de este parásito es la temprana y continua producción de gametocitos desde antes que el individuo infectado presente síntomas y acuda al centro de salud para el suministro de su medicación (4). Este factor es importante debido a que permite, incluso a las infecciones sintomáticas, poder evadir las estrategias establecidas para el control de la enfermedad. A esto se le debe agregar que las infecciones asintomáticas, tanto aquellas que son asintomáticas por que aún no llegan al umbral pirogénico del individuo infectado como aquellas infecciones asintomáticas producto de la inmunidad no esterilizante, son capaces de generar densidades gametocitarias suficientes como para poder transmitir estos estadios sexuales al mosquito y así continuar el ciclo de transmisión (40). Además, esta capacidad de transmitir la enfermedad antes de recibir el tratamiento, genera que el efecto de selección que ejerce la administración de la droga sobre las clonas resistentes y susceptibles no sea tan pronunciado, y que la fijación de alelos resistentes sea lenta y poco percibida (50). Este efecto no solo se observa en características genéticas favorables para el parásito sino también en otros caracteres neutrales, y por ello las poblaciones de *P. vivax* tienden a fijar con menor facilidad alelos, y mantienen su diversidad poblacional con mejor eficacia que las poblaciones de *P. falciparum*.

Por último, se destaca la capacidad de este parásito para poder mantener su ciclo biológico a temperaturas templadas (14.5°C). Esto le ha permitido poder tener una

distribución geográfica más amplia a nivel global, e invadir territorios no habitados por *P. falciparum* (33). Todas estas características en su conjunto facilitan la transmisión de *P. vivax* y generan serias dificultades para su control y eliminación; no obstante, para que la transmisión se dé de manera efectiva es necesario la interacción entre el vector y el hospedero humano bajo un contexto ecológico y epidemiológico favorable.

Los vectores que transmiten malaria pertenecen al género Anopheles. Existen ~50 especies que pueden transmitir malaria al humano, y *An. darlingi* es el principal transmisor de malaria en la Amazonía Peruana (24, 25, 51). Este es un mosquito ribereño preferentemente rural. Su ciclo biológico está conformado por 4 fases: huevo, larva, pupa y adulto; y las tres primeras fases se dan en medio acuático y se pueden prolongar hasta por dos semanas. Desde el punto de vista del vector (capacidad vectorial), la transmisión de malaria depende de la densidad poblacional de los mosquitos, la disponibilidad de fuentes de agua adecuadas para el crecimiento de sus larvas, su comportamiento alimenticio y preferencias de picadura, su ciclo gonotrófico (tiempo desde que el mosquito es fecundado, se alimenta y oviposita), su longevidad, la eficacia de infectividad de los mosquitos, su rango de vuelo, y los factores ambientales y micro-climatológicos que influyen a su ciclo biológico (24).

La densidad poblacional del vector tiende a ser estacional y estar fuertemente influenciada por el ciclo anual de lluvias, pero su patrón puede variar entre regiones geográficas, e incluso entre comunidades cercanas (52, 53). El incremento de lluvias trae consigo mayor disponibilidad de cuerpos de agua, los cuales son necesarios para

la crianza de las larvas; sin embargo, si las lluvias son intensas, la corriente de los ríos puede remover los sitios de crianza, y solo cuando las lluvias cesan, las zonas inundadas permiten la propagación del vector. Como se puede deducir este efecto del clima sobre el crecimiento de la población de mosquitos no es universal, sino más bien específico de la región geográfica. Otros factores exógenos (deforestación, piscigranjas, minería) que favorecen la formación de cuerpos de agua estables y adecuados para la crianza de las larvas, pueden también influenciar la dinámica temporal de la densidad poblacional del vector.

Además, las características físicas, químicas y biológicas de los cuerpos de agua donde crecen las larvas de *An. darlingi* son variables, y dependen del tipo de criadero y la topografía donde se ubican (lagos, lagunas, ríos grandes o chicos, bosques inundados, piscinas de agua entre otros) (25). Aun así, *An. darlingi* tiene preferencia por cuerpos de agua grandes y profundos, con poca cobertura vegetal (~24%), una temperatura de 20°C a 28°C, y un pH de 6.5 a 7.3.

En diversos estudios se ha mostrado que *An. darlingi* tiene una baja tasa de infección (proporción de mosquitos infectado), sin embargo, su capacidad vectorial es elevada debido a su alta tasa de picadura al humano (proporción de mosquitos que pican al humano), su comportamiento principalmente antropofílico, y su corto ciclo gonotrófico (~48 horas) (24).

Un factor importante en la selección de un método apropiado para el control vectorial es el comportamiento durante la alimentación y reposo de los mosquitos (54); y en

ese sentido *An. darlingi* muestra una alta diversidad y plasticidad en su comportamiento. Si bien este mosquito es principalmente antropofílico, diversos estudios han demostrado que este mosquito es oportunista, pudiéndose alimentar de aves de corral y otros animales domésticos en ausencia de sangre humana (51). Un estudio realizado en la Amazonía Peruana mostró que la segunda fuente de sangre más importante después de la humana es la sangre de gallinas (ave de corral bastante común en las viviendas de la amazonia), y además el nivel de antropofilia vario tanto a nivel geográfico como temporal (51). Esta plasticidad no solo es respecto a sus preferencias de fuente de sangre, sino también en su lugar y hora de alimentación y reposo (24, 54). Algunos mosquitos suelen picar entre las 21:00 – 24:00 horas en el interior de los hogares (comportamiento nocturno endofágico), mientras que otros pican en horas de la madrugada (04:00 – 06:00) y fuera de las casas (exofagia). Por otro lado, luego de la ingesta de sangre, algunos mosquitos prefieren reposar dentro de los hogares (endofilia), mientras que otros reposan fuera de ellos (exofilia). Este comportamiento varía a nivel temporal y geográfico, y está influenciado principalmente por las características del hogar, factores micro-climáticos, el ciclo gonotrópico, la disponibilidad de la fuente de sangre, y la presencia de agentes repelentes o tóxicos como los insecticidas. Prussing, Moreno (55) mostró que *An. darlingi* puede cambiar su comportamiento en respuesta a la presencia o ausencia de medidas de control vectorial. En presencia de redes impregnadas con insecticidas de larga duración en las habitaciones, los mosquitos mostraron un comportamiento exofágico/exofílico, pero cuando el efecto de las redes impregnadas se disipó (al cabo

de 6 meses) los mosquitos gradualmente cambiaron su comportamiento a endofágico/endofílico.

Finalmente, la conducta del hospedero humano y su demografía también juegan un papel importante en la transmisión de *P. vivax*. Desde un punto de vista operativo, la vigilancia y control de la enfermedad se dan a nivel distrital sin tomar en consideración que su transmisión es espacialmente heterogénea (2, 11, 14). Si bien el término foco de transmisión hace referencia a un área de varios kilómetros cuadrados que soporta o mantiene la transmisión de la malaria, incluso dentro de este limitado espacio geográfico la transmisión se da a lo largo de pequeñas unidades geográficas denominadas *hot-spots* (13). Estos *hot-spots* pueden ser una casa o grupo de casas donde la exposición a las picaduras de mosquitos es por encima del promedio, y el incremento en el riesgo de exposición se ve influenciado por factores como la proporción de niños en el hogar, susceptibilidad inmune del huésped, estatus socio-económico, la falta de uso de medidas de control vectorial, el tipo de vivienda o factores micro-ambientales. Estos factores que incrementan los niveles de exposición de los individuos, permiten la adquisición de una respuesta inmune protectora, incremento de infecciones con baja densidad parasitaria (infecciones sub-microscópicas) y ausencia de síntomas. La importancia de estas infecciones asintomáticas sub-microscópicas recae en su capacidad de ser los reservorios de esta enfermedad; sin embargo, aún no se conoce la densidad parasitaria y gametocitaria mínima requerida para que se dé la transmisión hacia el vector.

Vigilancia y control de *P. vivax*

El control efectivo de los casos de *P. vivax* depende del empleo adecuado de herramientas de detección de casos y monitoreo de la intensidad de transmisión, para el suministro oportuno de tratamiento y otras medidas de prevención. Las principales herramientas de monitoreo se basan en la detección de casos mediante el reporte rutinario por microscopia en los centros de salud (56). En el Perú la vigilancia se realiza de manera rutinaria a través de la detección pasiva de casos (DPC), y la implementación excepcional de detección activa de casos (DAC) en comunidades de alto riesgo basado en el índice de positividad de lámina y la prevalencia de la enfermedad (2). Es así, que a los individuos identificados como positivos por estas dos estrategias se les suministra el tratamiento recomendado por ministerio de salud (25 mg/kg de Cloroquina por 3 días, y 0.5mg/kg de Primaquina por 7 días). Sin embargo, esta aproximación subestima la transmisión real debido a la alta prevalencia de casos sub-microscópicos y asintomáticos, además del largo periodo de tratamiento y la falta de supervisión generan poca adherencia al tratamiento y elevadas tasas de recurrencias (57).

Otra medida de la transmisión es la tasa de inoculación entomológica (EIR: número de picaduras infecciosas de mosquitos por individuo por año); sin embargo, este método también carece de precisión por que la distribución geográfica de los mosquitos es heterogénea (58), el comportamiento del mosquito es bastante plástico y estimar la tasa de picaduras es sumamente difícil (59). La tasa de seroconversión es una medida robusta de la intensidad de transmisión, pero su efectividad es relativa al

tipo de antígeno utilizado, a su diversidad y a la inmunogenicidad de la población hacia el mismo. Por lo que el uso de esta herramienta debe ser adaptada a la población de interés (60).

Las herramientas previamente mencionadas permiten medir la intensidad de transmisión e identificar posibles focos transmisión residual en el tiempo y espacio. Sin embargo, para que el control de una enfermedad sea efectivo no solo es necesario coleccionar información del número de infecciones o personas expuestas, sino también el uso de herramientas que permitan identificar los patrones de transmisión y que consideren la heterogeneidad temporal y espacial de la misma, para así caracterizar y determinar las subpoblaciones o linajes que están adaptados a microambientes específicos y si estos se mantienen en el tiempo (16, 17). En este sentido los marcadores moleculares son capaces de mapear la ruta de dispersión de los parásitos, rastrear su origen, y discriminar entre infecciones locales e importadas.

Principios básicos de genética poblacional

En su concepto más amplio la genética poblacional estudia la variación genética dentro y entre las poblaciones. Esto involucra el análisis y modelamiento de los cambios en las frecuencias de las variantes, también denominadas alelos o incluso genes, en las poblaciones en el espacio y tiempo (61, 62). La genética poblacional también hace inferencias de los procesos evolutivos estocásticos que moldean la variabilidad genética en el tiempo. En los estudios genético poblacionales los individuos son caracterizados por su genotipo (constitución genética) y su fenotipo

(rasgos físicos observables manifestados por el genotipo). No obstante, la relación entre el genotipo y el fenotipo es compleja debido a que un fenotipo puede ser el resultado de la interacción entre muchos genes y el ambiente donde la población se desarrolla (61, 62). Es así que mediante la construcción de un marco de trabajo genético en donde la evolución toma lugar, los principios de genética poblacional son esenciales para el estudio de los procesos biológicos bajo una perspectiva evolutiva amplia (61).

La variabilidad genética es la base de la genética poblacional, y por ello es importante mencionar 4 características fundamentales de ella. Primero se debe resaltar que la variabilidad genética es necesaria para la evolución y supervivencia de las especies. Ella les permite adaptarse de manera eficiente a los cambios del entorno donde la especie se desarrolla (sean estos cambios perjudiciales, beneficiosos o neutrales). Es así que, desde una perspectiva conservacionista el objetivo es mantener el acervo genético de las poblaciones para permitirles una correcta adecuación a su ambiente; mientras que desde el punto de vista del estudio de enfermedades infecciosas (y malaria no escapa de ello) el objetivo es completamente opuesto, y lo que se busca es que las estrategias de control que se implementen tengan un efecto negativo sobre la diversidad genética del agente infeccioso, y que no se seleccionen rasgos que le permitan evadir las estrategias de intervención (63).

En segundo lugar, es importante precisar que esta variabilidad se da en distintos niveles que ocurren a lo largo del genoma. Estos niveles van desde mutaciones puntuales (ya sea en regiones codantes o no codantes), motivos repetidos de ADN

(microsatélites y minisatélites), inserciones o deleciones (indeles), hasta macro-mutaciones a nivel cromosómico (inserción, deleción, duplicación, inversión o trasposición de regiones largas de ADN) (61, 62). Cada una de estas variaciones, llamadas mutaciones, ocurre en una posición específica en la cadena de ADN a la cual se le denomina locus (loci en plural), y a las variantes de un locus se les denomina alelo. Al conjunto de alelos de distintos loci analizados de un mismo individuo se le denomina genotipo, y al conjunto de alelos de distintos loci analizados de un mismo cromosoma se le denomina haplotipo. En el caso de *Plasmodium* como el parásito es haploide en el humano, el genotipo es igual al haplotipo debido a que todo el genoma actúa como si fuese un solo cromosoma.

En tercer lugar, cada locus y cada nivel de variación posee una tasa mutacional distinta (tasa en la que se introducen nuevas variantes), la cual depende de la tasa de error del proceso de replicación ADN, la tasa de error del sistema reparación, y la tolerancia de la especie para permitir mutaciones en un locus determinado. Esta tolerancia está determinada por el efecto que pueda tener el alelo para el gameto o cigoto recién formado. Si el efecto es neutro o beneficioso, la mutación podrá transmitirse a las siguientes generaciones; en cambio, si el efecto es deletéreo, la nueva variante no podrá ser transmitida. Esta condición de deletéreo, neutro o beneficioso es circunstancial, y depende de la interacción del nuevo fenotipo con su ambiente.

La tasa de mutación de un locus es importante, porque ella es utilizada por la genética poblacional y la biología evolutiva para responder fenómenos biológicos a distintas

escalas de tiempo y espacio (61, 62). Si lo que se desea es conocer un evento histórico ocurrido mucho tiempo atrás (cientos de miles o millones de años atrás), loci con tasas de mutación baja, y poca variabilidad serán más útiles para poder hacer inferencias sobre las relaciones existentes entre los individuos analizados. Si el evento de interés es reciente y/o ocurre en periodos cortos de tiempo, se necesitan loci con tasas de mutación más elevadas. Los loci utilizados en el marco del estudio de algún proceso biológico se denominan marcadores, y en el contexto del estudio epidemiológico de malaria los marcadores más utilizados han sido los microsatélites y los polimorfismos de un solo nucleótido (SNPs) debido a su fácil implementación y alta tasa mutacional.

El uso de microsatélites se ha difundido ampliamente en los últimos 20 años en distintas áreas de la investigación y estudios genético poblacionales (64). Los microsatélites son repeticiones en tándem de 2 a 3 nucleótidos generalmente esparcidos a lo largo del genoma en regiones no codantes, por lo que se presume que no están sujetos a presión de selección por parte de factores externos. Gracias a esto los microsatélites tienden a tener un alto grado de polimorfismo y múltiples alelos por cada loci (hasta más de 50 por loci) y una elevada tasa de mutación ($10^{-2} - 10^{-6}$), que en el caso de *P. falciparum* es de 1.59×10^{-4} (65), y en el caso de *P. vivax* está en el rango de 5.35×10^{-3} y 3.77×10^{-2} (66). Sin embargo, su variabilidad depende de su cercanía a regiones codantes sujetas a presión de selección, a la longitud del motivo (repeticiones de di, tri o tetra nucleótidos), al número de repeticiones, y a la complejidad del motivo de repetición (motivos simples conformados por un solo tipo

de repetición, o motivos complejos conformados por varios tipos de motivos a la vez).

Muchos de los sistemas de análisis descritos en genética poblacional fueron diseñados para trabajar con marcadores bi-alélicos poco polimórficos, por lo que los análisis basados en microsatélites en muchos casos pueden ser malinterpretados. Además, debido al elevado polimorfismo y a la alta tasa mutacional que presentan estos marcadores, es muy difícil seguir la ruta evolutiva que define la frecuencia de un alelo (67), y esta saturación de las mutaciones hace, en algunos casos, imposible definir si un alelo determinado se originó producto del incremento, disminución o estabilización en el número de repeticiones de su alelo predecesor (Homoplasia, o alelos idénticos en estado pero que poseen orígenes evolutivos distintos) (64, 68). Es por este motivo que muchos de los trabajos en los cuales se emplean este tipo de marcadores solo realizan un análisis descriptivo de los datos, pero no realizan inferencias a nivel evolutivo. Aun así, si se tienen microsatélites con tasas mutacionales conocidas y modelos de evolución claros es posible hacer inferencias válidas (32, 69).

Por último, como se mencionó anteriormente, la genética poblacional hace inferencia de distintos procesos evolutivos estocásticos mediante el uso de datos empíricos y modelos evolutivos explícitos. Estos procesos evolutivos son la mutación, migración, recombinación, deriva génica, y selección (61). La mutación, la migración (introducción de individuos foráneos a la población), y la recombinación son capaces de introducir nuevas variantes alélicas a la población; no obstante, la recombinación

también permite generar nuevos haplotipos en base a las variantes alélicas ya existentes. En cambio, la deriva génica y la selección moldean las frecuencias de los alelos presentes en la población, ocasionando que estos alelos se fijen, se mantengan o se pierdan en el tiempo. La forma como estas dos fuerzas evolutivas actúan es distinta y depende del tamaño de la población.

La deriva genética es el cambio aleatorio en las frecuencias alélicas producto del proceso de muestreo en la reproducción (70). Esta aleatoriedad no tiene direccionalidad o preferencia hacia algún alelo o rasgo genético; sin embargo, cuando los tamaños poblacionales son pequeños (producto de subdivisión poblacional o cuellos de botella en la población), la deriva genética fija algún rasgo de manera aleatoria. Por otro lado, la selección tiende a fijar aquellos rasgos genéticos que tienen un mejor fitness con su ambiente, es decir, fija aquellos rasgos que confieren alguna ventaja reproductiva a los individuos que la poseen. Y es importante recalcar que ambos procesos solo actúan sobre aquellas variantes alélicas ya existentes, y no son capaces de generar nuevas variantes por si solos.

Indicadores genético poblacionales en el estudio de malaria

En el contexto epidemiológico de malaria, los 4 actores principales de la transmisión de esta enfermedad, el vector, el humano, la biología del parásito y el entorno o contexto ambiental (aquí podemos incluir a los factores climatológicos y a las estrategias de intervención), tienen un efecto directo sobre los procesos evolutivos que afectan a la población parasitaria. Por lo que, al hacer inferencias sobre estos

procesos evolutivos, también hacemos inferencias de manera indirecta sobre los factores del vector, del humano, del parásito, y del ambiente que favorecen a la transmisión. De esta manera en malaria se han empleado diversos indicadores genéticos que permiten medir el efecto de la transmisión y de cada factor sobre la población de parásitos.

Uno de estos indicadores es la prevalencia de infecciones policlonales, o coexistencia de dos o más clonas distintas dentro del individuo infectado (18). Como el genoma del parásito en el humano es haploide, al genotipificar una muestra infectada se espera la presencia de tan solo un alelo por cada locus analizado, es decir la presencia de un solo haplotipo. La presencia de dos o más alelos en uno o más loci es el producto de la presencia de más de una clona parasitaria (dos o más haplotipos distintos) en la muestra. Esta multiplicidad de la infección puede deberse a la transmisión de dos o más genotipos distintos a partir de una misma picadura (coinfección), o la adquisición de parásitos por múltiples infecciones o picaduras independientes (super-infección); y ambos procesos están relacionados a la intensidad de transmisión.

Cuando la transmisión es baja, hay pocas picaduras infectivas por parte del mosquito al humano y viceversa (tasa de inoculación entomológica y tasa de infección respectivamente), por lo que existen pocas probabilidades de que una persona pueda adquirir una super-infección o coinfección, y por ello la mayoría de infecciones son monoclonales (presencia de una sola clona en la infección) (71). Cuando se dan las condiciones necesarias para el incremento de la población de mosquitos y se favorece

la interacción entre este y el humano, también se incrementan las tasas de infección de los mosquitos y de inoculación entomológica. Esto ocasiona un incremento en la prevalencia de personas infectadas, y con ello se incrementan las posibilidades de ocurrencia de super-infecciones y coinfecciones, generando un aumento de la prevalencia de infecciones policlonales (49).

Otro indicador genético bastante utilizado en los estudios de genética poblacional de los parásitos de *Plasmodium* es el desequilibrio de ligamiento o asociación no al azar de alelos de loci distintos (18). Los mecanismos de herencia de los alelos siguen la ley de segregación independiente de Mendel (72, 73), por lo que si analizamos dos loci bi-alélicos con la presencia de los alelos A, a, B y b, esperamos observar cuatro posibles haplotipos en la población: AB, Ab, aB y ab. Como la segregación de alelos es independiente, la frecuencia de estos cuatro haplotipos está definida por la frecuencia de los alelos que los conforman (73), y así por ejemplo la frecuencia del haplotipo AB (f_{AB}) es igual a la multiplicación de las frecuencias de los alelos A (f_A) y B (f_B) en la población.

Sin embargo, existen muchos factores que pueden afectar la independencia de la herencia de los alelos en una población, ocasionando que los alelos de loci distintos se segreguen juntos y no al azar (73). De esta forma el desequilibrio de ligamiento se define, de manera formal, como la diferencia (D) entre la frecuencia del haplotipo observada (f_{AB}) menos la frecuencia del haplotipo esperada ($f_A f_B$). Si D es igual a cero, entonces los alelos de los dos loci analizados se encuentran en equilibrio de ligamiento; en cambio, si D es diferente de cero, los alelos de los dos loci analizados

están en desequilibrio de ligamiento (LD). Existen múltiples coeficientes empleados para medir el LD , siendo el índice de asociación estandarizado (I_A^S) el más empleado en el estudio de *Plasmodium* debido a que representa una medida de resumen del valor de D entre alelos de múltiples loci analizados (74).

Como observamos anteriormente, en escenarios de baja transmisión se favorecen las infecciones monoclonales, por lo que cuando una persona infectada es picada, un solo haplotipo es transmitido y se da la autofecundación del parásito en el mosquito (reproducción sexual de dos haplotipos genéticamente idénticos o relacionados). Este mecanismo de propagación y reproducción por autofecundación es denominado propagación clonal, y causa que todos los alelos de un haplotipo se segreguen de manera conjunta y no de forma independiente, incrementando así el desequilibrio de ligamiento. Por ello la presencia de LD o tener un valor de I_A^S distinto de cero refleja la presencia de autofecundación y propagación clonal en la población (18).

El tamaño efectivo de la población parasitaria (número de individuos de la población que explican la magnitud del efecto de la deriva genética en la población) y junto a ello su diversidad, también se ven influenciados por la transmisión (18, 75). El incremento en la transmisión es reflejo del éxito reproductivo de los distintos linajes de parásitos circulantes y el incremento del tamaño efectivo poblacional de los mismos. Bajo estas circunstancias se favorece la introducción de nuevas variantes en la población (incremento de la tasa mutacional poblacional o θ), lo cual incrementa el número de sitios polimórficos, la riqueza de alelos de cada sitio polimórfico (número de alelos por locus), y la diversidad genética en términos de heterocigosidad (H_{exp}) de

la población. En cambio, cuando baja la transmisión y se reduce el tamaño efectivo poblacional, la deriva genética ocasiona la fijación aleatoria de alelos, y se reduce la diversidad de la población. Sin embargo, el tamaño efectivo de la población y su diversidad también se ven fuertemente influenciados por la introducción de nuevas variantes producto de la migración (75) y la estructuración genética de las poblaciones parasitarias, por lo que la relación entre estas métricas y la transmisión no necesariamente es lineal.

Por último está la presencia de la estructuración genética o subdivisión poblacional. A diferencia del concepto de población que se maneja de manera tradicional en epidemiología o ecología, donde una población es el conjunto de individuos que coexisten en un mismo lugar y tiempo; en genética poblacional se define como población al conjunto de individuos que comparten una misma genealogía, están conectados por flujo genético, y donde el apareamiento entre todos los individuos que pertenecen a la población es aleatorio (72, 76). La subdivisión poblacional se debe a la presencia de aislamiento reproductivo (apareamiento no aleatorio) entre los individuos que conforman la población global. Este apareamiento no aleatorio se da como resultado de procesos naturales e históricos que incluyen el aislamiento geográfico, el comportamiento y preferencias reproductivas del parásito; y su interacción con la conducta de riesgo del hospedero humano, el comportamiento del vector, y los patrones de movilización de ambos hospederos (3, 16-18, 63, 77, 78).

El patrón de subdivisión se acentúa con la disminución de la transmisión; y en términos prácticos, observar infecciones causadas por parásitos que pertenecen a una

misma subpoblación, implica que los individuos que transportan estos parásitos comparten una misma conducta o factor de riesgo (ocupación, cercanía geográfica, hábitos de desplazamiento o recreación, entre otros). De esta manera, determinar el patrón de subdivisión poblacional permite incriminar y medir la contribución diferencial de cada uno de los factores de riesgos involucrados en el origen de los casos.

Existe una amplia variedad de métodos empleados para detectar la presencia de estructuración. Quizás el más tradicional, es el comúnmente llamado estadísticos F de Wright (F_{ST}), también denominados índices de fijación o índices de diferenciación (76). Cuando se da el proceso de subdivisión, la reducción del tamaño poblacional en las subpoblaciones fomenta la fijación de alelos de manera aleatoria en cada subpoblación debido la deriva génica, y esto conlleva a un incremento de la diferenciación genética entre las subpoblaciones. Esta diferenciación puede darse en niveles jerárquicos (infecciones dentro de comunidades, comunidades dentro de distritos, y distritos dentro de provincias), y de esta manera se describen tres coeficientes: F_{IT} que es la correlación de alelos dentro de los individuos de la población total; F_{IS} , correlación de alelos dentro de las subpoblaciones; y el F_{ST} , correlación de alelos entre las subpoblaciones. Muchas variaciones de estos estadísticos han sido propuestas para corregir sesgos debido al muestreo, a la tasa de mutación de los marcadores, o al modelo de sustitución de los mismos. Una de estas variaciones es el Análisis de Varianza Molecular (AMOVA), el cual ofrece un marco de trabajo para evaluar distintas hipótesis sobre los factores que favorecen la

subdivisión poblacional (79). Sin embargo, las principales limitantes de estas aproximaciones son que dependen de un conocimiento previo de los factores o fuentes que han ocasionado la subdivisión (rara vez se conoce esta información), y no permiten una asignación objetiva de los genotipos a las subpoblaciones propuestas.

Ante esta limitación se han propuesto múltiples métodos de agrupación basados en modelos, y entre ellos destaca la implementación de modelos basados en estadística bayesiana y el muestro mediante Cadenas de Markov Monte Carlo (MCMC) en el programa STRUCTURE (80). Si se muestrean N individuos de una población total, *a priori* no tenemos conocimiento del número de subpoblaciones (k) que puedan estar presentes, ni que individuos pertenecen a una misma población. Mediante el empleo MCMC, STRUCTURE simula distintos escenarios o formas de agrupar los individuos asumiendo diferentes valores de k , y en base a la probabilidad posterior de los parámetros de ancestría (coeficiente de ancestría o Q), y la proporción de mestizaje (α), este programa permite encontrar el modelo cuya probabilidad posterior se maximiza en base a la verosimilitud de los datos. De esta manera STRUCTURE estima las frecuencias alélicas dentro de cada subpoblación, el origen ancestral de cada individuo (Q), y el número de subpoblaciones presentes en la población total.

Conocer el patrón de estructuración, la asignación de infecciones a una subpoblación y la distribución geográfica de las subpoblaciones no solo permite conocer los factores epidemiológicos asociados a estas subpoblaciones, sino también permite conocer el origen de las infecciones y saber que infecciones su produjeron por transmisión local y que infecciones son importadas.

De esta manera en este trabajo de tesis se describe la distribución espacial a nivel micro-geográfico de distintas métricas genético poblacionales en tres comunidades de la cuenca del río Mazán, y de la comunidad de Urco Miraño en la cuenca del río Napo, todas estas comunidades ubicadas dentro de la jurisdicción del Distrito de Mazán. Para ello se estimaron y compararon los niveles de multiplicidad de infección, diversidad genética, desequilibrio de ligamiento y estructuración poblacional en las cuatro comunidades de interés. Además, se evaluaron los factores socio-demográficos asociados a sub-poblaciones genéticas específicas (denominadas haplogrupos en este trabajo) en cada micro-cuenca, y sus efectos sobre las manifestaciones clínicas de los individuos infectados.

IV. OBJETIVOS

Objetivo general

Describir la variabilidad genética de las poblaciones parasitarias de *P. vivax* en las comunidades de Libertad, Primero de Enero, Gamitanacocha y Urco Miraño ubicadas en el distrito de Mazán en la Amazonía Peruana; y determinar los factores demográficos y clínico-epidemiológicos asociados al patrón de estructuración genético poblacional del parásito observado. También se busca identificar las infecciones de origen autóctono, importado e introducido, y determinar su contribución y efecto en la transmisión de *P. vivax* en estas cuatro comunidades.

Objetivos específicos

- Genotipificación, con un panel de 16 microsatélites, de muestras positivas a *P. vivax* colectadas en los meses de Junio y Agosto del año 2015 en las comunidades de Gamitanacocha, Libertad, Primero de Enero, y Urco Miraño.
- Determinar las diferencias en la proporción de infecciones policlonales, diversidad genética y nivel de endogamia entre las cuencas y comunidades analizadas.
- Identificar las poblaciones genéticas de parásitos presentes en cada comunidad y cuenca, y discriminar entre **infecciones locales** (infecciones causadas por subpoblaciones genéticas de parásitos que se transmiten de manera activa dentro del área geográfica), **importadas** (infecciones causadas por

subpoblaciones o haplotipos que no se transmiten de manera activa en el área geográfica donde se identificó la infección) e **introducidas** (infecciones causadas por subpoblaciones genéticas que han ingresado hace una o dos generaciones del parásito al área geográfica donde se identificó la infección).

- Determinar los factores demográficos y clínico-epidemiológicos asociados a las subpoblaciones genéticas del parásito, y a la importación/introducción casos en cada cuenca analizada.
- Determinar el efecto del flujo migratorio (infecciones importadas e introducidas) a nivel micro-geográfico y de la introducción de nuevas subpoblaciones parasitarias, sobre las manifestaciones clínicas de la población susceptible.

V. JUSTIFICACIÓN

La efectividad del control de la malaria depende de la pertinencia de las estrategias de intervención, en base a un contexto epidemiológico específico. Para ello es necesario un conocimiento profundo del mecanismo y ruta de transmisión de esta enfermedad; conocer y entender los factores que afectan a este proceso; y determinar el impacto de las migraciones a nivel micro-geográfico en la transmisión, y en la eficacia de las intervenciones (16, 17). La transmisión de *P. vivax* en la Amazonía Peruana está caracterizada por tener un patrón heterogéneo a nivel espacial y temporal (en términos de su incidencia y prevalencia) (2, 11, 14, 20, 22, 47, 81), el cual es influenciado por múltiples factores (edad, sexo, características de las viviendas, ocupación humana, comportamiento del vector, presencia de cuerpos de agua favorables para la crianza de las larvas, entre otros) que varían también a nivel geográfico, micro-geográfico y temporal (11-13). Por lo que conocer donde y cuando un factor otorga una mayor contribución a la transmisión, ayuda a poder focalizar los recursos limitados que posee el estado para combatir esta enfermedad. En este sentido la genética poblacional complementa y enriquece la información proporcionada por las herramientas tradicionales empleadas en el monitoreo de la transmisión y podría contribuir a una mejor focalización y selección de los recursos necesarios para controlar y eliminar la transmisión de malaria.

VI. MATERIALES Y MÉTODOS

Consentimiento ético

Las muestras a utilizadas cuentan con la aprobación del Comité de Ética de la Universidad Peruana Cayetano Heredia a través de los proyectos denominados “Epidemiología molecular de *Plasmodium vivax* en cuatro comunidades ribereñas del distrito de Mazán en la Amazonía Peruana”, código SIDISI: 66638, “Estimación de la proporción de infecciones por malaria no detectadas por las intervenciones de detección activa de casos (DAC) en áreas endémicas de la Amazonía Peruana”, Código SIDISI: 64371; e “Impacto de los portadores asintomáticos en la epidemiología y control de malaria en la Amazonia Peruana”, código SIDISI: 57395. Los pacientes incluidos en el estudio cuentan con consentimiento informado por escrito según las Normas del Código de Ética y Bioseguridad. Todos los procedimientos fueron realizados siguiendo los lineamientos establecidos en la Declaración de Helsinki.

Área de estudio y población

Este estudio fue realizado en el distrito de Mazán, ubicado al norte de la ciudad de Iquitos en la provincia de Maynas del departamento de Loreto. El centro poblado Mazán (3.503° Sur, 73.094° Oeste, **Figura 1**) es la comunidad más grande del distrito y está ubicada en la cuenca del río Napo a unos 55 – 60 km de la ciudad de Iquitos (3 a 4 horas de viaje en bote). Las comunidades con mayor transmisión se encuentran

ubicadas a lo largo de los ríos Mazán y Napo, con distancias hasta la última comunidad dentro de la jurisdicción del distrito de 30 km y 160 km respectivamente (equivalente a recorridos entre 2 y 6 horas en bote desde el centro poblado).

Las comunidades seleccionadas para este estudio fueron Gamitanacocha (3.426° S, 73.318° O), Libertad (3.496° S, 73.234° O), Primero de Enero (3.494° S, 73.221° O) y Urco Miraño (3.361° S, 73.064° O). Un total de 154 casas y 820 individuos fueron integrados al estudio en las cuatro comunidades, y la distribución de las casas e individuos por comunidad fue la siguiente: Gamitanacocha (GC) = 19 casas y 92 individuos (ind); Libertad (LI) = 56 casas y 298 ind; Primero de enero (PE) = 20 casas y 67 ind, y Urco Miraño (UM) = 59 casas y 331 ind.

La población del distrito es predominantemente mestiza, de escasos recursos económicos, y entre las principales actividades laborales destacan la agricultura, la maderería y la pesca. La mayoría de las casas tienen al menos un cuarto cerrado, y los materiales predominantes de las paredes es madera, el techo es de hoja de palma o paja y del piso es madera. El clima es tropical lluvioso, con temperaturas mínimas de 17 a 20 °C en los meses de diciembre a marzo y una máxima de hasta 36 °C en los meses de Junio y Julio; y la humedad promedio es de 84% con fuertes lluvias de Noviembre a Mayo (82).

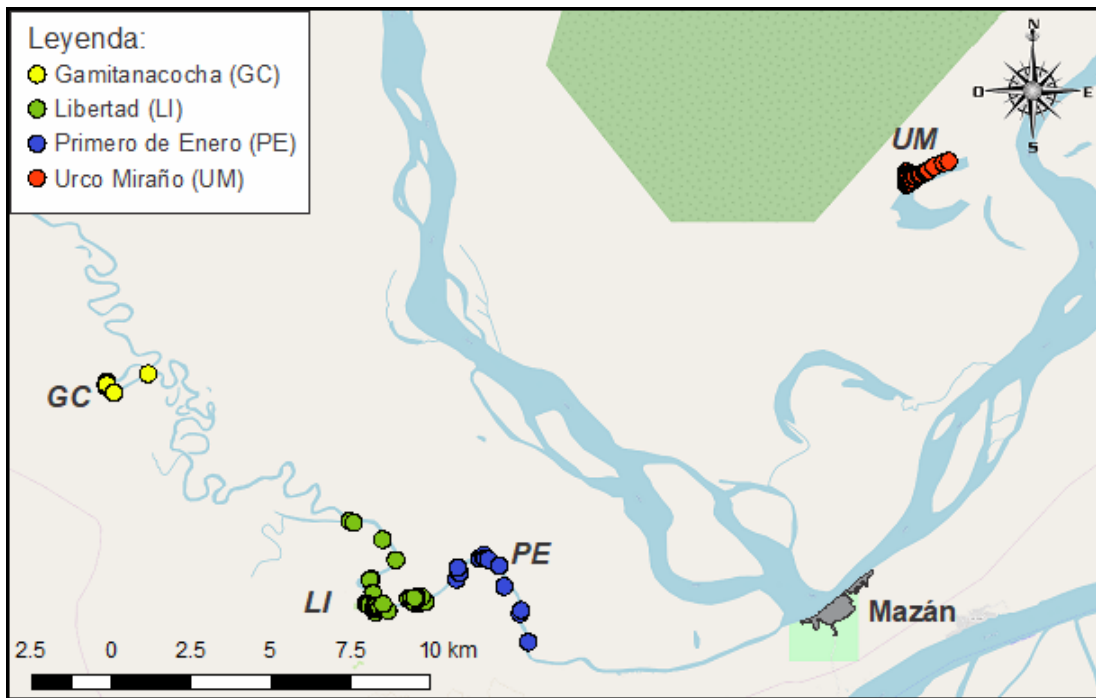


Figura 1: Distribución geográfica de las comunidades seleccionadas (Gamitanacocha, Libertad, Primero de Enero y Urco Miraflo), y de las muestras genotipificadas. En gris se muestra la comunidad de Mazán, centro poblado principal del distrito del mismo nombre, y lugar donde se referencian todos los casos febriles para realizar el diagnóstico y tratamiento respectivo. Cada punto en el gráfico representa una casa genotipificada.

Diseño de estudio

El presente estudio es del tipo analítico retrospectivo y emplea datos previamente colectados en un estudio transversal a nivel poblacional llevado a cabo en forma conjunta con la Dirección Regional de Salud de Loreto en el distrito (DRSL) de Mazán en la Amazonía Peruana. En marzo 2015, el estudio primario (SIDISI: 64371) realizó un censo y georreferenciación de viviendas en comunidades del distrito de Mazán clasificadas como de alto riesgo de malaria según DRSL. Durante el censo se

realizó la georreferenciación de las viviendas, y cada miembro del hogar fue invitado a participar del estudio. A cada participante se le asignó un código de 8 dígitos, y se procedió a la colección de datos de los individuos y las viviendas mediante el empleo de una encuesta (edad, sexo, ingreso familiar, ocupación, estudios, uso de medidas preventivas, tenencia de animales, características estructurales del hogar, entre otros).

Entre Marzo y Abril 2015 se realizó un seguimiento semanal para identificar comunidades con un incremento inusual de casos de malaria (**Figura 2**). Esta identificación se realizó en base a la proporción de individuos evaluados para la presencia de Malaria por microscopía (mayor de 20% del total de la población de la comunidad), y al índice de positividad de láminas o IPL (mayor a 5% por dos semanas consecutivas).

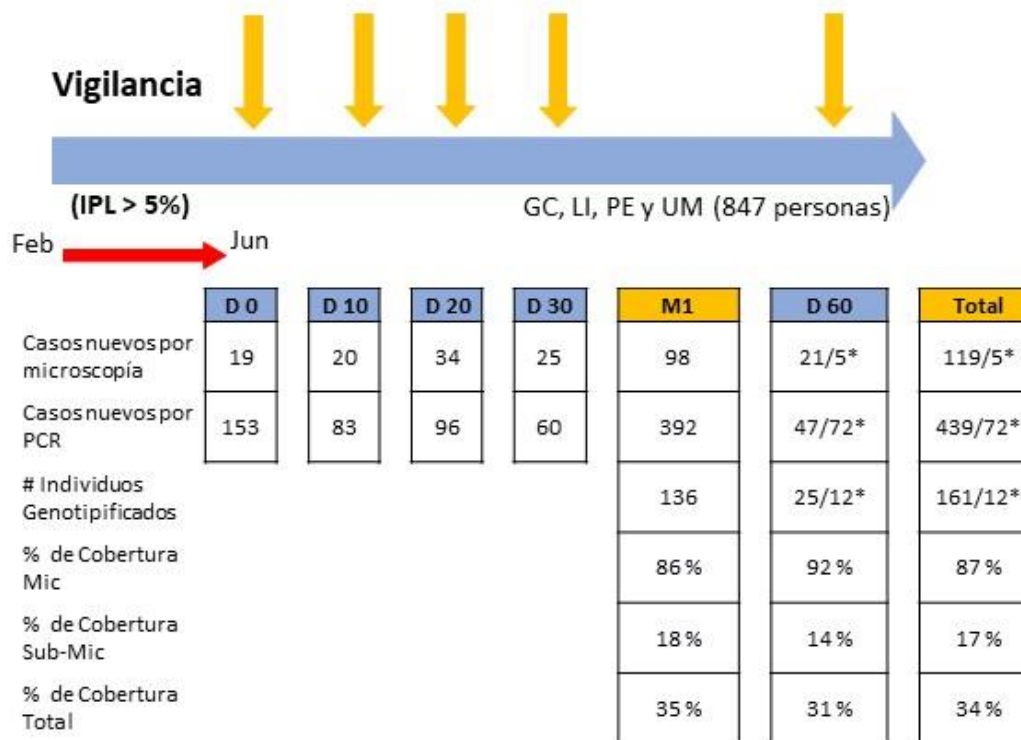


Figura 2: Diseño de estudio y colección de muestras. Casos nuevos detectados por microscopía y PCR en cada barrido; y número de individuos genotipificados en cada periodo. En el primer mes de muestreo (primeras 4 DACs), 392 individuos tuvieron al menos una muestra positiva por PCR (605 muestras positivas por PCR colectadas), y 98 de ellos fueron también positivos por microscopía. Mientras que, en la quinta DAC (después de 60 días desde la primera DAC), 119 individuos fueron PCR positivos (26 por microscopía); sin embargo, solo 47 de estos casos fueron episodios nuevos, los otros 72 fueron episodios recurrentes (muestras representadas por asterisco).

Así, se identificaron 4 comunidades (GC, LI, PE y UM) que presentaron un incremento inusual de casos de malaria en Mazán. En estas cuatro comunidades se procedió con la implementación de la estrategia de detección activa de casos (DAC) a cargo de la DRSL, la cual consistió en la ejecución de cuatro DAC consecutivas (con intervalos de 7-10 días), y una quinta al cabo de 60 días.

En cada DAC se visitó todas las casas de la comunidad, y se colectaron datos socio-demográficos del último mes, y manifestaciones clínicas de los individuos de los últimos siete días. Además, a cada participante se le realizó una evaluación clínica para determinar la presencia de fiebre u otros síntomas relacionados a malaria, y también se le realizó microscopía diagnóstica para descartar la presencia de parásitos. Solo los positivos por microscopía recibieron tratamiento, independientemente de la presencia de síntomas, de acuerdo a las normas nacionales. Adicionalmente se tomó una muestra de sangre en papel filtro Whatman-3 para el posterior diagnóstico molecular por PCR en tiempo real y genotipificación.

De esta manera cada participante del estudio otorgó hasta un máximo de 5 muestras, y así esta estrategia de muestreo permitió la recolección 724 muestra positivas a *P. vivax* por PCR, correspondientes a 439 infecciones únicas y 72 infecciones recurrentes detectadas en la quinta DAC. Trescientos noventa y dos de estas infecciones fueron identificadas en el primer mes de muestro (primeras 4 DACs), y 119 infecciones fueron detectas en la quinta DAC (47 nuevas infecciones y 72 infecciones recurrentes).

Estas 511 infecciones (439 individuos) fueron ordenadas y clasificadas en base a su concentración de ADN parasitario (ver detalle más adelante), y toda aquella muestra con una concentración por debajo de 0.8 mol/ μ L (equivalente a \sim 7 parásitos/ μ L) fue descartada por tener una concentración menor al límite de detección de los PCR utilizados para el proceso de genotipificación. Luego de realizados los PCRs, se analizó el porcentaje de amplificación de los marcadores por cada muestra, y aquellas

muestras que amplificaron para menos de 12 marcadores (amplificación de menos del 75% de los loci) también fueron descartadas del análisis (**Figura 3**).

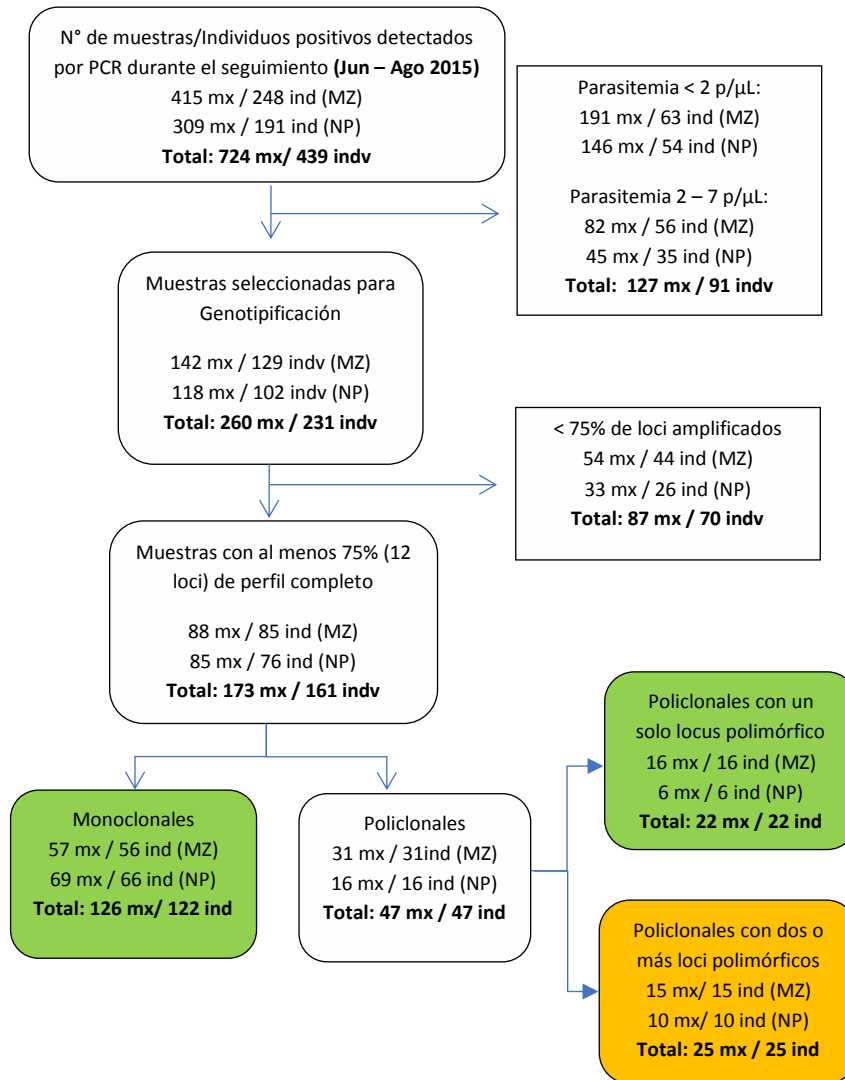


Figura 3: Criterios de inclusión y exclusión de muestras para el análisis genético poblacional. Las muestras pasaron por dos En verde se muestran las muestras empleadas en el Modelo 1 implementado en STRUCTURE, y en naranja son las muestras utilizadas en los Modelos 2 y 3.

Finalmente, las muestras fueron clasificadas en base a la multiplicidad de su infección en infecciones monoclonales e infecciones policlonales, y estas últimas en infecciones con un solo locus polimórfico (solo se observó un locus con dos o más alelos) e infecciones con dos o más loci polimórficos. En una primera etapa de los análisis genéticos, solo se emplearon las muestras monoclonales y las policlonales con un solo locus polimórfico debido a que en estas muestras se puede tener certeza de la constitución genética del haplotipo (el estado de fase o combinación de los alelos a lo largo de los cromosomas y genoma). Este conjunto de muestras se empleó para la estimación de los distintos análisis genéticos realizados, pero principalmente para definir el número de poblaciones genéticas presentes en las áreas de estudio.

Adicionalmente se utilizó otro grupo de infecciones provenientes de otras comunidades no incluidas en este estudio. Esto se realizó con dos propósitos, el primero fue hacer al modelo bayesiano de identificación de poblaciones genéticas (Modelo 1 ver más adelante) más robusto, es decir, disminuir la amplitud del intervalo de credibilidad de la probabilidad posterior del coeficiente de ancestría de cada infección. El segundo motivo fue poder identificar infecciones o subpoblaciones genéticas originadas fuera de las cuatro comunidades del distrito de Mazán inicialmente analizadas (Modelo 3, ver más adelante). Luego, los análisis genéticos se volvieron a realizar incluyendo las muestras policlonales con 2 o más loci polimórficos, y se verificó que no existieran diferencias significativas en ninguna de las estimaciones de los índices genéticos (datos no mostrados). Finalmente, se asoció la pertenencia a una subpoblación genética con las distintas variables

epidemiológicas, demográficas y clínicas colectadas durante el censo y la búsqueda activa de casos. El Flujo de trabajo completo se muestra en la **Figura 4**.

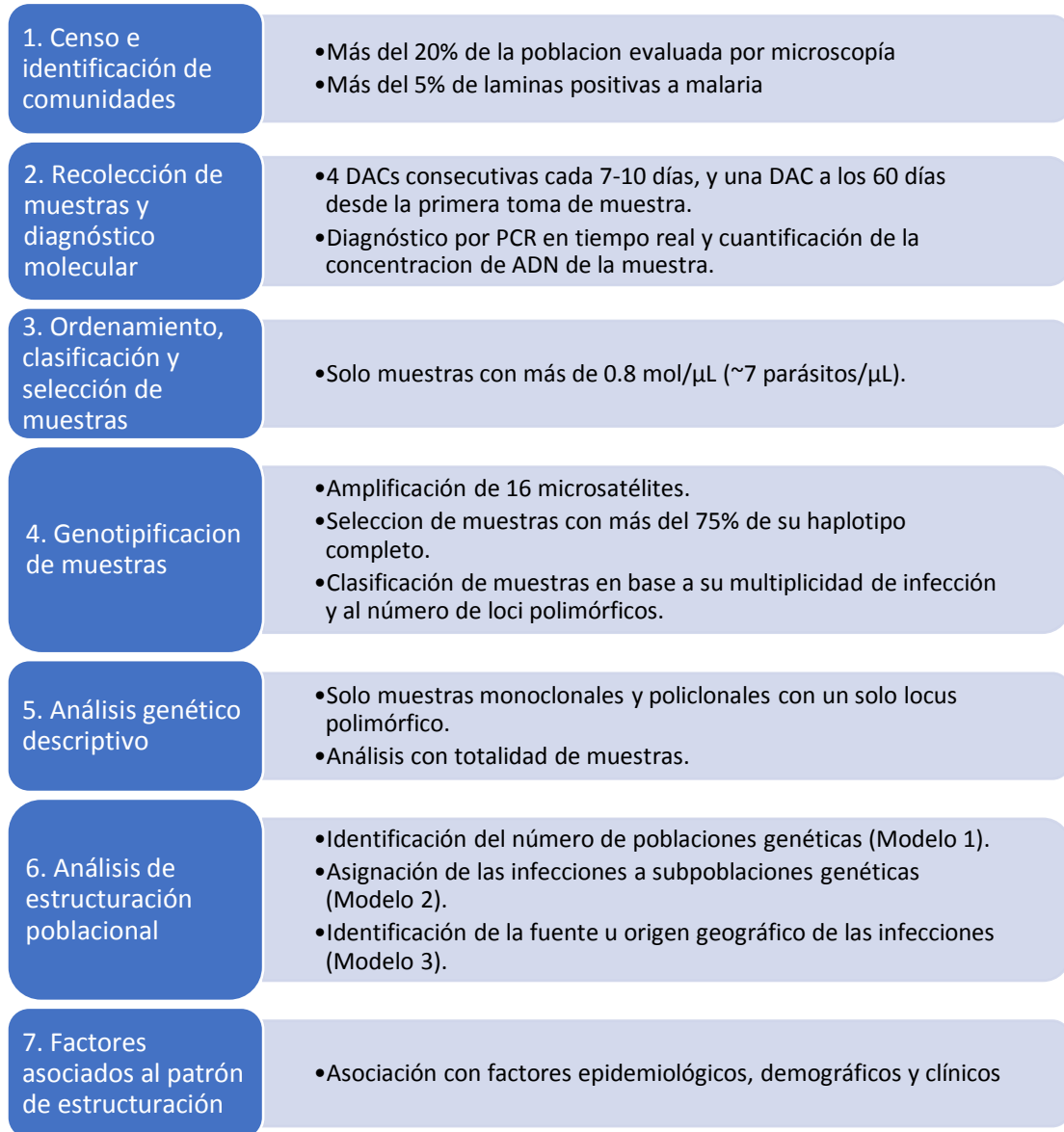


Figura 4: Flujograma de recolección, procesamiento y análisis de datos. En el esquema se muestran siete etapas empleadas para la recolección de información y muestras, selección y procesamiento de las mismas, y finalmente su análisis.

Procedimientos de laboratorio

Extracción de ADN y diagnóstico por PCR en tiempo real

La extracción de ADN de las muestras de papel filtro colectadas fue realizada mediante el uso del kit comercial E.Z.N.A.® Blood DNA Kit con ligeras modificaciones. Se cortó $\sim 6\text{mm}^2$ de papel filtro impregnado con sangre, y se colocó el papel en un tubo eppendorff conteniendo 250 μL de buffer TEN 2X (0.1 M Tris-HCl pH 8.0, 10 mM EDTA pH 8.0, 1 M NaCl), 50 μL de SDS (Dodecilsulfato sódico) al 10%, y 180 μL de H₂O ultra pura. Se incubó por una hora a 65°C con agitación constante. Luego se adicionó 25 μL de Proteasa OB provista por el kit y se incubó nuevamente a 65°C por 30 minutos adicionales. Luego el sobrenadante fue transferido a un nuevo tubo conteniendo 250 μL de Buffer BL y 250 μL de etanol absoluto. Todo el contenido de esta mezcla fue transferido a la columna de extracción HiBind DNA mini column, se realizó un lavado con 500 μL de HBC buffer, y dos lavados con 700 μL buffer de lavado. Finalmente, la elución del ADN se realizó con 50 μL de buffer de elución pre calentado a 65°C. El ADN fue almacenado a - 20 °C hasta su uso en los métodos moleculares posteriores.

El diagnóstico molecular por PCR en tiempo real se realizó siguiendo el protocolo modificado reportado Mangold, Manson (83) y utilizando los siguientes cebadores PL1473F18 [5'-TAA CgA ACg AgA TCT TAA-3'] y PL1679R18 [5'-gTT CCT CTA AgA AgC TTT-3'], dirigidos hacia el gen 18S rRNA. La reacción de PCR se realizó a un volumen final de 25 μL el cual incluyó 12.5 μL de PerfeCTa® SYBR®

Green FastMix®, 300 nM de cada primer y 5 µL de ADN. La amplificación se realizó en un termociclador Real-Time PCR CFX Connect™ Bio-Rad, y las condiciones de ciclado consistieron en una denaturación inicial a 95°C por 2 minutos, seguido por una amplificación por 40 ciclos de 20 segundos a 95°C, 20 segundos a 52°C, y 30 segundos a 68°C con adquisición continua de imagen, luego una extensión a 68°C por 3 minutos; y finalmente una curva de hibridación desde 65°C hasta 90°C con incremento de 0.5°C y adquisición de imagen cada 5 segundos.

La cuantificación absoluta de la concentración de ADN presente en cada muestra se realizó en base a una curva estándar elaborada a partir del plásmido pGEM-PV-18srRNA y 6 diluciones seriadas de factor 1:10, las cuales son expresadas como número de moléculas del plásmido por microlitro de muestra: 2×10^6 mol/µL, 2×10^5 mol/µL, 2×10^4 mol/µL, 2×10^3 mol/µL, 2×10^2 mol/µL, 2×10^1 mol/µL y 2×10^0 mol/µL. Cada punto de la curva fue evaluado por triplicado.

El PCR en tiempo real empleado para el diagnóstico en este trabajo posee un límite de detección de 0.2 genomas por reacción (0.04 mol/µL) (83), y debido a que los PCRs posteriores son menos sensibles que este PCR, las muestras fueron clasificadas de acuerdo a su concentración de ADN. Solo aquellas muestras con un número de copias mayor a 24 mol/µL fueron diluidas hasta tener 8 mol/µL en la muestra. Esta dilución fue utilizada de manera directa en la amplificación de marcadores empleados. Toda muestra que tuvo una concentración entre 1.5 y 24 mol/µL fue re-amplificada mediante el uso del kit ilustra GenomiPhi V2 DNA Amplification (el límite de detección de este método fue 4 mol/µL). Finalmente, las muestras con una

concentración por debajo de 1.5 mol/ μ L fueron amplificadas de manera directa (**Figura 3**).

Genotipificación por microsatélites

El proceso de genotipificación se realizó utilizando un panel de 16 microsatélites, 9 de ellos reportados en estudios previos realizados en la Amazonía Peruana y Brasileira (84, 85). Estos marcadores han sido seleccionados debido a que permiten individualizar cada muestra y discriminar entre muestras mono y policlonales (86). Adicionalmente, se utilizaron 7 marcadores de microsatélites en ligamiento físico en los cromosomas 2 y 14 con el objetivo de mejorar discriminación entre la identidad de estado y la identidad por descendencia. Estos nuevos marcadores fueron identificados y diseñados en base al genoma de la cepa referencial Sal1, y también han mostrado ser polimórficos y permitir determinar la multiplicidad de infección de las muestras, y las relaciones genéticas entre las infecciones primarias y los episodios de recurrencia (47). Uno de los cebadores de cada microsatélite fue marcado con un fluoróforo (6-FAM, VIC, NED, o PET) para poder analizar su producto por electroforesis capilar en el sistema ABI PRISM 3100.

La amplificación se realizó utilizando una unidad de la enzima ADN polimerasa AccuStart II, 2.5 μ L de Buffer de PCR, 1.5 mM de $MgCl_2$, 200 – 400 nM (dependiendo del cebador) de cada uno de los cebadores, y 2.5 μ L de la muestra de ADN. La reacción consistió de una denaturación inicial a 95°C por 5 minutos, seguido por 40 ciclos de 30 segundos a 94°C, 30 segundos a 55 – 67 °C (dependiendo

del cebador), y 30 segundos a 72°C; y finalmente una extensión final de 15 minutos a 72°C. En el **Anexo 1** se muestra la lista de cebadores para la amplificación de microsatélites contra *P. vivax*, y se especifica el fluoróforo, la concentración, y la temperatura de hibridación empleada para cada cebador.

Las muestras amplificadas fueron analizadas en el sistema ABI PRISM 3100 mediante electroforesis capilar, utilizando el marcador de pares de bases GeneScan™ 500 LIZ®. Para confirmar la precisión del equipo en la determinación de los alelos, 4 muestras (tres mono-infecciones, y una muestra de infección múltiple simulada) fueron genotipificadas 12 veces por el set de 16 microsatélites. Se definió la precisión de la genotipificación en base a la desviación estándar de la estimación del tamaño de cada alelo para cada locus. La desviación estándar recomendada para el equipo es de 0.5 pares de bases (87), por lo que este valor fue tomado como referencia para definir la precisión de la estimación del tamaño en pares de bases de cada alelo.

La depuración de los cromatogramas se realizó usando el programa GeneMapper® Software Version 4.0 (Applied Biosystems), y se consideró como un pico verdadero a todo aquel que presente una fluorescencia relativa mayor a 50 RFUs (Unidades relativas de fluorescencia). Todo pico secundario que fue mayor a un tercio del tamaño del pico principal y mayor a los 50 RFUs, también fue considerado como alelo; y aquellas muestras que presentaron más de un pico o alelo en un loci determinado fueron consideradas como muestras policlonales, es decir, que existe la presencia de más de una clona de parásitos en la muestra. Se consideró un umbral de

al menos 12 loci (75%) amplificados, para cada muestra, para los análisis genéticos basados en haplotipos (**Figura 3**).

Análisis genético poblacional

Los resultados de microsatélites de cada muestra fueron ingresados en una base de Excel en formato GenAlex, donde el código de cada alelo fue el tamaño en pares de bases del producto de PCR. Los análisis de diversidad haplotípica, genética y desequilibrio de ligamiento fueron realizados utilizando el paquete “*poppr*” (88, 89) desarrollado en el programa R v3.3.3, y la descripción de todos estos indicadores genéticos se realizó a nivel de la comunidad, cuenca y distrito. La diversidad haplotípica se definió en base a la riqueza de genotipos (*GML* o número de genotipos multilocus encontrados), al índice de diversidad de Shannon (*H*), y a la equidad (*E*.5, equidad en la distribución de las frecuencias de cada haplotipo). La diversidad genética fue medida en base a la Heterocigosidad esperada (H_{exp}), donde H_{exp} se define como la probabilidad de escoger dos alelos distintos al azar de un conjunto de datos (90). Para determinar si existen diferencias significativas en la diversidad genética entre los distintos grupos de comparación (por ejemplo entre cuencas, comunidades, o infecciones sintomáticas versus infecciones asintomáticas) se realizó un test de Monte Carlo con 10,000 remuestreos de haplotipos, implementado en el paquete “*adegenet*” (91) en R v3.3.3. Además. se reportó tanto la varianza total de la Heterocigosidad esperada (causada por la varianza intra e inter locus), como la varianza intra locus (varianza causada por el proceso de muestreo) (90). Como medida indirecta de endogamia se estimó el Índice de Asociación Estandarizado (I_A^S)

a partir de haplotipos únicos (para evitar una falsa endogamia debido a la propagación clonal) por cada población de interés, y se utilizó 10,000 permutaciones de alelos para determinar su significancia estadística, donde la hipótesis nula fue que el valor de I_A^S es igual a cero (74, 92). Con el fin de evitar una falsa endogamia producto al uso de marcadores ligados físicamente en un mismo cromosoma, también se determinó el Índice de Asociación Estandarizado pareado (*pairwise* I_A^S) entre cada par de loci.

El grado de diferenciación poblacional fue medido mediante el cálculo del G''_{ST} propuesto por Hedrick (93) e implementado en el paquete “*mmod*” en R v3.3.3 (94), mientras que la evaluación de la fuente de diferenciación poblacional se realizó mediante el análisis de varianza molecular (AMOVA) usando el paquete “*ade4*” (95), para lo cual se utilizó dos niveles jerárquicos (Cuenca y comunidad), y la significancia estadística de los índices de fijación ‘ ϕ ’ (phi) se calculó utilizando 10,000 permutaciones.

Estructuración poblacional e identificación de infecciones importadas e introducidas

Se empleó un modelo bayesiano implementado en el programa STRUCTURE v2.3 (80) para determinar el número de poblaciones o subpoblaciones genéticas presentes en el distrito de Mazán, y asignar cada infección a una subpoblación genética específica. Este análisis se realizó en tres etapas. En la primera etapa se determinó el número de subpoblaciones genéticas presentes en el área de estudio de interés

(**MODELO 1**). En la segunda etapa se asignaron todas las muestras analizadas a una subpoblación genética (**MODELO 2**). Y finalmente, en la tercera etapa se determinó la procedencia u origen geográfico de las infecciones (**MODELO 3**).

La determinación del número de poblaciones genéticas presentes se realizó utilizando solo las infecciones monoclonales e infecciones policlonales en un solo locus (**Figura 3**). El haplotipo de estas últimas se asignó en base al pico principal del locus polimórfico. Además, también se incluyeron 777 muestras provenientes de otras 3 cuencas (Nanay, Itaya y Tigre) para poder disminuir la amplitud del intervalo de credibilidad de la probabilidad posterior del coeficiente de ancestría de cada infección, y hacer la estimación del número de poblaciones genéticas más robusto. Estas 777 muestras también sirvieron para identificar infecciones migrantes que no procedan de alguna de las comunidades presentes en el distrito de Mazán (**MODELO 3**, ver más adelante). Para asegurar que la estructuración observada no se vea subestimada por la presencia de poblaciones altamente diferenciadas producto de aislamiento geográfico (poblaciones provenientes de distintas cuencas hidrográficas), el análisis se realizó de manera independiente para cada cuenca, así como para la población total.

Para esta primera etapa se utilizó un Modelo Ligado con Admixture (**MODELO 1**) realizado 20 veces para valores de k desde 1 a 15, con un periodo de burn-in de 50,000, seguido por 200,000 iteraciones de cadenas de Markov Monte Carlo. La distancia genética entre los loci (en centimorgans o cM) fue definida en base a la actividad de recombinación meiótica reportada en *P. falciparum* (96), para lo cual se

dividió la distancia física en pares de bases (pb) entre loci contiguos, entre 17Kpb/cM. Para obtener el número óptimo de poblaciones genéticas se realizó el análisis de la tasa de cambio de segundo orden de la probabilidad marginal ($\Delta\Delta K$) descrito por Evanno, Regnaut (97), e implementado en STRUCTURE HARVESTER (98). El coeficiente de ancestría promedio de las 20 réplicas realizadas para cada infección fue obtenido usando CLUMPP version 1.1.2.

Aquellas muestras con un coeficiente de ancestría elevado (muestras cuyo intervalo de credibilidad abarcó la unidad y no abarcó el cero), fueron utilizadas como muestras de entrenamiento para estimar el coeficiente de ancestría de las muestras policlonales en dos o más loci que no fueron incluidas en el análisis anterior. Para ello se habilitó el POPFLAG = 1 solo en las muestras de entrenamiento, y se utilizó un Modelo de Admixture (**MODELO 2**) realizado 20 veces con un periodo de burn-in de 50,000, seguido por 200,000 iteraciones de cadenas de Markov de Monte Carlo (99). El valor de k utilizado para este modelo fue el obtenido en el **MODELO 1** (en este caso 6).

Por último, se empleó la información de las 950 muestras genotipificadas en las 5 cuencas analizadas (Mazán, Napo, Nanay, Itaya y Tigre) para identificar infecciones de origen migrante dentro del distrito de Mazán. La estructuración poblacional de las poblaciones parasitarias de *P. vivax* resulta de la interacción entre sus dos hospederos (determinada por el comportamiento de ambos), y la escala geográfico – temporal en donde esta interacción ocurre (3, 16-18, 63, 77, 78). Tomando en consideración que la fecha procedencia de las muestras de las cuencas Nanay, Itaya y Tigre es distinta a

la fecha de procedencia de las muestras de las cuencas Mazán y Napo; y que, además la comunidad Urco Miraño ubicada en Napo es una población Nativa con muy poca movilización humana fuera de su territorio; se optó por definir como **Haplogrupo** a aquella sub-población genética o híbrido entre ellas, que haya mostrado expansión clonal en una cuenca, y que su frecuencia sea más de un 10% en el área geográfica de procedencia.

Esta información de grupos predefinidos (haplogrupos) fue utilizada para evaluar la presencia de migrantes o híbridos en cada una de las cuencas estudiadas. Así, nuevamente se utilizó un Modelo de Admixture (**MODELO 3**) realizado 5 veces con un periodo de burn-in de 50,000, seguido por 150,000 iteraciones de cadenas de Markov Monte Carlo (99), donde el valor de k para este modelo se determinó en base al número de haplogrupos definidos (según nuestros datos el valor de k fue 23). Además, se habilitó el uso de la ubicación de muestro (definido como el haplogrupo) para calcular la probabilidad de que la infección haya sido originada en la población predefinida (dentro de su haplogrupo preestablecido o definido en el MODELO 2), y se estableció un valor de 0.05 como probabilidad a priori de que la infección sea migrante (MIGRPRIOR = 0.05). También se calculó la probabilidad de que el evento de migración haya ocurrido en esta generación actual (GENSBACK = 0), hace una generación atrás (GENSBACK = 1), o hace dos generaciones atrás (GENSBACK = 2).

De esta manera, se definió como infección **Autóctona** a aquellas infecciones causadas por subpoblaciones genéticas de parásitos que se transmiten de manera

activa dentro del área geográfica; **híbrido**, cuando su origen es producto de la recombinación genética de dos o más poblaciones autóctonas en el área geográfica; **introducido**, haplotipos que recientemente han ingresado al área geográfica; e **importado** a infecciones causadas por subpoblaciones o haplotipos que no se transmiten de manera activa en el área geográfica donde se identificó la infección.

Análisis estadístico

Se utilizaron modelos de regresión logística para identificar factores demográficos y clínicos-epidemiológicos asociados a los haplogrupos genéticos identificados; y también para evaluar las consecuencias directas sobre la presencia de sintomatología en la población humana expuesta. La construcción de los modelos se realizó de manera independiente para cada haplogrupo y para la presencia o ausencia de síntomas en cada una de las cuencas analizadas; donde las variables dependientes fueron la pertenencia o no a un haplogrupo específico (1: sí pertenece, 0: no pertenece), y la presencia de síntomas (fiebre, cefalea, escalofríos, o sudoración). Las variables independientes fueron edad (categorizada en < 7 años, entre 7 y 12 años, entre 12 y 18 años, y > 18 años), sexo, ocupación (Estudiante, Agricultor, Otros), haber realizado actividad agrícola en el último mes, haber realizado actividad de pesca en el último mes, haber realizado actividad maderera en el último mes, realizar alguna actividad ocupacional fuera de la comunidad, convivir con un agricultor en el hogar, convivir con un pescador en el hogar, convivir con un maderero en el hogar, haber viajado fuera de la comunidad en el último mes, convivir con alguien que haya viajado en el último mes, multiplicidad de la infección, presencia de síntomas en los

últimos 7 días, la pertenencia a algún haplogrupo, la patencia de la infección (microscopia positivo no negativo) y parasitemia de la infección medida por PCR.

De esta manera se construyeron modelos bivariados para cada haplogrupo y para la sintomatología, específicos para cada cuenca. La construcción de los modelos multivariados se realizó mediante un proceso de escalonamiento progresivo (“Forward stepwise process”), y una evaluación de la razón de verosimilitud (Likelihood Ratio Test) para modelos anidados. Los factores fueron ordenados y añadidos al modelo multivariado en base al logaritmo de su verosimilitud en los modelos bivariados, y se retuvieron si tuvieron un p-valor < 0.2 en el modelo ajustado.

Debido a que este es un estudio retrospectivo, y el tamaño de muestra para cada categoría depende de la disponibilidad y calidad de la muestra de ADN obtenida en un estudio previo, solo se evaluó el poder estadístico de cada análisis realizado, mas no se realizó un cálculo de tamaño de muestra. Los análisis estadísticos de prueba de hipótesis fueron realizados en STATA 14 (StataCorp. 2015. Stata Statistical Software: Release 14. College Station, TX), y el cálculo de poder estadístico para cada análisis binomial se realizó mediante simulaciones implementadas en el paquete “*Hmisc*” en el programa R v3.3.3. La significancia estadística fue definida en un p-valor < 0.05 y los intervalos de confianza fueron reportados al 95%.

VII. RESULTADOS

El porcentaje de amplificación de los marcadores microsatélites empleados varió de manera decreciente respecto a la concentración de ADN inicial de la muestra (**Anexo 2**). Así, las muestras con una concentración inicial mayor a 8 mol/ μ L (70 parásitos/ μ L) tuvieron un porcentaje de amplificación mayor a 98%; mientras que, las muestras con concentraciones iniciales entre 1.5 – 8 (15 – 70 parásitos/ μ L), y 0.8 – 1.5 mol/ μ L (7 – 15 parásitos/ μ L) tuvieron porcentajes de amplificación por encima de 64% (promedio: 73%) y 40% (promedio: 57%) respectivamente. Ninguna muestra por debajo de la concentración de 0.8 mol/ μ L (7 parásitos/ μ L) pudo ser amplificada con éxito para más de un 40% del total de marcadores utilizados, lo cual generó que las infecciones sub-microscópicas representen solo 39% del total de muestras genotipificadas con éxito (muestras con más de 75 % de loci amplificados), y ocasiona un sesgo en la interpretación de los resultados en este grupo.

Por otro lado, la desviación estándar máxima del tamaño asignado a los alelos genotipificados fue de 0.1977 (coeficiente de variación menor a 0.05%) evidenciando una alta precisión en la estimación del tamaño de los fragmentos de PCR; y además solo 0.18% (1/560) de los controles positivos y 3.75% (9/240) de los controles negativos mostraron amplificación de alelos no correspondientes con su genotipo (en el caso de los controles negativos no debieron mostrar amplificación) durante el procesamiento de las muestras. Además, la asignación de alelos principales y secundarios en muestras policlonales fue consistente cuando el alelo principal estuvo

en una proporción de 2 a 1 (95.65%) y de 3 a 1 (100%) respecto a sus alelos secundarios. Cuando la proporción fue 1/1, en el 56% de los casos uno de los alelos se mantuvo como alelo principal.

En los 5 barridos realizados en las cuatro comunidades seleccionadas se identificaron 511 individuos infectados con *P. vivax* (392 en el primer mes, y 119 en la quinta DAC), 38 en GC, 206 en LI, 45 en PE, y 222 en UM; además, del total de individuos infectados 72 mostraron una infección recurrente en el quinto barrido (**Figura 2**). Ciento treinta y seis infecciones detectadas en los 4 primeros barridos, y 37 del quinto barrido fueron genotipificadas con un porcentaje de amplificación mayor a 75%, y sólo estas 173 infecciones fueron incluidas en los análisis posteriores, lo cual equivale a una tasa de inclusión de 34% respecto al total de infecciones detectadas por PCR, y 87% respecto solo a las infecciones detectadas por microscopía.

Complejidad de infección, diversidad genética y desequilibrio de ligamiento

Los resultados obtenidos en este trabajo muestran una distribución espacial heterogénea de la proporción de infecciones policlonales, diversidad, y endogamia a nivel micro-geográfico, con diferencias muy marcadas entre las dos cuencas analizadas. Se observaron diferencias significativas en la proporción de infecciones policlonales entre las comunidades pertenecientes a la cuenca Mazán y la comunidad de Urco Miraño ubicada en la cuenca Napo (p -valor = 0.008, **Tabla 1**, los resultados por mes se muestran en la **Anexo 3**). En las comunidades de la cuenca Mazán la

proporción de muestras policlonales fue 35.23% (IC95% 25.4 – 45.2), mientras que en Urco Miraño fue 18.82% (IC95% 10.40 - 27.24).

A nivel distrital se observó una elevada diversidad haplotípica ($H = 4.155$) y genotípica ($H_{exp} = 0.615$) de la población parasitaria, y a nivel de las dos cuencas analizadas la riqueza de genotipos y la diversidad genética fue mayor en la cuenca Mazán ($p\text{-valor} < 0.001$). En las 88 infecciones analizadas en la cuenca Mazán se identificaron 69 haplotipos únicos y la diversidad genética (H_{exp}) fue 0.707. Por otro lado, en Urco Miraño sólo se encontraron 41 haplotipos únicos y la diversidad genética fue 0.392.

En contraste, el desequilibrio de ligamiento observado fue más acentuado en Urco Miraño (I_A^S 0.21, $p\text{-valor} < 0.001$) respecto a las otras tres comunidades de la cuenca Mazán (I_A^S 0.065, $p\text{-valor} < 0.001$). El elevado nivel de desequilibrio observado se apreció tanto entre marcadores ubicados en un mismo cromosoma, como entre marcadores de distintos cromosomas (**Anexo 4**), indicando que el desequilibrio es producto de la endogamia y no por el empleo de marcadores ligados físicamente.

Tabla 1: Diversidad Haplotípica y Genética por comunidad y por cuenca en el distrito de Mazán

Cuenca	Comunidad	N	% Policlonales	GML	H	E.5	H_{exp}	σ_{He}	σ^2_{Int}	I_A^S	p-valor
Mazán	GC	7	57.14% (17.27 - 97.02)	7	1.946	1.000	0.696	0.282	0.021	0.090	0.041
	LI	63	38.10% (25.92 - 50.27)	49	3.757	0.816	0.705	0.126	0.002	0.076	0.001
	PE	18	16.67% (0.00 - 34.51)	15	2.630	0.890	0.611	0.131	0.008	0.173	0.001
Cuenca Mazán		88	35.23% (25.12 - 45.34)	69	4.110	0.833	0.707	0.128	0.001	0.065	0.001
Napo	UM	85	18.82% (10.40 - 27.24)	41	2.790	0.370	0.392	0.197	0.001	0.210	0.001
Distrito Mazán		173	27.17% (20.47 - 33.86)	110	4.155	0.370	0.615	0.126	0.001	0.086	0.001

N: Número muestras, **GML**: Número de genotipos Multilocus, **H**: Índice de Shannon, **E.5**: Equidad, **H_{exp}** : Heterocigosidad esperada, **σ_{He}** : desviación estándar de la heterocigosidad esperada, **σ^2_{Int}** : varianza intra-locus de la heterocigosidad esperada, **I_A^S** : Índice de Asociación estandarizado, **p-valor** del Índice de Asociación.

Incluso dentro de la micro-cuenca mazán, los distintos indicadores genéticos analizados presentaron variación a nivel espacial. En la comunidad de Libertad el porcentaje de infecciones policlonales fue 38.1% (25.92 - 50.27) y la diversidad observada fue 0.705. En cambio, la comunidad de Primero de Enero, la cual se ubica contigua a Libertad a una distancia no mayor de 2 km, presentó valores más bajos de infecciones policlonales 16.67% (0.00 - 34.51) y menor diversidad genética ($H_{exp} = 0.611$). Y si bien, no se observaron diferencias significativas en la proporción de infecciones policlonales (debido al bajo poder estadístico y bajo tamaño de muestra), las diferencias en la diversidad genética si fueron significativas ($p\text{-valor} < 0.005$).

Diferenciación genética, estructuración poblacional e identificación del origen de las infecciones

Se observaron valores elevados de diferenciación poblacional entre las cuatro comunidades, siendo las más altas las observadas entre Urco Miraño y las tres

comunidades pertenecientes a la cuenca Mazán ($G''_{ST} = 0.43$), y la observada entre Gamitanacocha y Primero de Enero ($G''_{ST} = 0.51$). El análisis de varianza molecular (AMOVA) fue realizado utilizando la cuenca hidrográfica y la comunidad de procedencia de cada infección como causantes de la diferenciación y variación genética entre las infecciones. Este análisis mostró que la mayor variación se encuentra dentro cada comunidad (79.43%), seguido por la variación entre Cuencas (12.26%) y la variación entre Comunidades dentro de cada Cuenca (8.3%). Los valores de Φ_{ST} , Φ_{CT} y Φ_{SC} estimados fueron de 0.12, 0.09, y 0.21, respectivamente, mostrando que incluso dentro de cada comunidad hay una alta presencia de estructuración poblacional.

El análisis de estructuración poblacional y la asignación de cada infección a una población genética ancestral, se realizó mediante el uso de un método bayesiano. Este método fue implementado en el programa STRUCTURE usando un modelo de ligamiento genético (**MODELO 1**) solo con aquellas muestras que fueron monoclonales y con las policlonales en un solo marcador (**Figura 3**). Así, el análisis de la tasa de cambio de segundo orden de la probabilidad marginal (ΔK) mostró que en el set de datos del distrito de Mazán los números posibles de poblaciones ancestrales presentes son $k = 3$ y $k = 7$. Al agregar los sets de datos de las áreas de estudio de San José de Lupuna, Cahuide, y Santa Emilia (cuencas Nanay, Itaya y Tigre respectivamente), los números posibles de poblaciones ancestrales presentes en las cuatro áreas de estudio (5 cuencas) fueron $k = 6$, $k = 8$ y $k = 10$ (ver **Anexo 5**). Pese a los distintos números de poblaciones probables encontradas, se observó una

alta consistencia en la asignación de las infecciones a una población ancestral al comparar el coeficiente de ancestría obtenido a partir de 6, 8 y 10 poblaciones ancestrales (**Anexo 6**). A su vez, se observó consistencia entre las poblaciones genéticas encontradas y la expansión clonal observada mediante una red de expansión mínima de haplotipos (**Figura 5** y **Anexo 7**), lo cual sugiere que los distintos niveles de estructuración genética son causados por eventos de expansión clonal de las poblaciones genéticas ancestrales de parásitos dentro de cada cuenca, o expansión clonal de los híbridos entre ellas. Por este motivo, se asumió que el número de poblaciones ancestrales presentes en las 4 áreas (5 cuencas) de estudio es 6, y fue este número el utilizado para estimar el coeficiente de ancestría en las infecciones policlonales restantes (**MODELO 2, Anexo 8**), y también en la descripción de la diversidad dentro de cada población genética.

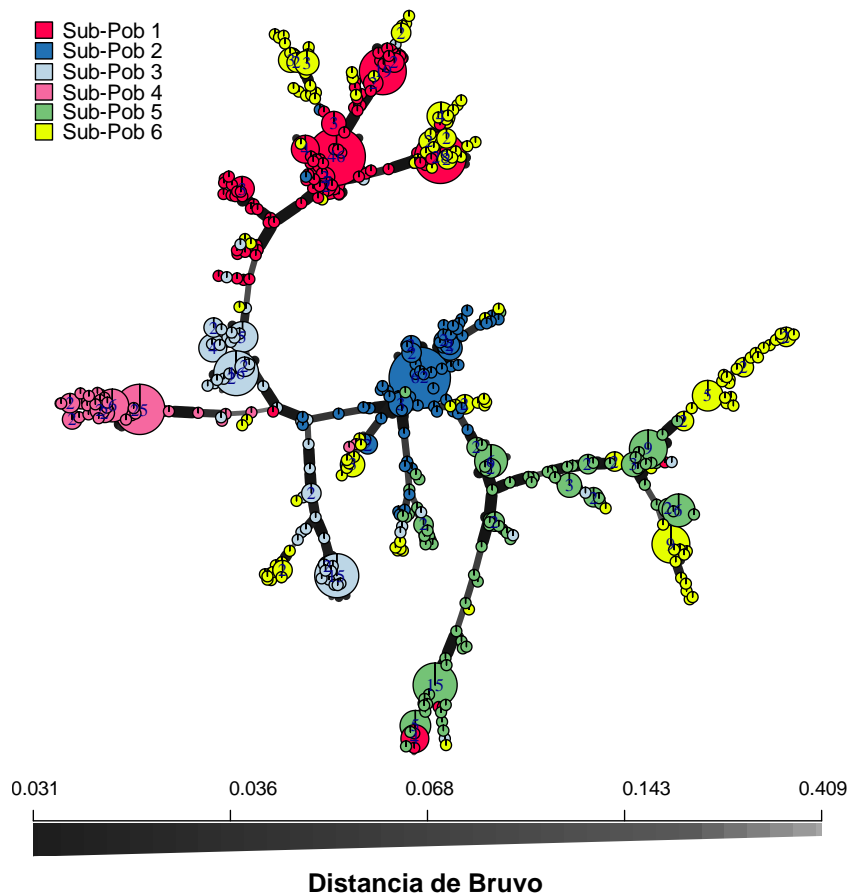


Figura 5: Red mínima de expansión de haplotipos del total de infecciones genotipificadas en las 4 áreas de estudio (Mazán, Lupuna, Cahuide y Santa Emilia). Los colores han sido asignados de acuerdo a la pertenencia de cada infección a una Sub-Población Genética asumiendo la presencia de 6 Sub-Poblaciones. Cada círculo representa un haplotipo, y su tamaño guarda relación directa con el número de muestras que presentan ese haplotipo. Las relaciones genéticas entre los haplotipos están representadas por líneas que conectan cada círculo, y su distancia genética se muestra en la barra inferior del gráfico.

La diversidad genética en 5 de las 6 subpoblaciones fue menor a 0.5 ($H_{exp} = 0.182 - 0.405$); sin embargo, la diversidad de la subpoblación 6 fue 0.724. Además, la distribución de los haplotipos de la subpoblación 6 en la red de expansión mínima fue aleatoria y no se observó una agrupación clara en alguna rama de la red (**Figura 5**).

Esto podría indicar que este grupo no sea una población genética real, sino que es una agrupación de haplotipos que por su falta de parentesco con las otras 5 poblaciones fueron asignadas a un sexto grupo. Por otro lado, se observó una amplia distribución geográfica de cada población genética ancestral, y el grado de mestizaje y diversificación de estas poblaciones varió en las áreas geográficas analizadas (**Anexo 8 y Anexo 9**). Por ello se optó por definir como **Haplogrupo** a aquella población genética o híbrido presente en una cuenca con más de 10% de frecuencia (**Anexo 10**).

Es así que en la cuenca Mazán se encontraron 5 haplogrupos originados por mestizaje entre las 6 poblaciones genéticas ancestrales originales. El haplogrupo 1 está conformada principalmente por infecciones híbridas de la población genética 2 con la población 5, y tuvo una diversidad de 0.419. El haplogrupo 2 está conformado por híbridos de la población genética 3 con las poblaciones 1 y 5, y su diversidad fue 0.378. El haplogrupo 3 es un híbrido de la población 4 con la población 5, con diversidad de 0.385. El haplogrupo 4 se conforma por híbridos entre las poblaciones 5 y 6, y su diversidad fue 0.325; y finalmente la haplogrupo 5 está conformada infecciones híbridas entre la población genética 6 principalmente y el resto de poblaciones genéticas con una diversidad de 0.668.

En contraste, en Napo se observó la presencia de solo dos haplogrupos (haplogrupos 6 y 7). El haplogrupo 6 es un híbrido entre las poblaciones genéticas 1 y 6 con una diversidad de 0.12, mientras que el haplogrupo 7 corresponde a infecciones cuyo origen proviene de la población genética 3 y su diversidad fue 0.092. Ambos

haplogrupos estuvieron presentes en las dos fechas de muestreo; sin embargo, el haplogrupo 7 disminuyó su frecuencia notoriamente en el segundo mes (**Figura 6**).

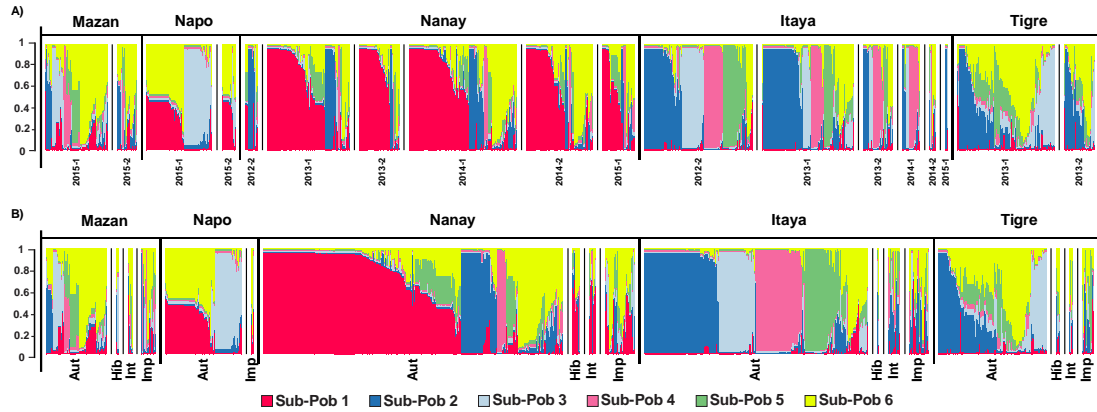


Figura 6: Estructuración poblacional de *P. vivax* en cuatro cuencas de la Amazonía Peruana. Cada color corresponde a una población ancestral, y cada barra indica la proporción del genotipo de un individuo que proviene de cada población ancestral (Coeficiente de Ancestría o Q). El coeficiente de ancestría (Q) en ambos gráficos (A y B) fue calculado mediante el MODELO 2. A) Distribución temporal de las sub-poblaciones en cada cuenca. B) Identificación de infecciones Autóctonas, Híbridas, Introducidas e Importadas en cada cuenca.

Otros 16 haplogrupos fueron identificados en las otras cuencas analizadas y, en conjunto con los 7 haplogrupos anteriores, se emplearon como base para la identificación infecciones de origen **autóctono, híbrido, introducido e importado** (**Figura 6**). De esta manera se estimó la probabilidad de que la infección se haya originado en el haplogrupo al cual fue asignado en el modelo 2 o $P(Hap_{Modelo2})$, y también se estimó la probabilidad de que la infección se haya originado en un haplogrupo distinto (GB_0), la probabilidad de que su ancestro inmediato provenga de un haplogrupo distinto (GB_1), y la probabilidad de que el ancestro de su ancestro provenga de un haplogrupo distinto (GB_2).

Veintitrés infecciones pertenecientes al distrito Mazán mostraron una probabilidad menor a 0.75 de pertenecer al haplogrupo al cual habían sido asignados en el Modelo 2 (**Tabla 2**). De hecho, 18 de estas infecciones su probabilidad fue por debajo de 0.07, indicando que ni un solo alelo de su genotipo pertenece al haplogrupo asignado previamente. En cambio, en las otras 5 infecciones la probabilidad estuvo entre 0.118 y 0.513 indicando que al menos uno de sus alelos proviene del haplogrupo asignado en el modelo 2, lo cual sugiere que el genotipo migrante se recombinó con algún genotipo autóctono (infecciones introducidas). Ninguna de las 18 infecciones de origen completamente foráneo tuvo su origen inmediato (GB_0) en otro haplogrupo, sino más bien su origen provino de hace una (GB_1) o dos (GB_2) generaciones atrás. Pese a que es posible que estas infecciones si hayan sido introducidas, pero aún no haya habido tiempo suficiente para que ocurra recombinación con algún genotipo autóctono, se optó por definir a estas infecciones como importadas.

Tabla 2: Probabilidad de procedencia a una población.

Comunidad	Hap _{Modelo2}	P(Hap _{Modelo2})	Hap _{Modelo3}	GB ₀	GB ₁	GB ₂
Gamitanacocha	5	0.001	12	0	0	0.991
	1	0.047	17	0	0	0.486
	3	0	16	0.32	0.641	0.01
Libertad	4	0	21	0.004	0.008	0.988
	5	0.024	9	0	0.001	0.974
	5	0.05	9	0	0	0.947
	3	0.002	21	0	0	0.862
	5	0.028	18	0.058	0.117	0.74
	3	0.052	7	0	0	0.453
	1	0.014	14	0.222	0.443	0.031
	5	0	13	0.327	0.655	0.011
	5	0	13	0.33	0.661	0.009
	5	0	13	0.332	0.663	0.005
	5	0.118	9	0.011	0.021	0.437
	5	0.15	23	0.107	0.214	0.265
	1	0.513	13	0.067	0.135	0.234
Primero de Enero	2	0.01	13	0	0	0.885
	3	0	11	0.315	0.63	0.024
	5	0.284	17	0	0	0.651
	1	0.161	8	0	0	0.561
Urco Mirañó	7	0.067	8	0	0	0.734
	6	0.018	5	0.004	0.008	0.536
	6	0	5	0.034	0.068	0.533

Hap_{Modelo2}: Haplogrupo asignado en el Modelo 2. *P(Hap_{Modelo2})*: probabilidad posterior de haber sido asignado correctamente en el modelo 2. *Hap_{Modelo3}*: Haplogrupo asignado en el Modelo 3. *GB₀*: probabilidad de que la infección se haya originado en el haplogrupo asignado por el modelo 3. *GB₁*: probabilidad de que su ancestro inmediato provenga del haplogrupo asignado por el modelo 3. *GB₂*: probabilidad de que el ancestro de su ancestro provenga del haplogrupo asignado por el modelo 3. Los haplogrupos del 1 al 5 pertenecen a la cuenca Mazan, el 6 y 7 a la cuenca Napo, del 8 al 13 a la cuenca Nanay (Lupuna), del 14 al 18 a la cuenca Itaya (Cahuide), y del 19 al 23 a la cuenca Tigre (Santa Emilia).

Así, en la cuenca Mazán, 20 (22.73% IC95% 13.9 – 31.5%) de las infecciones genotipificadas fueron infecciones foráneas de origen introducido 5 (5.68 %) o importado 15 (17.05 %, ver **Tabla 3**), y 10 de estas 20 infecciones habían sido

previamente asignadas al haplogrupo 5 (perteneciente a la subpoblación genética 6, ver **Tabla 2**). Además, 12 de las 20 infecciones foráneas tuvieron su origen en el área de estudio de San José de Lupuna (Cuenca Nanay), la cual está geográficamente más cercana a las comunidades del distrito de Mazán que las otras dos áreas de estudio, Cahuide y Santa Emilia (Cuenca Itaya y Tigre respectivamente, ver **Anexo 9**). Dentro de la cuenca Mazán, la comunidad de Libertad presentó 13 (20.64 % IC95% 10.49 - 30.78%) infecciones de origen foráneo, introducidas 3 (4.76%) o importadas 10 (15.87 %); mientras que las comunidades de Primero de enero y Gamitanacocha presentaron 4 (2 introducidas y 2 importadas) y 3 (todas importadas) infecciones foráneas respectivamente,

En cambio, en Urco Miraño solo 3 infecciones (3.5% IC95% 0 – 7.5%) fueron de origen importado, y dos de estas infecciones tuvieron su origen en la cuenca Mazan, y la tercera tuvo su origen en San José de Lupuna (cuenca Nanay). En las otras cuencas, 11.58 % (9.33 – 13.84%) fueron de origen importado o introducido (**Anexo 11**).

Tabla 3: Fuente u origen de las infecciones en el Distrito de Mazán

Cuenca	Mazán		Napo	
	# Indv	Frec.% (IC95%)	# Indv	Frec.% (IC95%)
Autóctono	65	73.86 (64.62 – 83.11)	82	96.47 (92.52 – 100.00)
Híbrido	3	3.41 (0.00 – 7.23)	-	-
Introducido	5	5.68 (0.81 – 10.55)	-	-
Importado	15	17.05 (9.13 – 24.96)	3	3.53 (0.00 – 7.48)
Total	88		85	

Por otro lado, la distribución geográfica de los 7 haplogrupos identificados en las dos cuencas del distrito de Mazán no fue homogénea a lo largo de sus comunidades (**Figura 7**); y es así que, los haplogrupos 1, 3 y 4 por ejemplo, solo estuvieron presentes en la comunidad de LI, con la excepción de 4 casos importados encontrados en GC y PE (**Anexo 12 - 14**). El haplogrupo 2 se ubicó en la zona limítrofe entre LI y PE (**Anexo 17**); mientras que el haplogrupo 5 estuvo a lo largo de toda la cuenca Mazán (**Anexo 15**). Por otro lado, como se mencionó anteriormente, los haplogrupos 6 y 7 fueron específicos de UM (**Anexo 16 y 17**). En Primero de enero 4 de las 18 infecciones fueron de origen foráneo (dos introducidas y 2 importadas), mientras que en Gamitanacocha 3 de las 7 infecciones fueron importadas.

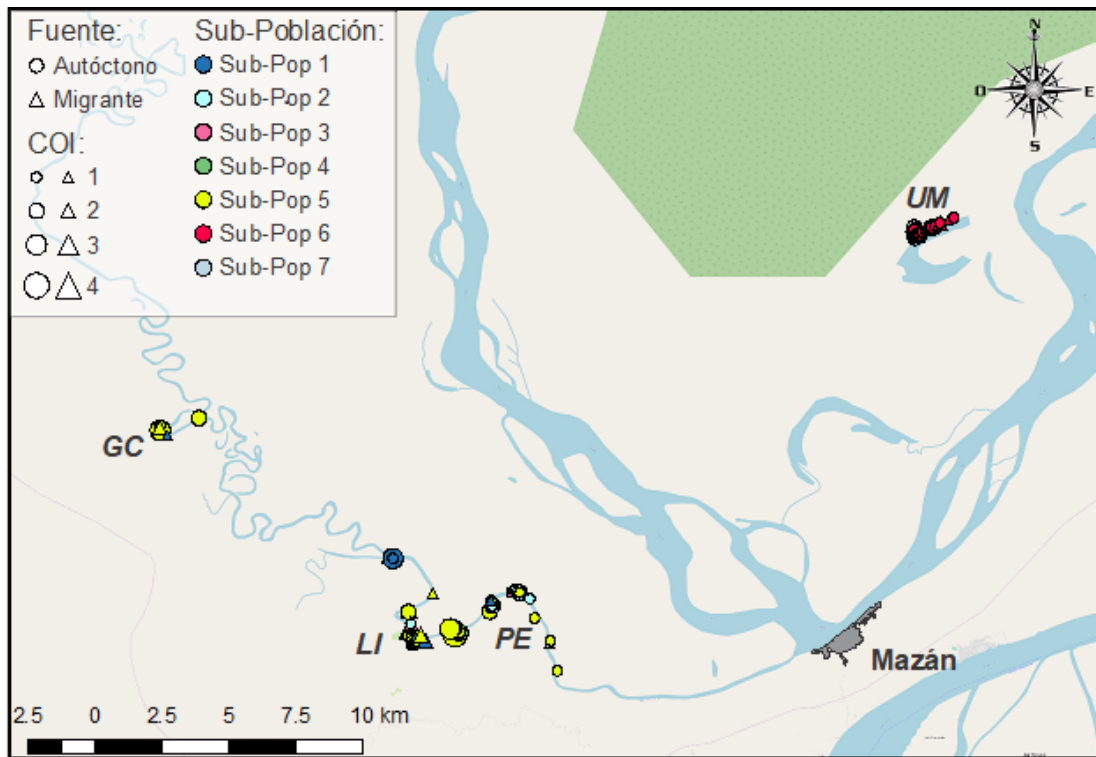


Figura 7: Distribución geográfica de los haplogrupos, casos autóctonos, e infecciones policlonales en las cuatro comunidades estudiadas: Gamitanacocha (GC), Libertad (LI), Primero de Enero (PE) y Urco Mirañó (UM). Las infecciones autóctonas causadas por cada haplogrupo están representadas por puntos, mientras que las infecciones foráneas se representan con triángulos; y su color y tamaño depende del haplogrupo y de la complejidad de la infección respectivamente. La distribución espacial de cada haplogrupo se muestra de manera individual en los Anexos 12 - 17.

Por otro lado, doce de las 37 infecciones genotipificadas en el segundo barrido realizado en este estudio (60 días después de la primera toma de muestra) fueron infecciones recurrentes; 4 de ellas en Libertad, y 8 en Urco Mirañó. Las 4 infecciones en libertad fueron recurrencias heterólogas respecto a la primera infección (el haplotipo de la primera infección pertenece a un haplogrupo distinto que el haplotipo de la infección recurrente); mientras que 6 de las ocho recurrencias en Urco Mirañó

fueron homólogas (haplotipos similares o que pertenecen a un mismo haplogrupo). A pesar de ello, el diseño de estudio de este trabajo no permite diferenciar entre nuevas infecciones y episodios de relapsos, por lo que no es posible poder evaluar de manera directa y objetiva la contribución de estos últimos en la diversidad y patrones de estructuración observados.

Factores demográficos y clínicos-epidemiológicos asociados a los haplogrupos genéticos

Los modelos de regresión logística bivariados mostraron que haber viajado en el último mes está asociado e incrementa en 4.33 veces (p -valor = 0.044, IC95% 1.04 – 18.10, poder = 0.824) el riesgo de tener una infección causada por el haplogrupo 2 en la cuenca Mazán (**Tabla 4** y **Anexo 19**). Otros factores como la edad (OR = 4.04, p -valor = 0.124, poder = 0.484), ser varón (OR = 3.28, p -valor = 0.062, poder = 0.518), convivir con una persona que realice alguna ocupación fuera de la comunidad (OR = 2.61, p -valor = 0.132, poder = 0.412), convivir con alguien que haya realizado actividad maderera en el último mes (OR = 2.87, p -valor = 0.1, poder = 0.434), convivir con alguien que haya salido de la comunidad en el último mes (OR = 3.35, p -valor = 0.052, poder = 0.597), y tener presencia de sintomatología (OR = 2.78, p -valor = 0.086, poder = 0.427) mostraron una tendencia a incrementar el riesgo de tener una infección causada por el haplogrupo 2; sin embargo, esta tendencia no fue estadísticamente significativa por la falta de poder estadístico (poder < 0.8).

En esta misma cuenca, otros haplogrupos también mostraron tendencia a incrementar su probabilidad de ser la causa de la infección ante la presencia de síntomas (haplogrupo 1, OR = 3.65, p-valor = 0.055, poder = 0.524, **Anexo 18**), ser infecciones monoclonales (haplogrupo 1, OR = 6.25, p-valor = 0.085, poder = 0.659, **Anexo 18**), haber realizado pesca en el último mes (haplogrupo 3, OR = 6.42, p-valor = 0.141, poder = 0.808, **Anexo 20**), o ser infecciones asintomáticas (haplogrupo 4, Prueba exacta de Fisher p-valor = 0.025); sin embargo, la asociación de algunas de ellas no fue estadísticamente significativa debido al bajo tamaño de muestra y la falta de poder estadístico (Todos los modelos se describen en las **Anexo 18 – 25**).

Tabla 4: Modelos bivariados y Multivariado de regresión logística para la presencia del Haplogrupo 2 en la cuenca Mazán.

Factor de exposición	Modelo Bivariado				
	OR	p-valor	Poder	[IC 95%]	
Edad (Ref = < 7 años)					
7 - <12	4.04	0.124	0.484	0.68	23.94
Sexo (Ref = Femenino)					
Masculino	3.28	0.062	0.518	0.94	11.42
Ocupación fuera de la comunidad (Ref = No)					
Sí	2.61	0.132	0.412	0.75	9.07
Maderero en el hogar (Ref = No)					
Sí	2.87	0.1	0.434	0.82	10.08
Viaje en el último mes (Ref = No)					
Sí	4.33	0.044	0.824	1.04	18.1
Viaje de algún miembro de la casa (Ref = No)					
Sí	3.35	0.052	0.597	0.99	11.31
Sintomatología (Ref = No)					
Sí	2.78	0.086	0.427	0.86	8.96

Razón de probabilidades (OR) de los modelos bivariados. Solo se muestran lo OR de aquellas variables que poseen un p-valor menor a 0.2.

Por otro lado, en la cuenca Napo, tener una infección causada por el haplogrupo 7 es 3.96 veces más probable (p-valor = 0.048, IC95% 1.01 – 15.52, poder = 0.509) en el grupo etario entre 12 – 18 años respecto a los niños menores de 7 años (**Anexo 25**).

Las infecciones foráneas presentes en la cuenca Mazán no se encontraron asociadas a ninguno de los factores mencionados anteriormente, pero si se observó una tendencia a incrementar su presencia en personas menores de 7 años (Referencia = 7-12 años, OR = 7.69, p-valor = 0.074, poder = 0.557), y mujeres (OR = 3.03, p-valor = 0.082, poder = 0.462) (**Anexo 23**). Por último, ninguno de los haplogrupos identificados se encontró asociado a algún factor en los modelos multivariados construidos, principalmente por el bajo tamaño de muestra que se genera al subcategorizar los datos.

Asociación de haplogrupos genéticos, factores demográficos y clínicos-epidemiológicos con la presencia de síntomas

Los factores genéticos (haplogrupos), así como los demográficos y clínico-epidemiológicos anteriormente mencionados, pueden tener consecuencias directas sobre la presencia de sintomatología en la población humana expuesta, y su efecto o magnitud puede variar a nivel geográfico (entre cuencas o incluso comunidades). Los modelos de regresión logística bivariados y el modelo multivariado a nivel distrital mostraron que tener una infección patente o microscópica (OR = 2.35, p-valor = 0.01, poder = 0.597) y convivir con una persona que haya realizado un viaje fuera de la

comunidad en el último mes (OR = 2.74, p-valor = 0.01, poder = 0.757) incrementa la probabilidad de presentar sintomatología (**Tabla 5** y **Anexo 26**).

Tabla 5: Modelos bivariados y multivariado de regresión logística para la presencia de Sintomatología en el distrito de Mazán.

Factor de exposición	Modelo Bivariado				Modelo Multivariado			
	OR	p-valor	Poder	[IC 95%]	OR	p-valor	[IC 95%]	
Patencia (Ref = Sub-Patente)	2.01	0.04	0.597	1.05 3.86	2.35	0.01	1.19 4.66	
Clonalidad de la infección (Ref = Monoclonal)	0.60	0.17	0.355	0.29 1.25				
Haplogrupo 1 (Ref = No)	2.86	0.11	0.932	0.80 10.18				
Haplogrupo 2 (Ref = No)	2.17	0.17	0.943	0.72 6.56				
Ocupación fuera de la comunidad (Ref = No)	1.80	0.19	0.494	0.75 4.37				
Maderero en el hogar (Ref = No)	2.11	0.16	0.656	0.74 5.96				
Viaje en el último mes (Ref = No)	2.26	0.18	0.75	0.69 7.45				
Viaje de algún miembro de la casa (Ref = No)	2.29	0.03	0.757	1.08 4.83	2.74	0.01	1.25 6.00	

En los modelos bivariados y en el modelo multivariado se muestran todas aquellas variables con un p-valor menor a 0.2.

A nivel de la cuenca Mazán (**Tabla 6** y **Anexo 27**), la parasitemia de la infección (OR = 1.10, p-valor = 0.04), estar infectado por un parásito perteneciente al haplogrupo 1 (OR = 4.83, p-valor = 0.03), y convivir con alguien que haya salido de la comunidad en el último mes (OR = 4.48, p-valor = 0.01) incrementaron la probabilidad de presentar síntomas en el modelo multivariado. Estar infectado con un parásito perteneciente al haplogrupo 2 también incrementa la probabilidad de presentar síntomas (OR = 3.75, p-valor = 0.04, OR obtenido al sacar la convivencia con personas que hayan salido de la comunidad del modelo multivariado); sin embargo, esta variable se encontró asociada con convivir con alguien que haya salido de la comunidad en el último mes y por ello fue excluida del modelo multivariado.

Finalmente, en la cuenca Napo (**Anexo 28**), solo las infecciones patentes incrementaron el riesgo de presentar síntomas (OR = 5.00, p-valor = 0.002).

Tabla 6: Modelos bivariados y multivariado de regresión logística para la presencia de Sintomatología en la cuenca Mazán.

Factor de exposición	Modelo Bivariado				Modelo Multivariado				
	OR	p-valor	Poder	[IC 95%]	OR	p-valor	[IC 95%]		
Parasitemia (1/1000)	1.07	0.13	0.424	0.98	1.16	1.10	0.04	1.00	1.20
Clonalidad de la infección (Ref = Monoclonal)	0.42	0.09	0.442	0.16	1.16				
Complejidad de infección (Ref = 1)									
2	0.42	0.11	0.433	0.15	1.21				
Haplogrupo 1 (Ref = No)	3.65	0.06	0.855	0.97	13.67	4.83	0.03	1.19	19.61
Haplogrupo 2 (Ref = No)	2.78	0.09	0.653	0.86	8.96				
Ocupación fuera de la comunidad (Ref = No)	2.35	0.12	0.506	0.80	6.92				
Maderero en el hogar (Ref = No)	2.75	0.07	0.672	0.91	8.33				
Viaje de algún miembro de la casa (Ref = No)	3.16	0.03	0.768	1.11	9.05	4.48	0.01	1.45	13.78

En los modelos bivariados y en el modelo multivariado se muestran todas aquellas variables con un p-valor menor a 0.2.

Además, se observó que las infecciones sub-microscópicas ($H_{exp} = 0.649$) poseen una mayor diversidad genética que las infecciones microscópicas ($H_{exp} = 0.584$), siendo esta diferencia estadísticamente significativa (p-valor = 0.043). Las infecciones asintomáticas ($H_{exp} = 0.622$) también mostraron mayor diversidad genética en comparación a las infecciones sintomáticas ($H_{exp} = 0.591$); sin embargo, no se observaron diferencias significativas (p-valor = 0.092). Esta tendencia fue más notoria en el haplogrupo 6 presente en Urco Miraño, donde las infecciones sub-microscópicas mostraron una diversidad genética de 0.2762, mientras que las

infecciones microscópicas tuvieron una diversidad de 0.0401, con diferencias significativas entre ambos grupos (p -valor = 0.001). De igual manera, las infecciones asintomáticas infectadas con el haplogrupo 6 ($H_{exp} = 0.161$) también mostraron mayor diversidad genética en comparación a las infecciones sintomáticas ($H_{exp} = 0.060$) causada por el mismo grupo genético (p -valor = 0.034). Al representar las relaciones genéticas de los haplotipos del haplogrupo 6 en una red de expansión mínima (**Figura 8**), se observó que tanto las infecciones sub-microscópicas como las asintomáticas poseen una mayor dispersión a lo largo de la red que las infecciones microscópicas y sintomáticas. De hecho, las infecciones sintomáticas están representadas preferentemente por aquellos haplotipos más frecuentes en el grupo de los asintomáticos (tanto sub-microscópicos como microscópicos). Esto podría sugerir que la ausencia de control sobre este grupo de infecciones (asintomáticas y/o sub-microscópicas) permite el incremento y diversificación poblacional del parásito, y además sirven como reservorio y fuente a la población humana no expuesta o susceptible.

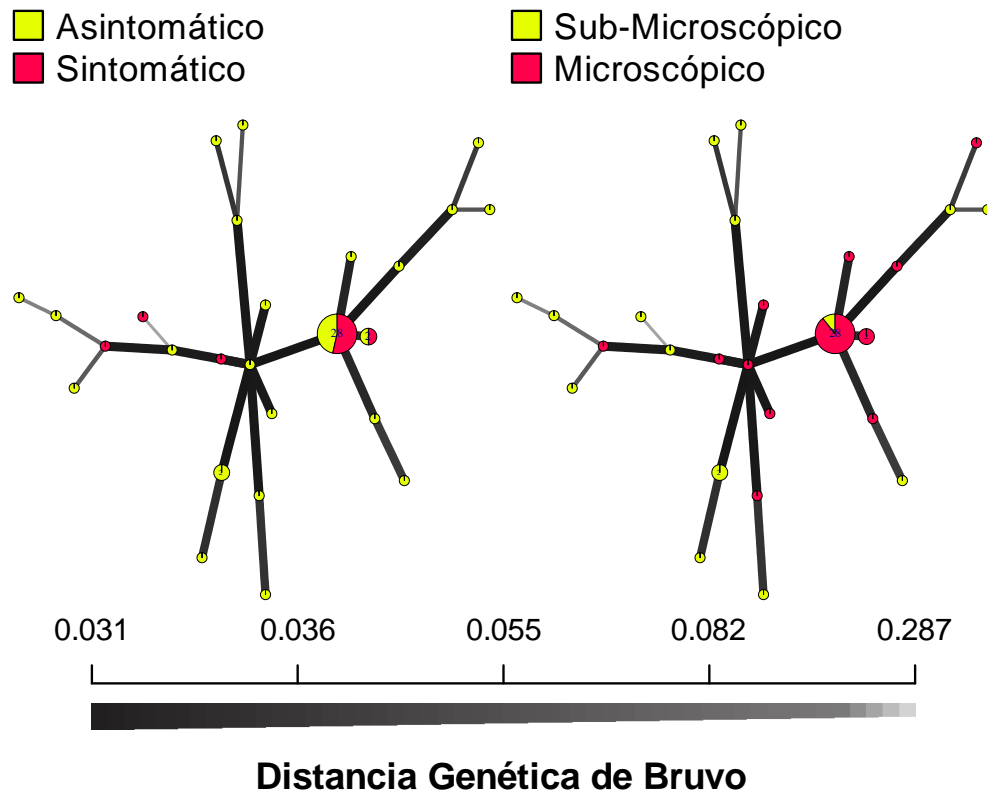


Figura 8: Red mínima de expansión de haplotipos del haplogrupo 6 presente en UM. El gráfico representa las relaciones genéticas existentes entre las infecciones sintomáticas y asintomáticas (gráfico de la izquierda), y las infecciones microscópicas y sub-microscópicas (derecha).

VIII. DISCUSIONES

La intensidad de transmisión de Malaria tiene efectos sobre la población parasitaria, los cuales pueden evidenciarse en la proporción de infecciones policlonales, en la diversidad de las poblaciones parasitarias, y en los niveles de auto-fecundación, propagación clonal, y patrones estructuración observados (3, 16-18). Por ello en el presente estudio se describió la distribución espacial a nivel micro-geográfico de distintas métricas genético poblacionales en tres comunidades de la cuenca del río Mazán, y de la comunidad de Urco Miraño en la cuenca Napo, todas estas comunidades ubicadas dentro de la jurisdicción del Distrito de Mazán.

Los resultados obtenidos en este trabajo muestran una distribución espacial heterogénea de la diversidad, expansión clonal y estructuración poblacional a nivel micro-geográfico, con diferencias muy marcadas entre las dos cuencas analizadas. Así por ejemplo, la comunidad de Urco Miraño presentó niveles bajos de diversidad genética ($H_{exp} = 0.392$), incluso ligeramente menores a los reportados previamente en la comunidad de San Carlos por Delgado-Ratto, Soto-Calle (20) y Van den Eede, Van der Auwera (22), en donde se encontró valores de diversidad de 0.49 y 0.44 respectivamente. San Carlos es una comunidad situada a 10 km de la carretera principal Iquitos-Nauta, y a 21 km al sureste de la ciudad de Iquitos, a seis horas en barco por el río Itaya. En esta comunidad también se reportó niveles bajos de infecciones policlonales (11%) y un fuerte desequilibrio de ligamiento ($IAS = 0.15$, $p\text{-valor} = 0.001$). Al igual que San Carlos, Urco Miraño es una comunidad nativa con

poco o casi escaso desplazamiento humano fuera de la comunidad; lo cual fomenta el aislamiento reproductivo del parásito, incrementa el efecto de la deriva génica, y con ello se reduce la diversidad genética, la proporción de infecciones policlonales y se incrementa la endogamia (3).

Un escenario muy distinto es el que se observa en las comunidades de la cuenca Mazan, en donde estas no solo mostraron niveles elevados de diversidad e infecciones policlonales; si no también, un bajo nivel endogamia y un alto número de poblaciones genéticas que coexisten incluso dentro de una misma comunidad. Un ejemplo de ello es la comunidad de Libertad, donde el porcentaje de infecciones policlonales fue 38.1% y la diversidad observada fue 0.705. Estos resultados son consistentes con los niveles de diversidad ($Hex = 0.66$) e infecciones policlonales (44%) reportadas en este mismo distrito con muestras colectadas en el año 2006 (22). Esta ausencia de disminución en la diversidad genética en esta zona, pese a los casi 9 años de diferencia en la realización de los dos estudios, sugiere una poca eficacia por parte de las estrategias de intervención para controlar y reducir los casos en Mazan.

Además, en el presente estudio también se identificaron las subpoblaciones genéticas presentes en las distintas comunidades analizadas. La presencia de estructuración poblacional es el resultado de procesos naturales e históricos que generan aislamiento reproductivo en los parásitos (3, 16-18, 63, 77, 78). En términos prácticos, las infecciones parasitarias que pertenecen a una misma subpoblación genética comparten factores de riesgo similares que les dieron origen; mientras que infecciones que pertenecen a subpoblaciones genéticas distintas son producto de

factores o causas distintas, es decir, tienen un origen distinto. De esta manera en la comunidad de Libertad se observó hasta 5 sub-poblaciones genéticas (haplogrupos) distintas, lo cual sugiere la confluencia de múltiples factores de riesgo que dan origen a las infecciones en un espacio geográfico reducido. Esta comunidad posee la mayor tasa de prevalencia parasitaria entre las cuatro comunidades analizadas; sin embargo, en un estudio reciente no se encontró asociación significativa entre la prevalencia parasitaria y algún factor de riesgo del hogar o del individuo (11). Lo cual demuestra que cuando coexisten múltiples fuentes de poblaciones parasitaria en una región geográfica limitada, la estructuración poblacional es una herramienta útil para identificar y evaluar la contribución de estos factores riesgo favorables a la transmisión de parásitos.

Esta estructuración poblacional también fue empleada para identificar y estimar la proporción de infecciones migrantes en cada comunidad y cuenca. Definir el patrón de distribución y dispersión de la enfermedad es clave para la selección de medidas de control, ya que en base a ello se puede redirigir los recursos de manera más efectiva (78). Existen múltiples aproximaciones para poder definir el flujo genético y las rutas de dispersión geográficas de las poblaciones parasitarias, siendo aquellas basadas en coalescencia las más utilizadas (19, 78, 100). Estas permiten medir la tasa de migración de manera bidireccional (importación y exportación de casos) entre las poblaciones geográficas analizadas, lo cual es útil para determinar la direccionalidad de la dispersión.

Un estudio realizado en el 2016 con muestras provenientes de 5 poblaciones de la amazonia propuso trece posibles modelos que explican el flujo genético entre las poblaciones analizadas (19). Sin embargo, el modelo con mayor soporte fue el modelo de panmixia, o modelo de apareamiento reproductivo aleatorio, y sugiere que en las comunidades analizadas en ese estudio no existe subdivisión geográfica de la población de parásitos, lo cual es inconsistente con el registro de desplazamiento humano en esa región. Si bien los modelos basados en coalescencia permiten determinar la direccionalidad del flujo genético, estos no permiten definir que infecciones son autóctonas y cuales son migrantes, lo cual es clave para determinar las causas que propiciaron este flujo.

Por este motivo en este trabajo se optó por elegir una aproximación distinta, en la cual se propuso un modelo bayesiano que permita establecer la probabilidad de cada infección de ser autóctona o migrante (importada o introducida). De esta manera nuestro modelo propuesto permitió identificar 20 (22.73%) infecciones foráneas en la cuenca Mazán, siendo 12 de ellas procedentes de la comunidad de San José de Lupuna (cuenca Nanay), lo cual es consistente con la cercanía geográfica de estas dos regiones respecto a las otras áreas analizadas. No obstante, el algoritmo empleado para poder diferenciar entre infecciones autóctonas y migrantes, solo es capaz de detectar importaciones o introducciones que involucren un bajo número de genotipos (101). Es decir, los genotipos migrantes similares entre sí deben representar un bajo porcentaje del total de la población donde han ingresado; ya que si el porcentaje es elevado, estos genotipos similares son identificados como si fuesen una población

autóctona. Esto trae como consecuencia que las infecciones reportadas como migrantes en este trabajo no posean ninguna relación entre sí; que no se encuentre asociación entre estas infecciones y alguna variable socio-demográfica debido a que los orígenes de estas infecciones pueden ser múltiples; y que el efecto real de la movilización humana en esta cuenca posiblemente esté subestimado.

Además de esta gran proporción de infecciones migrantes, los resultados también mostraron una amplia dispersión geográfica de al menos 6 sub-poblaciones genéticas de parásitos; y expansión clonal de ellas y de sus híbridos en las distintas cuencas analizadas. Existen dos características biológicas importantes de *P. vivax* que favorecen los patrones de estructuración, intercambio génico y diversidad observados. Esas características son la presencia de hipnozoítos y las infecciones asintomáticas.

Los hipnozoítos son estadios durmientes que permanecen ocultos en el hígado, los cuales pueden causar relapsos después de semanas o incluso después de meses de haberse dado la infección (45). Esto conlleva a un aumento de las infecciones policlonales, incrementando la probabilidad de recombinación meiótica en el mosquito, e incremento la diversidad (8, 9, 36, 43, 44, 102). En la Amazonía Peruana se ha reportado hasta un 57% de episodios recurrentes luego de un tratamiento de cura radical supervisado (46); sin embargo, la baja sensibilidad de los seromarcadores de infección reciente hace difícil discriminar que porcentaje del total de infecciones recurrentes se deben a verdaderos episodios de relapsos. Lamentablemente el diseño de estudio de este trabajo no permite diferenciar entre nuevas infecciones y episodios de relapsos, por lo que no es posible poder evaluar de

manera directa y objetiva la contribución de estos últimos en la diversidad y patrones de estructuración observados.

El segundo factor involucra a las infecciones asintomáticas las cuales escapan a las herramientas y estrategias de vigilancia y control actuales, por lo que su falta de manejo puede también incrementar el número de infecciones policlonales, la recombinación, y diversidad parasitaria (103-106). Según los registros colectados por el proyecto ICEMR Amazonía realizado desde diciembre del 2012 hasta diciembre del 2015, en el Perú las infecciones asintomáticas representan un 83.2% del total de casos, de los cuales el 70.6% son infecciones sub-microscópicas. Considerando que el total de individuos que poseen una infección de malaria (sea sintomática o asintomática) representa el 15.5% de la población total en riesgo, las estrategias de manejo y detección de casos actualmente empleadas dejan desatendidas a un 13.9% de la población total (14).

Además, alrededor de un 50% del total de infecciones causadas por *P. vivax* poseen densidades parasitarias por debajo de 10 parásitos/ μ L (107), lo cual dificulta el rendimiento incluso de los métodos moleculares de diagnóstico (108) y de caracterización genética de parásitos (100). No obstante, aún no se conoce el nivel de contribución de las infecciones sub-microscópicas asintomáticas a la transmisión. Si bien se ha detectado la presencia de gametocitos en este tipo de infecciones (105, 109), aún no se sabe si existe un umbral mínimo de densidad parasitaria y/o gametocitaria para que un paciente infectado pueda transmitir los estadios sexuales al mosquito.

En este trabajo se observaron diferencias significativas (p -valor = 0.034) en la diversidad genética de las infecciones asintomáticas y sintomáticas. Además, el análisis de red de expansión mínima de haplotipos sugiere que la ausencia de control sobre este grupo de infecciones (asintomáticas y/o sub-microscópicas), ha permitido el incremento y diversificación poblacional del parásito, y además sirven como reservorio y fuente a la población humana no expuesta o susceptible. Esta ausencia de control también podría permitir un mayor desplazamiento geográfico de las poblaciones parasitarias cuando se combinan con factores sociales, epidemiológicos y ambientales que fomentan el desplazamiento humano y del vector. Sin embargo, estos análisis son preliminares y es necesario el empleo de métodos genéticos más robustos y adecuados, y tamaños de muestra más representativos que permitan hacer un mejor rastreo de las infecciones.

Por último, mediante la asociación de las subpoblaciones genéticas encontradas con distintos factores demográficos y clínico-epidemiológicos de los pacientes infectados, se midió la contribución de estos factores de riesgo a la transmisión de la enfermedad en el distrito y en cada cuenca estudiada, y las consecuencias clínicas sobre la población susceptible. Se encontró asociación entre las personas que realizaron algún viaje fuera de su comunidad en el último mes, y la presencia de infecciones causadas por el haplogrupo 2. Además, también se observó asociación entre convivir con alguien que haya realizado un viaje reciente, y pertenecer al haplogrupo 1, con la presencia de síntomas, sugiriendo que ambos haplogrupos han sido introducidos dentro de esta cuenca recientemente. Estos resultados revelan en primer lugar, que las

comunidades de la cuenca Mazán no solo son vulnerables por la elevada tasa de importación de infecciones, si no también, que son altamente receptivas para poder transmitir estos nuevos genotipos dentro de la comunidad (11, 110). En segundo lugar, los resultados sugieren que el ingreso de nuevas poblaciones parasitarias (haplogrupos 1 y 2) podría traer consigo el ingreso de nuevos repertorios antigénicos, lo cual tiene efecto sobre las manifestaciones clínicas de la población humana afectada. No obstante, en este trabajo no se analizó la diversidad genética de regiones codantes o genes con capacidad inmunogénica, por lo que es recomendable este tipo de estudios para confirmar y complementar este hallazgo.

Finalmente, este trabajo resalta la utilidad de la información proporcionada por los parámetros genético poblacionales, los cuales, si son empleados e interpretados de manera adecuada, brindan información valiosa a los decisores de políticas públicas (3, 16, 18, 111). Los resultados mostrados en este estudio sugieren la presencia de múltiples factores que favorecen la dispersión de las poblaciones parasitarias a lo largo de las cuencas que pertenecen a la jurisdicción del distrito de Mazán. La contribución de cada uno de estos factores varía a nivel geográfico, por lo que conocer donde y cuando un factor otorga una mayor contribución a la transmisión, ayuda a poder focalizar los recursos, muchas veces limitados, que posee el estado para combatir esta enfermedad. Uno de los factores más importantes identificado en este estudio fue la movilización humana, la cual ha ocasionado una elevada importación e introducción de poblaciones parasitarias específicamente en la comunidad de Libertad. Esta información debe ser tomada en cuenta en la

elaboración de los planes de control, no solo para focalizar los recursos, si no también, para determinar que estrategias son las más adecuadas para atacar este problema.

Pese a ello, aún hay algunas limitaciones que las herramientas genético poblacionales deben superar para poder ser integradas en los programas de vigilancia (3, 17, 111). Quizás la limitación más importante es la baja sensibilidad mostrada en infecciones sub-microscópicas, el cual es un problema recurrente en este tipo de estudios, pero que sin embargo ha sido poco documentado (100). Otra limitación es la falta de integración de los estudios genéticos y los estudios epidemiológicos en malaria. Pese a que la genética poblacional y la epidemiología poseen más cosas en común que diferencias, ambas disciplinas han trabajado de manera aislada por mucho tiempo (17, 111). Sin embargo, recientemente se han incrementado los esfuerzos por construir un lenguaje común entre las distintas que están involucradas en el estudio y prevención de esta enfermedad, lo cual ha realzado el rol de la genética poblacional como disciplina que complementa y enriquece los métodos tradicionales. Finalmente es necesario incrementar los esfuerzos para el desarrollo de métodos que otorguen información pertinente en un menor tiempo de manera costo-efectiva. El escalonamiento de los métodos basados en secuenciamiento de nueva generación, y la focalización de estas herramientas a preguntas de interés epidemiológico concretas, hará posible determinar un rol claro de la genética poblacional en la toma de decisiones en políticas de salud pública.

IX. CONCLUSIONES

La selección y focalización óptima las estrategias de intervención es indispensable para el control y eliminación de malaria. Para lograrlo es necesario que las herramientas utilizadas para la vigilancia den luces acerca de la contribución de los distintos factores que fomentan la transmisión de los parásitos.

Los resultados obtenidos en este estudio mostraron que la comunidad de Libertad tuvo mayor diversidad genética ($H_{exp} = 0.705$), confluencia de un mayor número de subpoblaciones genéticas parasitarias (5 subpoblaciones o haplogrupos), y menores niveles de endogamia, con respecto a las otras dos comunidades ubicadas en la misma micro-cuenca, y a la comunidad de Urco Miraño ubicada en el mismo distrito. Esto confirma la variación en los factores que favorecen la transmisión de malaria, tanto a nivel intra-distrital como dentro de una misma microcuenca; y enfatiza la necesidad de tener estrategias que tomen en consideración esta variabilidad geográfica y micro-geográfica para hacer que las medidas de control sean realmente efectivas.

Por otro lado, nuestros resultados también indican que las comunidades de la cuenca Mazan poseen una mayor proporción de importación e introducción de nuevos genotipos (22.73% de infecciones foráneas) que la comunidad de Urco Miraño (solo tres infecciones importadas), y que el haplogrupo 2 es una subpoblación genética recientemente introducida en la zona limítrofe entre la comunidad de Libertad y la comunidad de Primero de enero. Además, se encontró que la presencia de síntomas

en la cuenca Mazán es más probable en personas infectadas que conviven con alguien que haya salido fuera de la comunidad recientemente.

Por lo que estos resultados muestran que las comunidades de la cuenca Mazán, a diferencia de la comunidad de Urco Miraño en Napo, son vulnerables al ingreso y propagación de nuevas subpoblaciones parasitarias con consecuencias clínicas sobre la población humana susceptible. Esta información tiene un alto impacto en las medidas de prevención y control que deben ser implementadas en las comunidades de la cuenca Mazán, las cuales deben incluir el desplazamiento humano como factor en el diseño de las estrategias para esta zona y sus zonas colindantes, para así evitar la reintroducción de la enfermedad en escenarios aparentemente controlados, y también evitar la dispersión de la enfermedad hacia otras zonas también vulnerables y receptoras.

X. REFERENCIAS BIBLIOGRÁFICAS

1. Griffing SM, Gamboa D, Udhayakumar V. The history of 20th century malaria control in Peru. *Malar J*. 2013;12:303.
2. 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*. 2016;95(6 Suppl):133-44.
3. Barry AE, Waltmann A, Koepfli C, Barnadas C, Mueller I. Uncovering the transmission dynamics of Plasmodium vivax using population genetics. *Pathog Glob Health*. 2015;109(3):142-52.
4. Bousema T, Drakeley C. Epidemiology and infectivity of Plasmodium falciparum and Plasmodium vivax gametocytes in relation to malaria control and elimination. *Clin Microbiol Rev*. 2011;24(2):377-410.
5. Guerra CA, Snow RW, Hay SI. Defining the global spatial limits of malaria transmission in 2005. *Adv Parasitol*. 2006;62:157-79.
6. McKenzie FE, Jeffery GM, Collins WE. Gametocytemia and fever in human malaria infections. *J Parasitol*. 2007;93(3):627-33.
7. McKenzie FE, Jeffery GM, Collins WE. Plasmodium vivax blood-stage dynamics. *J Parasitol*. 2002;88(3):521-35.
8. Chen N, Auliff A, Rieckmann K, Gatton M, Cheng Q. Relapses of Plasmodium vivax infection result from clonal hypnozoites activated at predetermined intervals. *J Infect Dis*. 2007;195(7):934-41.
9. Imwong M, Snounou G, Pukrittayakamee S, Tanomsing N, Kim JR, Nandy A, et al. Relapses of Plasmodium vivax infection usually result from activation of heterologous hypnozoites. *J Infect Dis*. 2007;195(7):927-33.
10. Koepfli C, Rodrigues PT, Antao T, Orjuela-Sanchez P, Van den Eede P, Gamboa D, et al. Plasmodium vivax Diversity and Population Structure across Four Continents. *PLoS Negl Trop Dis*. 2015;9(6):e0003872.
11. 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*. 2017;7(1):8082.
12. Bousema T, Griffin JT, Sauerwein RW, Smith DL, Churcher TS, Takken W, et al. Hitting hotspots: spatial targeting of malaria for control and elimination. *PLoS Med*. 2012;9(1):e1001165.
13. Mosha JF, Sturrock HJ, Greenhouse B, Greenwood B, Sutherland CJ, Gadalla N, et al. Epidemiology of subpatent Plasmodium falciparum infection: implications for detection of hotspots with imperfect diagnostics. *Malar J*. 2013;12:221.
14. 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*. 2017;16(1):312.

15. Sturrock HJ, Bennett AF, Midekisa A, Gosling RD, Gething PW, Greenhouse B. Mapping Malaria Risk in Low Transmission Settings: Challenges and Opportunities. *Trends Parasitol.* 2016;32(8):635-45.
16. Arnott A, Barry AE, Reeder JC. Understanding the population genetics of *Plasmodium vivax* is essential for malaria control and elimination. *Malar J.* 2012;11:14.
17. Chenet SM, Schneider KA, Villegas L, Escalante AA. Local population structure of *Plasmodium*: impact on malaria control and elimination. *Malar J.* 2012;11:412.
18. Escalante AA, Ferreira MU, Vinetz JM, Volkman SK, Cui L, Gamboa D, et al. Malaria Molecular Epidemiology: Lessons from the International Centers of Excellence for Malaria Research Network. *Am J Trop Med Hyg.* 2015;93(3 Suppl):79-86.
19. Delgado-Ratto C, Gamboa D, Soto-Calle VE, Van den Eede P, Torres E, Sanchez-Martinez L, et al. Population Genetics of *Plasmodium vivax* in the Peruvian Amazon. *PLoS Negl Trop Dis.* 2016;10(1):e0004376.
20. Delgado-Ratto C, Soto-Calle VE, Van den Eede P, Gamboa D, Rosas A, Abatih EN, et al. Population structure and spatio-temporal transmission dynamics of *Plasmodium vivax* after radical cure treatment in a rural village of the Peruvian Amazon. *Malar J.* 2014;13:8.
21. Sutton PL, Neyra V, Hernandez JN, Branch OH. *Plasmodium falciparum* and *Plasmodium vivax* infections in the Peruvian Amazon: propagation of complex, multiple allele-type infections without super-infection. *Am J Trop Med Hyg.* 2009;81(6):950-60.
22. Van den Eede P, Van der Auwera G, Delgado C, Huyse T, Soto-Calle VE, Gamboa D, et al. Multilocus genotyping reveals high heterogeneity and strong local population structure of the *Plasmodium vivax* population in the Peruvian Amazon. *Malar J.* 2010;9:151.
23. Feo Istúriz O, Tobar Arias K, Pineda Restrepo BL, Rosas Aguirre A, Vargas Pachérrez D, Soto Alvariano G. Compartiendo lecciones aprendidas. Proyecto control de malaria en zonas fronterizas de la región andina: un enfoque comunitario-PAMAFRO. 2009.
24. Hiwat H, Bretas G. Ecology of *Anopheles darlingi* Root with respect to vector importance: a review. *Parasit Vectors.* 2011;4:177.
25. Sinka ME, Rubio-Palis Y, Manguin S, Patil AP, Temperley WH, Gething PW, et al. The dominant *Anopheles* vectors of human malaria in the Americas: occurrence data, distribution maps and biometric precis. *Parasit Vectors.* 2010;3:72.
26. Bruce-Chwatt LJ. History of malaria from prehistory to eradication. *Malaria Principles and Practice of Malariology.* 1988;1:1-59.
27. Sherman I. A brief history of malaria and discovery of the parasite's life cycle. *Malaria: parasite biology, pathogenesis, and protection.* 1998:3-10.
28. Sherman IW. The life of plasmodium: An overview. *Molecular approaches to malaria: American Society of Microbiology;* 2005. p. 1-11.

29. Ross R. *Memoirs: with a Full Account of the Great Malaria Problem and its Solution*. 1923.
30. de Kruif P. *The Microbe Hunters*. 1926. New York: Harcourt; 1953.
31. Harrison G. *Mosquitoes, malaria and man: a history of the hostilities since 1880*. 1978.
32. Zilversmit M, Hartl DL. Evolutionary history and population genetics of human malaria parasites. *Molecular approaches to malaria: American Society of Microbiology*; 2005. p. 95-109.
33. Escalante AA, Cornejo OE, Freeland DE, Poe AC, Durrego E, Collins WE, et al. A monkey's tale: the origin of *Plasmodium vivax* as a human malaria parasite. *Proc Natl Acad Sci U S A*. 2005;102(6):1980-5.
34. Cornejo OE, Escalante AA. The origin and age of *Plasmodium vivax*. *Trends in parasitology*. 2006;22(12):558-63.
35. Nacher M, Stefani A, Basurko C, Lemonnier D, Djossou F, Demar M, et al. The burden of *Plasmodium vivax* relapses in an Amerindian village in French Guiana. *Malar J*. 2013;12:367.
36. White NJ. Why Do Some Primate Malarial Relapse? *Trends Parasitol*. 2016;32(12):918-20.
37. Oakley MS, Gerald N, McCutchan TF, Aravind L, Kumar S. Clinical and molecular aspects of malaria fever. *Trends Parasitol*. 2011;27(10):442-9.
38. Schofield L, Grau GE. Immunological processes in malaria pathogenesis. *Nat Rev Immunol*. 2005;5(9):722-35.
39. Lopez C, Yepes-Perez Y, Hincapie-Escobar N, Diaz-Arevalo D, Patarroyo MA. What Is Known about the Immune Response Induced by *Plasmodium vivax* Malaria Vaccine Candidates? *Front Immunol*. 2017;8:126.
40. Eichner M, Diebner HH, Molineaux L, Collins WE, Jeffery GM, Dietz K. Genesis, sequestration and survival of *Plasmodium falciparum* gametocytes: parameter estimates from fitting a model to malariatherapy data. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. 2001;95(5):497-501.
41. Enger A, Strand OA, Ranheim T, Hellum KB. Exflagellation of microgametocytes in *Plasmodium vivax* malaria: a diagnostic conundrum. *Med Princ Pract*. 2004;13(5):298-300.
42. Bousema T, Dinglasan RR, Morlais I, Gouagna LC, van Warmerdam T, Awono-Ambene PH, et al. Mosquito feeding assays to determine the infectiousness of naturally infected *Plasmodium falciparum* gametocyte carriers. *PLoS One*. 2012;7(8):e42821.
43. Betuela I, Rosanas-Urgell A, Kiniboro B, Stanistic DI, Samol L, de Lazzari E, et al. Relapses contribute significantly to the risk of *Plasmodium vivax* infection and disease in Papua New Guinean children 1-5 years of age. *J Infect Dis*. 2012;206(11):1771-80.
44. de Araujo FC, de Rezende AM, Fontes CJ, Carvalho LH, Alves de Brito CF. Multiple-clone activation of hypnozoites is the leading cause of relapse in *Plasmodium vivax* infection. *PLoS One*. 2012;7(11):e49871.

45. Kim JR, Nandy A, Maji AK, Addy M, Dondorp AM, Day NP, et al. Genotyping of *Plasmodium vivax* reveals both short and long latency relapse patterns in Kolkata. *PLoS One*. 2012;7(7):e39645.
46. Van den Eede P, Soto-Calle VE, Delgado C, Gamboa D, Grande T, Rodriguez H, et al. *Plasmodium vivax* sub-patent infections after radical treatment are common in Peruvian patients: results of a 1-year prospective cohort study. *PLoS One*. 2011;6(1):e16257.
47. Alarcón Baldeón JJ. Dinámica de las parasitemias recurrentes de *Plasmodium vivax* determinadas por PCR en pobladores de las comunidades de San José de Lupuna y Cahuide de la región de Loreto. 2016.
48. Koepfli C, Colborn KL, Kiniboro B, Lin E, Speed TP, Siba PM, et al. A high force of *plasmodium vivax* blood-stage infection drives the rapid acquisition of immunity in papua new guinean children. *PLoS Negl Trop Dis*. 2013;7(9):e2403.
49. Koepfli C, Ross A, Kiniboro B, Smith TA, Zimmerman PA, Siba P, et al. Multiplicity and diversity of *Plasmodium vivax* infections in a highly endemic region in Papua New Guinea. *PLoS Negl Trop Dis*. 2011;5(12):e1424.
50. Schneider KA, Escalante AA. Fitness components and natural selection: why are there different patterns on the emergence of drug resistance in *Plasmodium falciparum* and *Plasmodium vivax*? *Malaria journal*. 2013;12(1):15.
51. Moreno M, Saavedra MP, Bickersmith SA, Prussing C, Michalski A, Tong Rios C, et al. Intensive trapping of blood-fed *Anopheles darlingi* in Amazonian Peru reveals unexpectedly high proportions of avian blood-meals. *PLoS Negl Trop Dis*. 2017;11(2):e0005337.
52. Parker BS, Paredes Olortegui M, Penataro Yori P, Escobedo K, Florin D, Rengifo Pinedo S, et al. Hyperendemic malaria transmission in areas of occupation-related travel in the Peruvian Amazon. *Malar J*. 2013;12:178.
53. Shililu J, Ghebremeskel T, Seulu F, Mengistu S, Fekadu H, Zerom M, et al. Seasonal abundance, vector behavior, and malaria parasite transmission in Eritrea. *J Am Mosq Control Assoc*. 2004;20(2):155-64.
54. Durnez L, Coosemans M. Residual transmission of malaria: an old issue for new approaches. *Anopheles mosquitoes-New insights into malaria vectors: Intech*; 2013.
55. Prussing C, Moreno M, Saavedra MP, Bickersmith SA, Gamboa D, Alava F, et al. Decreasing proportion of *Anopheles darlingi* biting outdoors between long-lasting insecticidal net distributions in peri-Iquitos, Amazonian Peru. *Malar J*. 2018;17(1):86.
56. Cibulskis RE, Aregawi M, Williams R, Otten M, Dye C. Worldwide incidence of malaria in 2009: estimates, time trends, and a critique of methods. *PLoS Med*. 2011;8(12):e1001142.
57. Corran P, Coleman P, Riley E, Drakeley C. Serology: a robust indicator of malaria transmission intensity? *Trends in parasitology*. 2007;23(12):575-82.
58. Drakeley C, Schellenberg D, Kihonda J, Sousa C, Arez A, Lopes D, et al. An estimation of the entomological inoculation rate for Ifakara: a semi-urban area in a region of intense malaria transmission in Tanzania. *Tropical Medicine & International Health*. 2003;8(9):767-74.

59. Smith DL, Dushoff J, McKenzie FE. The risk of a mosquito-borne infection in a heterogeneous environment. *PLoS Biol.* 2004;2(11):e368.
60. A research agenda for malaria eradication: diagnoses and diagnostics. *PLoS Med.* 2011;8(1):e1000396.
61. Hartl D, Clark A. Genetic and phenotypic variation. *Principles of population genetics.* Sunderland, Massachusetts: Fourth Edition Sinauer Associates Google Scholar 2007. p. 3-39.
62. Nielsen R, Slatkin MW. Introduction. *An introduction to population genetics: theory and applications:* Sinauer Associates; 2013. p. 1-19.
63. de Meeus T, McCoy KD, Prugnolle F, Chevillon C, Durand P, Hurtrez-Bousses S, et al. Population genetics and molecular epidemiology or how to "debusquer la bete". *Infect Genet Evol.* 2007;7(2):308-32.
64. Putman AI, Carbone I. Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and evolution.* 2014;4(22):4399-428.
65. 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*. *Molecular biology and evolution.* 2000;17(10):1467-82.
66. Chenet SM. *Plasmodium Population Structure in the Context of Malaria Control and Elimination:* Arizona State University; 2014.
67. Meirmans PG, Hedrick PW. Assessing population structure: FST and related measures. *Molecular Ecology Resources.* 2011;11(1):5-18.
68. Slatkin M. A measure of population subdivision based on microsatellite allele frequencies. *Genetics.* 1995;139(1):457-62.
69. Zhivotovsky LA. Estimating divergence time with the use of microsatellite genetic distances: impacts of population growth and gene flow. *Molecular Biology and Evolution.* 2001;18(5):700-9.
70. Hartl D, Clark A. Random genetic drift. *Principles of population genetics.* Sunderland, Massachusetts: Fourth Edition Sinauer Associates Google Scholar 2007. p. 95-148.
71. Volkman SK, Neafsey DE, Schaffner SF, Park DJ, Wirth DF. Harnessing genomics and genome biology to understand malaria biology. *Nat Rev Genet.* 2012;13(5):315-28.
72. Hartl D, Clark A. Organization of genetic variation. *Principles of population genetics.* Sunderland, Massachusetts: Fourth Edition Sinauer Associates Google Scholar 2007. p. 45-91.
73. Slatkin M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nat Rev Genet.* 2008;9(6):477-85.
74. Brown A, Feldman M, Nevo E. Multilocus structure of natural populations of *Hordeum spontaneum*. *Genetics.* 1980;96(2):523-36.
75. Charlesworth B, Charlesworth D, Barton NH. The effects of genetic and geographic structure on neutral variation. *Annual Review of Ecology, Evolution, and Systematics.* 2003;34(1):99-125.

76. Meirmans PG, Hedrick PW. Assessing population structure: F(ST) and related measures. *Mol Ecol Resour.* 2011;11(1):5-18.
77. Mita T, Jombart T. Patterns and dynamics of genetic diversity in *Plasmodium falciparum*: what past human migrations tell us about malaria. *Parasitol Int.* 2015;64(3):238-43.
78. Lo E, Zhou G, Oo W, Lee MC, Baum E, Felgner PL, et al. Molecular inference of sources and spreading patterns of *Plasmodium falciparum* malaria parasites in internally displaced persons settlements in Myanmar-China border area. *Infect Genet Evol.* 2015;33:189-96.
79. Meirmans PG. Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution.* 2006;60(11):2399-402.
80. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000;155(2):945-59.
81. Rosas-Aguirre A, Llanos-Cuentas A, Speybroeck N, Cook J, Contreras-Mancilla J, Soto V, et al. Assessing malaria transmission in a low endemicity area of north-western Peru. *Malar J.* 2013;12:339.
82. Senami. 2015 [Available from: <http://www.senamhi.gob.pe>].
83. Mangold KA, Manson RU, Koay ES, Stephens L, Regner M, Thomson RB, Jr., et al. Real-time PCR for detection and identification of *Plasmodium* spp. *J Clin Microbiol.* 2005;43(5):2435-40.
84. Imwong M, Sudimack D, Pukrittayakamee S, Osorio L, Carlton JM, Day NP, et al. Microsatellite variation, repeat array length, and population history of *Plasmodium vivax*. *Mol Biol Evol.* 2006;23(5):1016-8.
85. Karunaweera N, Ferreira M, Hartl D, Wirth D. Fourteen polymorphic microsatellite DNA markers for the human malaria parasite *Plasmodium vivax*. *Molecular Ecology Notes.* 2007;7(1):172-5.
86. Sutton PL. A call to arms: on refining *Plasmodium vivax* microsatellite marker panels for comparing global diversity. *Malar J.* 2013;12:447.
87. Manrique P, Hoshi M, Fasabi M, Nolasco O, Yori P, Calderon M, et al. Assessment of an automated capillary system for *Plasmodium vivax* microsatellite genotyping. *Malar J.* 2015;14:326.
88. Kamvar ZN, Brooks JC, Grunwald NJ. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front Genet.* 2015;6:208.
89. Kamvar ZN, Tabima JF, Grunwald NJ. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ.* 2014;2:e281.
90. Nei M, Roychoudhury AK. Sampling variances of heterozygosity and genetic distance. *Genetics.* 1974;76(2):379-90.
91. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics.* 2008;24(11):1403-5.
92. Agapow PM, Burt A. Indices of multilocus linkage disequilibrium. *Molecular Ecology Resources.* 2001;1(1- 2):101-2.
93. Hedrick PW. A standardized genetic differentiation measure. *Evolution.* 2005;59(8):1633-8.

94. Winter DJ. MMOD: an R library for the calculation of population differentiation statistics. *Molecular Ecology Resources*. 2012;12(6):1158-60.
95. Dray S, Dufour A-B. The ade4 package: implementing the duality diagram for ecologists. *Journal of statistical software*. 2007;22(4):1-20.
96. Su X, Ferdig MT, Huang Y, Huynh CQ, Liu A, You J, et al. A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum*. *Science*. 1999;286(5443):1351-3.
97. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol*. 2005;14(8):2611-20.
98. Earl DA. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation genetics resources*. 2012;4(2):359-61.
99. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*. 2003;164(4):1567-87.
100. Fola AA, Nate E, Abby Harrison GL, Barnadas C, Hetzel MW, Iga J, et al. Nationwide genetic surveillance of *Plasmodium vivax* in Papua New Guinea reveals heterogeneous transmission dynamics and routes of migration amongst subdivided populations. *Infect Genet Evol*. 2017;58:83-95.
101. Hubisz MJ, Falush D, Stephens M, Pritchard JK. Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour*. 2009;9(5):1322-32.
102. Adekunle AI, Pinkevych M, McGready R, Luxemburger C, White LJ, Nosten F, et al. Modeling the dynamics of *Plasmodium vivax* infection and hypnozoite reactivation in vivo. *PLoS Negl Trop Dis*. 2015;9(3):e0003595.
103. Karl S, Gurarie D, Zimmerman PA, King CH, St Pierre TG, Davis TM. A sub-microscopic gametocyte reservoir can sustain malaria transmission. *PLoS One*. 2011;6(6):e20805.
104. Lindblade KA, Steinhardt L, Samuels A, Kachur SP, Slutsker L. The silent threat: asymptomatic parasitemia and malaria transmission. *Expert Rev Anti Infect Ther*. 2013;11(6):623-39.
105. Rovira-Vallbona E, Contreras-Mancilla JJ, Ramirez R, Guzman-Guzman 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*. 2017;11(7):e0005674.
106. Sturrock HJ, Hsiang MS, Cohen JM, Smith DL, Greenhouse B, Bousema T, et al. Targeting asymptomatic malaria infections: active surveillance in control and elimination. *PLoS Med*. 2013;10(6):e1001467.
107. Carrasco-Escobar G, Miranda-Alban J, Fernandez-Minope C, Brouwer KC, Torres K, Calderon M, et al. High prevalence of very-low *Plasmodium falciparum* and *Plasmodium vivax* parasitaemia carriers in the Peruvian Amazon: insights into local and occupational mobility-related transmission. *Malar J*. 2017;16(1):415.
108. 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.

109. Vallejo AF, Garcia J, Amado-Garavito AB, Arevalo-Herrera M, Herrera S. Plasmodium vivax gametocyte infectivity in sub-microscopic infections. Malar J. 2016;15:48.

110. Reiner RC, Le Menach A, Kunene S, Ntshalintshali N, Hsiang MS, Perkins TA, et al. Mapping residual transmission for malaria elimination. Elife. 2015;4.

111. Greenhouse B, Smith DL. Malaria genotyping for epidemiologic surveillance. Proceedings of the National Academy of Sciences. 2015;112(22):6782-3.

XI. ANEXOS

ANEXO 1

Tabla suplementaria 1: Listado de cebadores y sus condiciones de hibridación y concentración.

MS	Cromosoma	Inicio	Motivo de repetición	Cebador Derecho	Cebador Izquierdo	Tamaño bp	Th	[nM]	Grupo
11.162*	Pv_Sal1_chr11	1859261	ATAC	6FAM-TTTGTTAGGAG ATCCGTCCTG	TAAATGACACTTTGG CTTC	222	60	200	1
14.297*	Pv_Sal1_chr14	2965458	AAG	PET-TACACCCCTTTAG GTCCCTCGT	TGAAAAATGTTCCGC TACTT	200	57	200	1
Ch2.121***	Pv_Sal1_chr02	121392	TTA	NED-CCTTGCTCTCC TGGCTCTTC	GTATGAGTGACCACC CTGGC	172	67	200	1
Ch2.152***	Pv_Sal1_chr02	152435	AAT	VIC-GTCATTTTCTAG GCAAAGACAGC	TACGTTTCGCCATGT TCCTG	153	63	400	1
Ch14.2986***	Pv_Sal1_chr14	2986814	AT	NED-TTC TAAATAA AGCACCATG	ACCATTCTGTTAAAA TGTATGT	175	57	200	2
Ch14.3010***	Pv_Sal1_chr14	3010226	AT	PET-GATAAGATAAAA ACAACAAATGCG	AGGATGCTTCGGTTT GCT	207	60	200	2
Ch14.3021***	Pv_Sal1_chr14	3021021	TA	6FAM-GCATTGTGTTAC ATTGTTTTCGT	GCAAGATGTGTCCTT TACA	228	57	400	2
Ch2.122***	Pv_Sal1_chr02	122021	ATA	VIC-AGCTGGATTGCC CTAACCTGC	CGCATCATCGCATACT GCTTG	197	60	200	2
13.239*	Pv_sal1_chr13	1601741	TTTA	NED-CAACAACAAA ATGAATGACG	CCTACCTTTACGCGT TTTTA	210	58	200	3
Ch14.2981***	Pv_Sal1_chr14	2981291	AT	PET-CGATGAATCAT TTTCGTCCGCA	ACTCTTCTTTCAGCAC TTTTGGC	161	63	400	3
MS6**	Pv_Sal1_chr11	1760698	(TCC)2(TCT)3(CCT)2(TCC)2 GCTTCT(TCC)10	VIC-GGTTCTTCGGT GATCTCTGC	CTGTCTTGAGGAGACA TCAACGGGATT	241	57	200	3
MS9**	Pv_Sal1_chr08	1394242	(GGA)18	6FAM-AGATGCCTAC ACGTTGACGA	CTGTCTTGAAGCTGC CCAATGTGGAAT	188	57	200	3
3.502*	Pv_sal1_chr03	451099	AACGGATG	PET-GTGGACCGATG GACCFAT	TCCTACTCAGGGGGA ATACT	168	55	400	4
MS15**	Pv_Sal1_chr05	1110660	(TCT)10	NED-TGTTTGCAAAG GAATCCACA	CTGTCTTCGGCCAGA TGAAAAGGATAA	243	57	400	4
MS20**	Pv_Sal1_chr10	1345613	(GAA)11GAG(GAA)13(CAA)4 GAA(CAA)5	VIC-GCACAACAAAT GCAAGATCC	CTGTCTTGTGGCAGT GGCTCATCTTCT	211	60	200	4
MS4**	Pv_Sal1_chr06	390261	(AGT)18	6FAM-CGATTTACTGT TGACGCTGAA	CTGTCTCAAAGGAA CATGCTCGATGA	220	57	200	4

* Marcadores reportados por Imwong, Sudimack (84)

** Marcadores reportados por Karunaweera, Ferreira (85)

*** Marcadores reportados por Alarcón Baldeón (47)

ANEXO 2

Tabla suplementaria 2: Porcentaje de amplificación promedio de los marcadores en cada rango de concentración de ADN del parásito

Categoría de Parasitemia	N	% amplificación MZN	% amplificación Global
> 8 mol/uL (Cq < 28.61)	93/251	98.12 % (min = 96.80%)	99.05 % (min = 98.01%)
4 – 8 mol/uL (28.65 < Cq < 31.5)	68/203	73.40 % (min = 64.20%)	82.80 % (min = 73.90%)
1.5 – 4 mol/uL (31.5 < Cq < 32.91)	41/140	77.10 % (min = 63.40%)	77.80 % (min = 69.10%)
0.8 – 1.5 mol/uL (Cq < 34.07)	56/98	56.80 % (min = 39.60%)	59.60 % (min = 44.20%)

N: Número de muestras totales genotipificadas en Mazán, y en las otras áreas de estudio.

% amplificación MZN: Respecto al total de muestras genotipificadas en MZN.

% amplificación Global: Respecto a la totalidad de muestras genotipificadas en todas las áreas de estudio (MZN más las 3 áreas de estudio adicionales).

ANEXO 3

Tabla suplementaria 3: Diversidad Haplotípica y Genética por comunidad, por cuenca en las cuatro áreas de estudio, en cada fecha de muestreo.

Área de Estudio	Cuenca	Comunidad	N	% Policlonales	MLG	H	E.5	Hexp	rbarD	p.rD	% Imp & Int (IC 95%)
Mazán		GC Junio 2015	4	50.00 (0.00 - 100.00)	4	1.386	1.000	0.802	0.059	0.369	50% (0 - 100%)
		GC Agosto 2015	3	66.67 (0.87 - 100.00)	3	1.099	1.000	0.542	1.000	0.001	33.3% (0 - 99.1%)
		Total GC	7	57.14 (17.27 - 97.02)	7	1.946	1.000	0.696	0.090	0.041	42.9% (3.0 - 82.7%)
		LI Junio 2015	47	38.30 (24.15 - 52.45)	38	3.509	0.810	0.705	0.072	0.001	17.0% (6.1 - 28.0%)
		Mazán LI Agosto 2015	16	37.50 (12.83 - 62.17)	15	2.686	0.967	0.663	0.145	0.001	31.3% (7.6 - 54.9%)
		Total LI	63	38.10 (25.92 - 50.27)	49	3.757	0.816	0.705	0.076	0.001	20.6% (10.5 - 30.8%)
		PE Junio 2015	15	6.67 (0.00 - 19.83)	12	2.396	0.880	0.588	0.204	0.001	20.0% (0 - 41.1%)
		PE Agosto 2015	3	66.67 (0.87 - 100.00)	3	1.099	1.000	0.750	0.167	0.175	33.3% (0 - 99.1%)
		Total PE	18	16.67 (0.00 - 34.51)	15	2.630	0.890	0.611	0.173	0.001	22.2% (2.3 - 42.1%)
		Cuenca Mazán Junio 2015	66	31.82 (20.41 - 43.22)	53	3.847	0.826	0.707	0.070	0.001	19.7% (10.0 - 29.4%)
		Cuenca Mazán Agosto 2015	22	45.45 (24.01 - 66.90)	21	3.028	0.975	0.678	0.104	0.001	31.8% (11.8 - 51.9%)
		Total Cuenca Mazán	88	35.23 (25.12 - 45.34)	69	4.110	0.833	0.707	0.065	0.001	22.3% (13.9 - 31.5%)
	Napo	UM Junio 2015	70	14.29 (5.97 - 22.60)	33	2.662	0.426	0.411	0.243	0.001	4.3% (0 - 9.1%)
	Napo	UM Agosto 2015	15	40.00 (14.16 - 65.84)	9	1.800	0.584	0.205	0.203	0.001	-
	Napo	Total UM	85	18.82 (10.40 - 27.24)	41	2.790	0.370	0.392	0.210	0.001	3.5% (0 - 7.5%)
	Area de Estudio Mazán	Junio 2015	136	22.79 (15.67 - 29.92)	86	3.930	0.418	0.610	0.092	0.001	11.8% (6.3 - 17.2%)
	Area de Estudio Mazán	Agosto 2015	37	43.24 (26.95 - 59.54)	30	3.205	0.672	0.621	0.107	0.001	18.9% (6.0 - 31.8%)
	Total area de Estudio Mazán		173	27.17 (20.47 - 33.86)	110	4.155	0.370	0.615	0.086	0.001	13.3% (8.2 - 18.4%)
Lupuna	Nanay	LUP	366	19.67 (15.59 - 23.76)	232	4.827	0.312	0.544	0.142	0.001	10.7% (7.5 - 13.9%)
Cahuide	Itaya	CAH	274	26.28 (21.05 - 31.51)	165	4.411	0.308	0.591	0.115	0.001	12.4% (8.5 - 16.3%)
Santa Emilia	Tigre	STE	137	40.15 (31.90 - 48.39)	111	4.558	0.729	0.596	0.161	0.001	12.4% (6.7 - 18.0%)
Total			950	25.89 (23.10 - 28.69)	602	5.763	0.277	0.655	0.076	0.001	11.9% (9.8 - 14.0%)

ANEXO 4

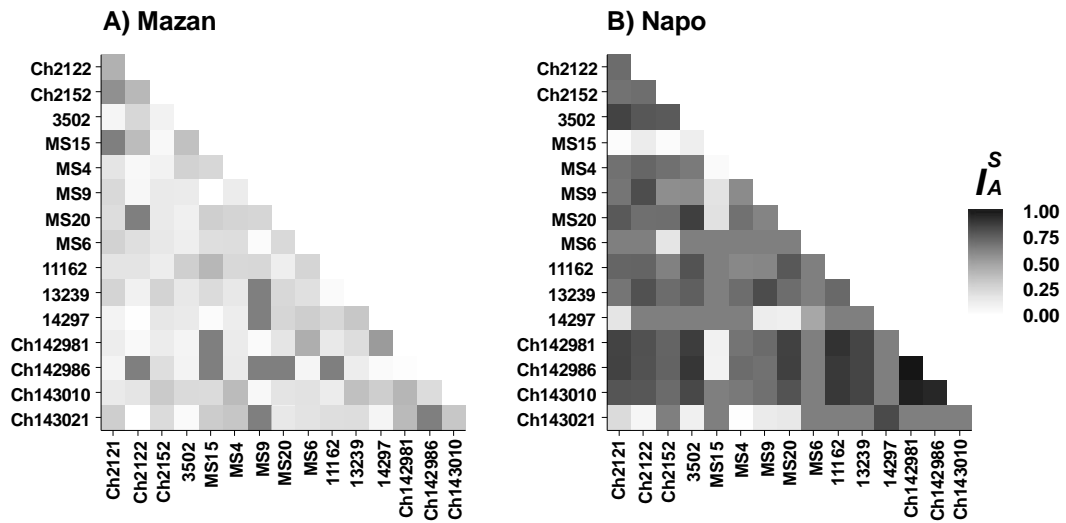


Figura suplementaria 4: Desequilibrio de ligamiento pareado en las cuencas Mazán (A) y Napo (B). La escala en grises representa el nivel de disequilibrio de ligamiento (de cero a uno), y cada recuadro representa la comparación entre los alelos de cada par de loci.

ANEXO 5

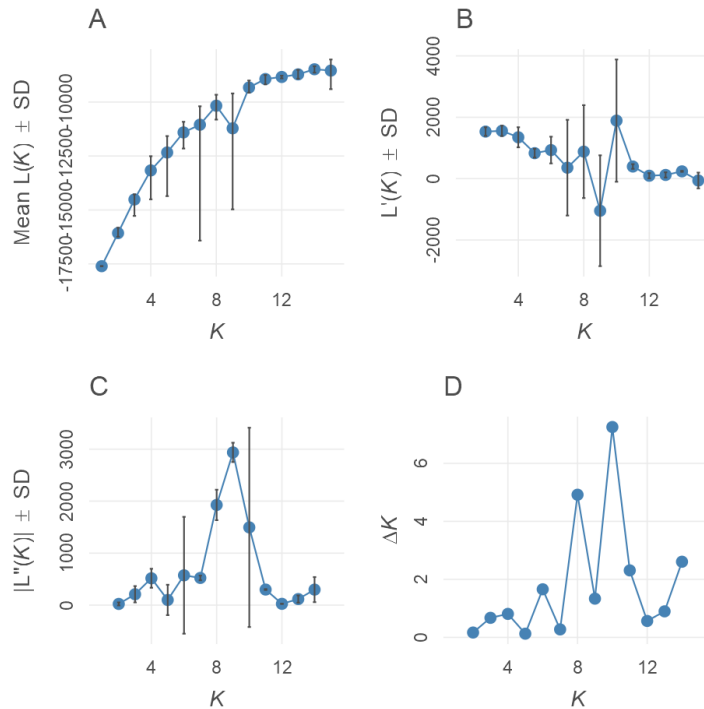


Figura suplementaria 2: Análisis de la tasa de cambio de segundo orden de la probabilidad marginal. En todos los gráficos el eje x representa el número de poblaciones posibles en el set de datos. A) Media aritmética del logaritmo de la probabilidad marginal de las 15 corridas realizadas asumiendo la presencia de 1 hasta 15 poblaciones. B) Primera diferencial del logaritmo de la probabilidad marginal. C) Segunda diferencial del logaritmo de la probabilidad marginal. D) Tasa de cambio de segundo orden de la probabilidad marginal normalizado con la varianza de las 250 mil iteraciones realizadas. Los picos en la figura D representan el número de poblaciones más probables presentes en el set de datos.

ANEXO 6

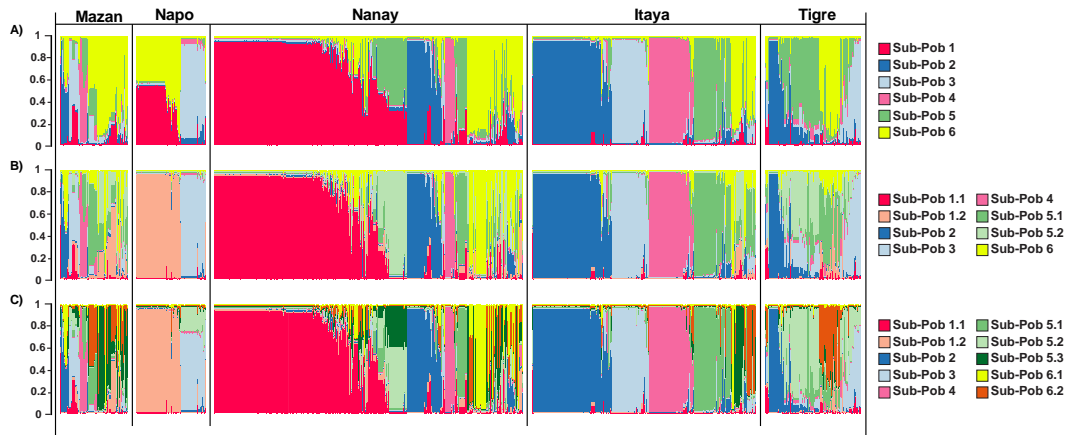


Figura suplementaria 3: Comparación de la Estructuración poblacional de *Plasmodium vivax* en cuatro cuencas de la Amazonía Peruana (Mazan, Napo, Nanay, Itaya y Santa Emilia) usando el modelo de Ligamiento (Modelo 1) asumiendo tres números de poblaciones ancestrales distintos. A) Estructuración en base a $K = 6$, B) $K = 8$, y C) $K = 10$. Cada color corresponde a una población ancestral, y cada barra indica la proporción del genotipo de un individuo que proviene de cada población ancestral (Coeficiente de Ancestría o Q).

ANEXO 7

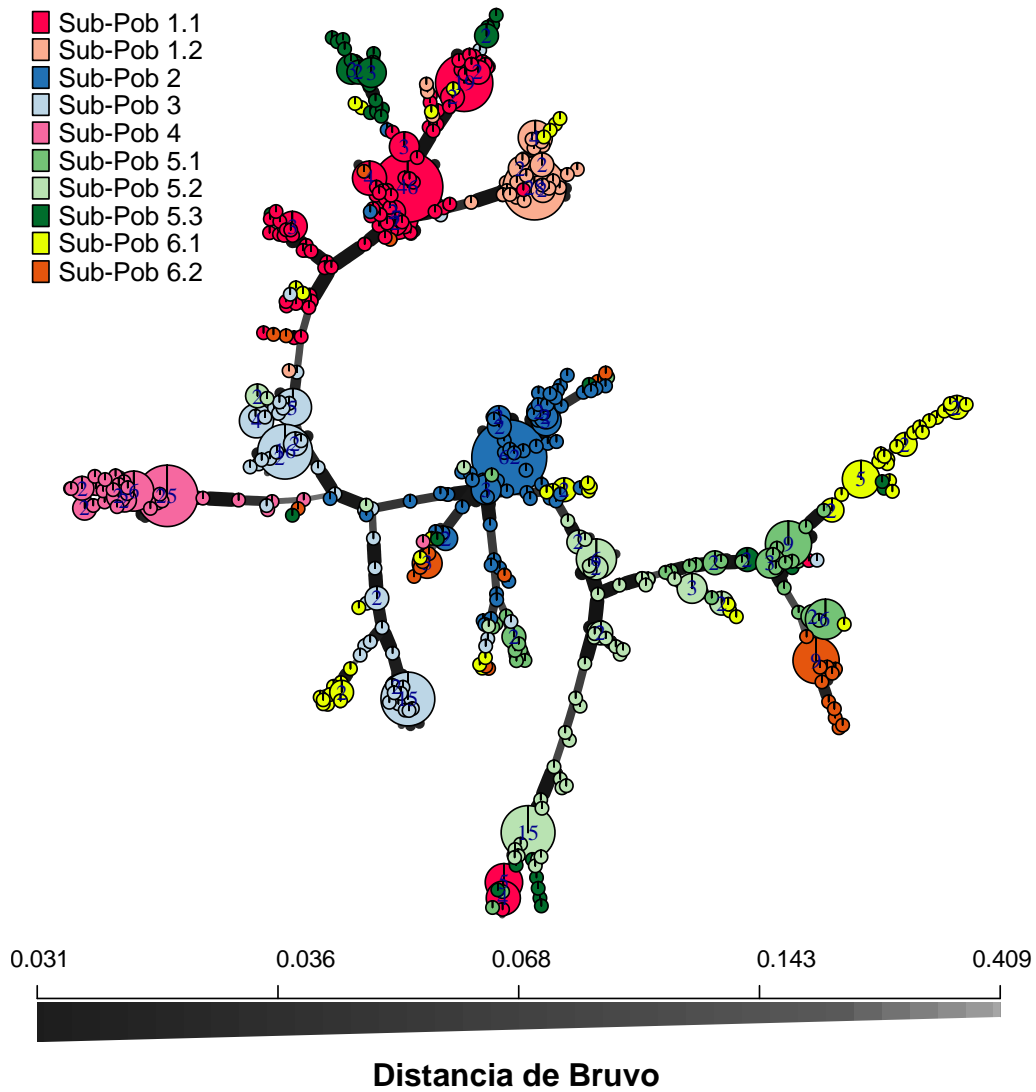


Figura suplementaria 4: Red mínima de expansão de haplotipos del total de infecciones genotipificadas en las 4 áreas de estudio (Mazan, Lupuna, Cahuide y Santa Emilia). Los colores han sido asignados de acuerdo a la pertenencia de cada infección a una Sub-Población Genética asumiendo la presencia de 10 Sub-Poblaciones. Cada círculo representa un haplotipo, y su tamaño guarda relación directa con el número de muestras que presentan ese haplotipo. Las relaciones genéticas entre los haplotipos están representadas por líneas que conectan cada círculo.

ANEXO 8

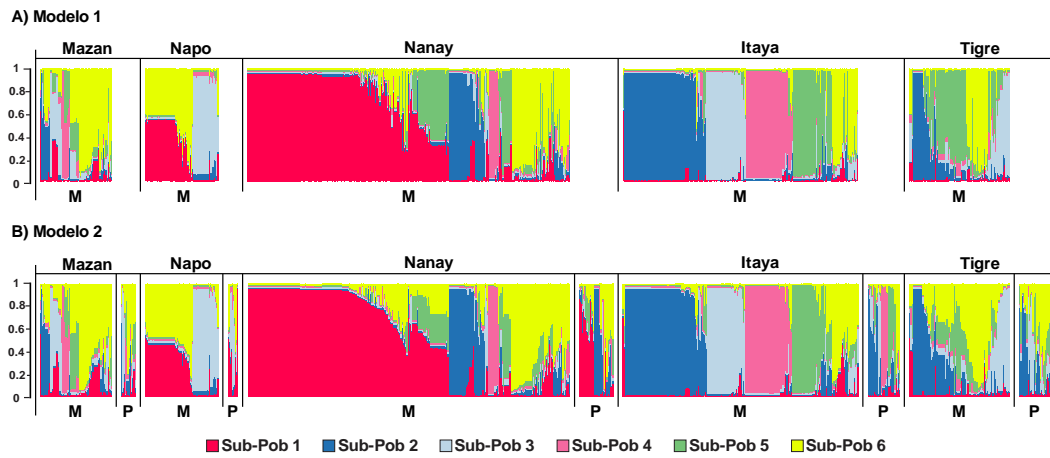
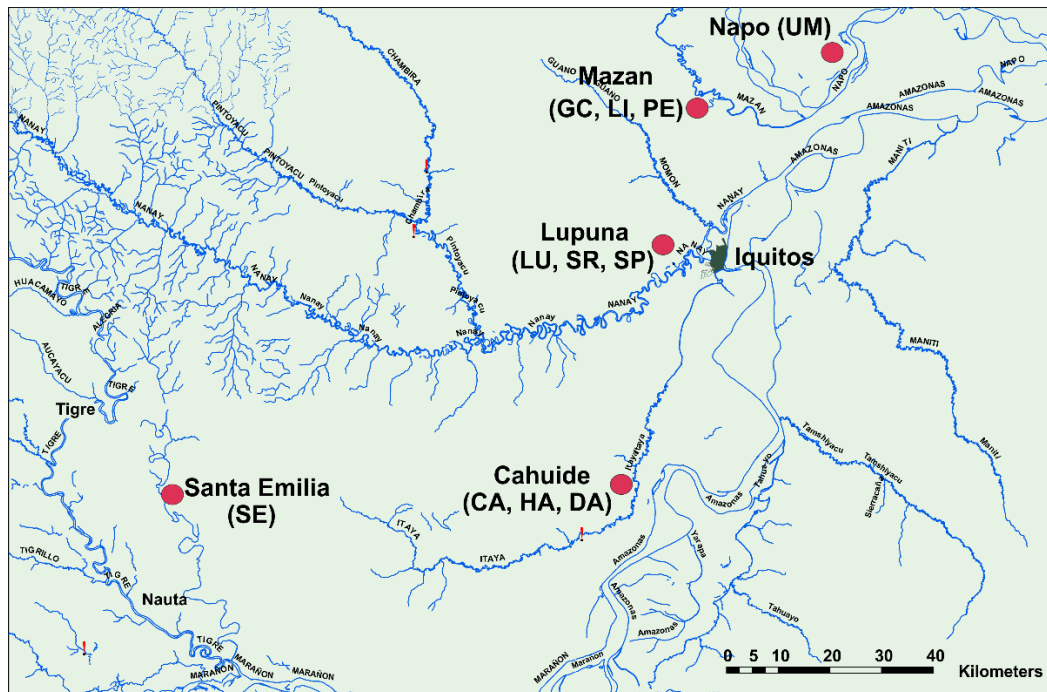


Figura suplementaria 5: Consistencia en el coeficiente de ancestría de cada infección de *Plasmodium vivax* asignado por A) el modelo de Ligamiento (Modelo 1) y B) el modelo USEPOPINFO (Modelo 2). Ambos modelos se compararon asumiendo $K = 6$. M: Monoclonales y policlonales en un solo loci, P: Policlonales en dos o más loci.

ANEXO 9



Áreas de estudio

1:560,198

Figura suplementaria 6: Distribución geográfica de las áreas de estudio utilizadas para la identificación de las infecciones foráneas dentro del distrito de Mazán. Las áreas de estudio de Lupuna, Cahuide y Santa Emilia están ubicadas en las cuencas de los ríos Nanay, Itaya y Tigre respectivamente.

ANEXO 10

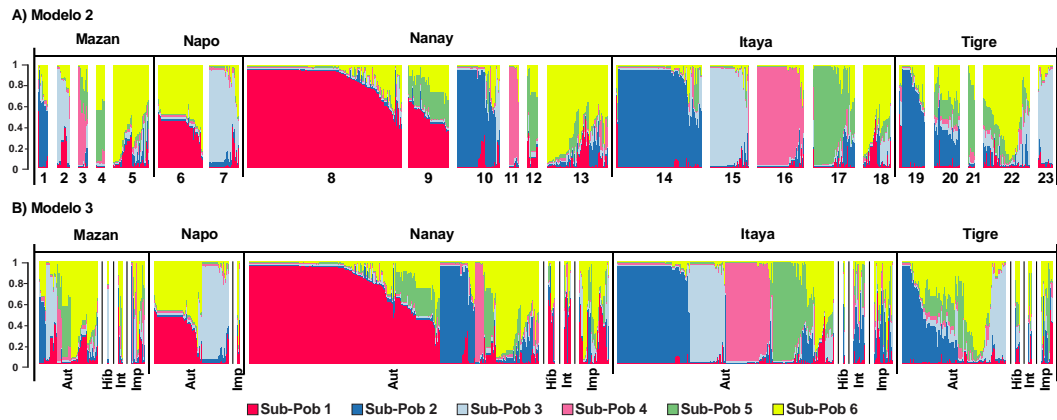


Figura suplementaria 7: A) Identificación y asignación de sub-poblaciones presentes en cada cuenca; el coeficiente de ancestría (Q) fue calculado usando el MODELO 2. B) El MODELO 3 corresponde a la identificación de infecciones Autóctonas, Híbridas, Introducidas e Importadas mediante el uso conjunto del lugar de muestreo, y la población ancestral definida en el modelo 2 (23 Haplogrupos asignados). El modelo 3 no calcula un coeficiente de ancestría (Q), por lo que el valor Q de cada infección mostrado fue calculado en el modelo 2.

ANEXO 11

Tabla suplementaria 4: Fuente u origen de las infecciones en cuatro cuencas de

Cuenca	Mazán		Napo		Nanay		Itaya		Tigre	
	# Indv	Frec. %	# Indv	Frec. %	# Indv	Frec. %	# Indv	Frec. %	# Indv	Frec. %
Autoctono	65	73.86	82	96.47	319	87.16	238	86.86	115	83.94
Híbrido	3	3.41	-	-	8	2.19	2	0.73	5	3.65
Introducido	5	5.68	-	-	7	1.91	13	4.74	4	2.92
Importado	15	17.05	3	3.53	32	8.74	21	7.66	13	9.49
Total	88		85		366		274		137	

la Amazonía Peruana

ANEXO 12

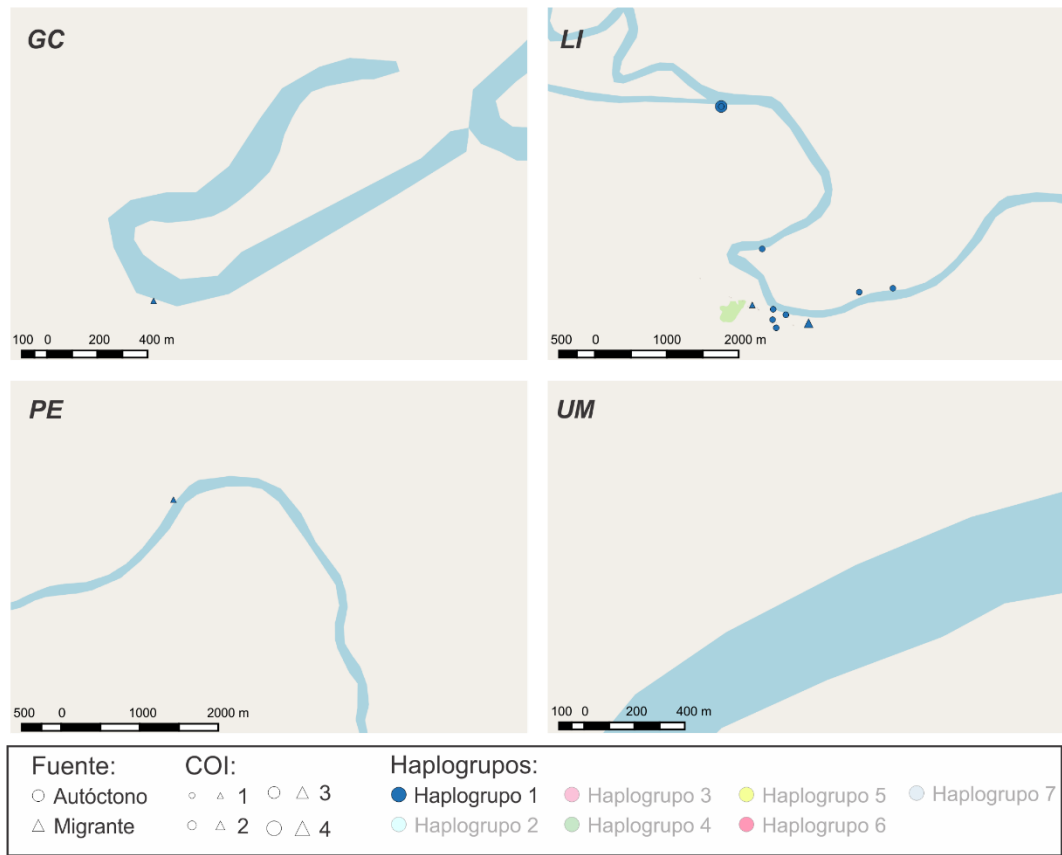


Figura suplementaria 8: Distribución geográfica de los casos autóctonos e infecciones policlonales causadas por el haplogrupo 1 en las cuatro comunidades estudiadas: Gamitanacocha (GC), Libertad (LI), Primero de Enero (PE) y Urco Miraño (UM). Las infecciones autóctonas causadas por este haplogrupo están representadas por puntos azules, mientras que las infecciones foráneas se representan con triángulos; y su tamaño depende de la complejidad de la infección.

ANEXO 13

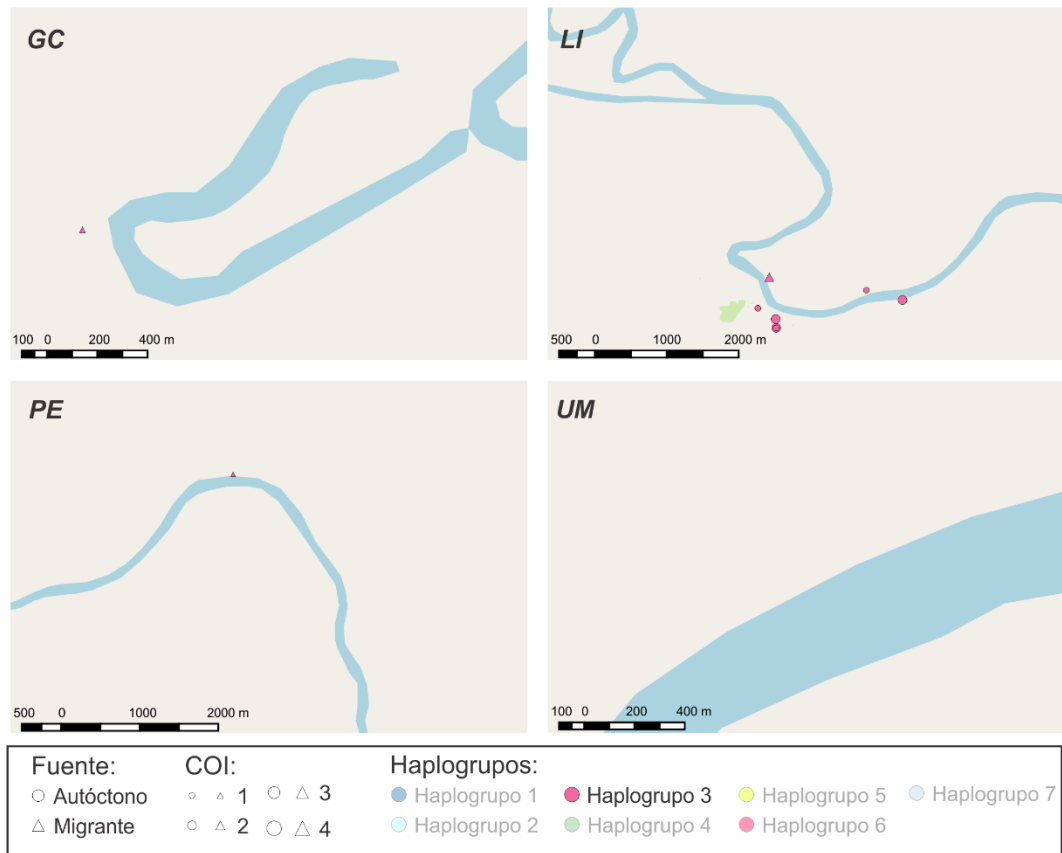


Figura suplementaria 9: Distribución geográfica de los casos autóctonos e infecciones policlonales causadas por el haplogrupo 3 en las cuatro comunidades estudiadas: Gamitanacocha (GC), Libertad (LI), Primero de Enero (PE) y Urco Miraño (UM). Las infecciones autóctonas causadas por este haplogrupo están representadas por puntos fucsias, mientras que las infecciones foráneas se representan con triángulos; y su tamaño depende de la complejidad de la infección.

ANEXO 14

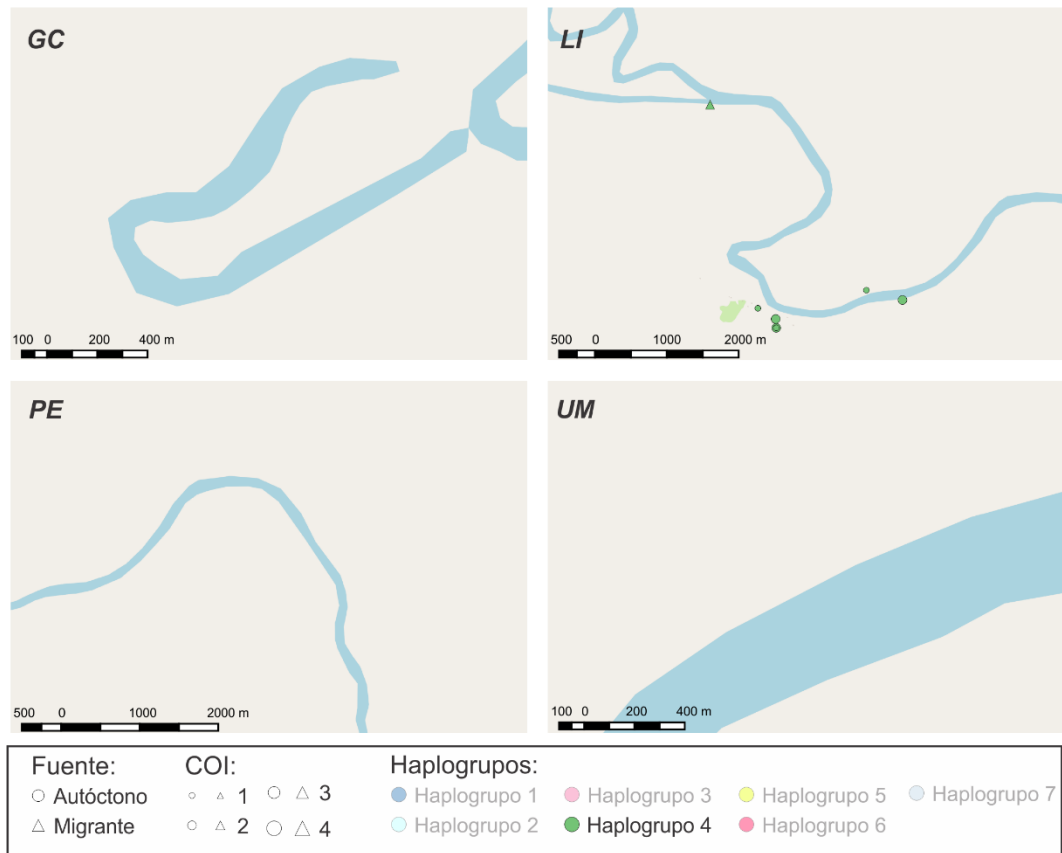


Figura suplementaria 10: Distribución geográfica de los casos autóctonos e infecciones policlonales causadas por el haplogrupo 4 en las cuatro comunidades estudiadas: Gamitanacocha (GC), Libertad (LI), Primero de Enero (PE) y Urco Miraño (UM). Las infecciones autóctonas causadas por este haplogrupo están representadas por puntos verdes, mientras que las infecciones foráneas se representan con triángulos; y su tamaño depende de la complejidad de la infección.

ANEXO 15

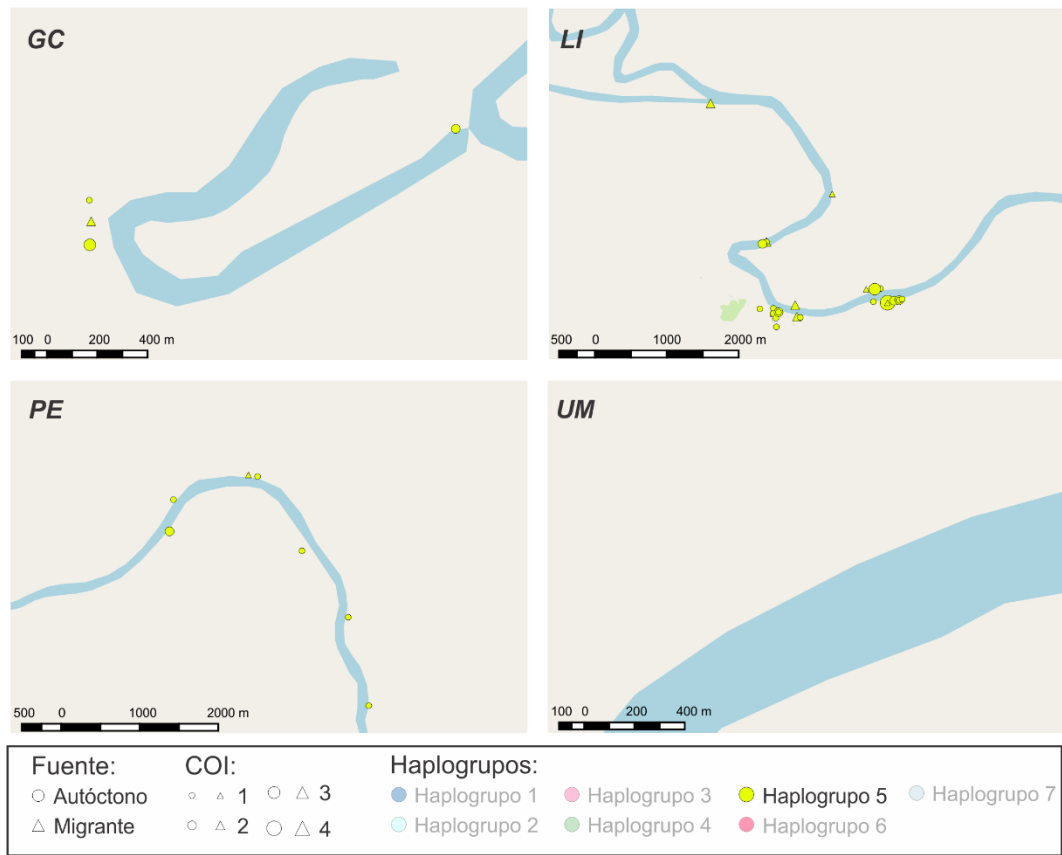


Figura suplementaria 11: Distribución geográfica de los casos autóctonos e infecciones policlonales causadas por el haplogrupo 5 en las cuatro comunidades estudiadas: Gamitanacocha (GC), Libertad (LI), Primero de Enero (PE) y Urco Miraño (UM). Las infecciones autóctonas causadas por este haplogrupo están representadas por puntos amarillos, mientras que las infecciones foráneas se representan con triángulos; y su tamaño depende de la complejidad de la infección.

ANEXO 16

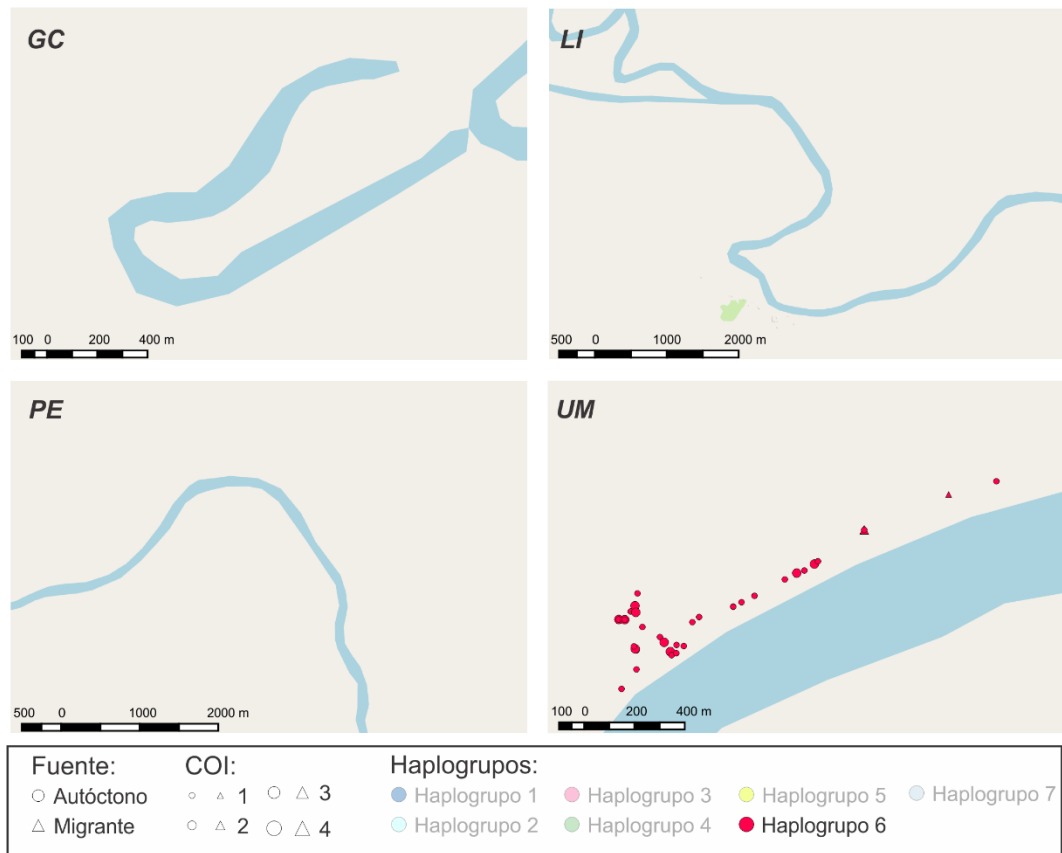


Figura suplementaria 12: Distribución geográfica de los casos autóctonos e infecciones policlonales causadas por el haplogrupo 6 en las cuatro comunidades estudiadas: Gamitanacocha (GC), Libertad (LI), Primero de Enero (PE) y Urco Miraño (UM). Las infecciones autóctonas causadas por este haplogrupo están representadas por puntos rojos, mientras que las infecciones foráneas se representan con triángulos; y su tamaño depende de la complejidad de la infección.

ANEXO 17

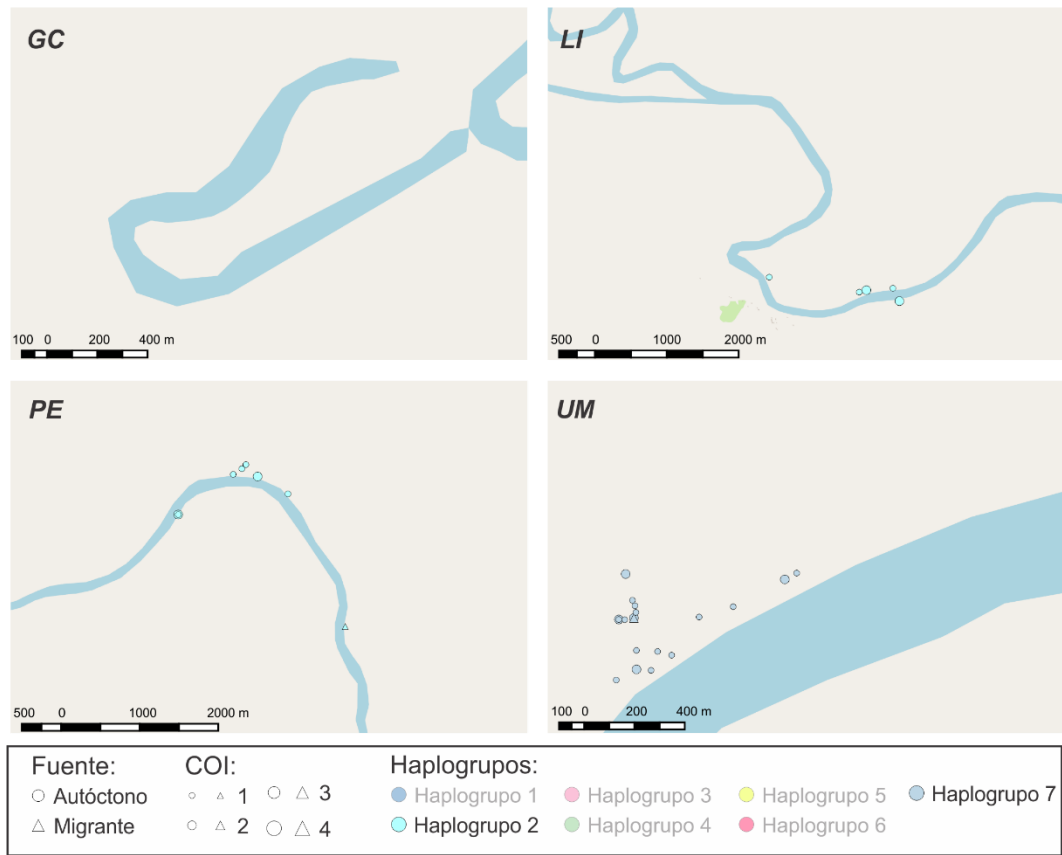


Figura suplementaria 13: Distribución geográfica de los casos autóctonos e infecciones policlonales causadas por los haplogrupos 2 y 7 en las cuatro comunidades estudiadas: Gamitanacocha (GC), Libertad (LI), Primero de Enero (PE) y Urco Miraño (UM). Las infecciones autóctonas causadas por este haplogrupo están representadas por puntos celestes, mientras que las infecciones foráneas se representan con triángulos; y su tamaño depende de la complejidad de la infección.

ANEXO 18

Tabla suplementaria 5: Modelos bivariados de regresión logística para la presencia del Haplogrupo 1 en la cuenca Mazán.

Factor de exposición	Modelo Bivariado			
	OR	p-valor	[IC 95%]	
Edad (Ref = < 7 años)				
7 - <12	1.31	0.796	0.17	10.35
12 - <18	2.63	0.373	0.31	21.92
> 18	1.64	0.575	0.29	9.25
Sexo (Ref = Femenino)				
Masculino	0.90	0.872	0.25	3.20
Ocupación (Ref = Menor de Edad)				
Estudiante	1.20	0.834	0.22	6.61
Agricultor	0.70	0.705	0.11	4.55
Otros	2.40	0.33	0.41	13.98
Agricultor en el último mes (Ref = No)				
Sí	0.74	0.669	0.18	3.02
Pescador en el último mes (Ref = No)				
Sí				
Maderero en el último mes (Ref = No)				
Sí				
Ocupación fuera de la comunidad (Ref = No)				
Sí	0.35	0.337	0.04	2.96
Agricultor en el hogar (Ref = No)				
Sí	0.73	0.621	0.20	2.59
Pescador en el hogar (Ref = No)				
Sí				
Maderero en el hogar (Ref = No)				
Sí	0.38	0.374	0.05	3.20
Viaje en el último mes (Ref = No)				
Sí	0.72	0.769	0.08	6.33
Viaje de algún miembro de la casa (Ref = No)				
Sí	0.75	0.723	0.15	3.78
Complejidad de Infección (Ref = 1)				
Policlonal	0.16	0.085	0.02	1.29
Sintomatología (Ref = No)				
Sí	3.65	0.055	0.97	13.67
Patencia (Ref = Sub-Patente)				
Patente	0.91	0.89	0.26	3.26
Parasitemia				
	0.97	0.714	1.00	1.00

ANEXO 19

Tabla suplementaria 6: Modelos bivariados y Multivariado de regresión logística para la presencia del Haplogrupo 2 en la cuenca Mazán.

Factor de exposición	Modelo Bivariado			Modelo Multivariado		
	OR	p-valor	[IC 95%]	OR	p-valor	[IC 95%]
Edad (Ref = < 7 años)						
7 - <12	4.04	0.124	0.68	23.94		
12 - <18	1.17	0.905	0.09	14.56		
> 18	2.03	0.412	0.37	11.05		
Sexo (Ref = Femenino)						
Masculino	3.28	0.062	0.94	11.42	2.98	0.10 0.83 10.70
Ocupación (Ref = Menor de Edad)						
Estudiante	2.22	0.315	0.47	10.54		
Agricultor	1.09	0.92	0.20	5.98		
Otros	2.40	0.33	0.41	13.98		
Agricultor en el último mes (Ref = No)						
Sí	0.50	0.323	0.13	1.97		
Pescador en el último mes (Ref = No)						
Sí						
Maderero en el último mes (Ref = No)						
Sí	5.38	0.244	0.32	91.64		
Ocupación fuera de la comunidad (Ref = No)						
Sí	2.61	0.132	0.75	9.07		
Agricultor en el hogar (Ref = No)						
Sí	0.81	0.724	0.25	2.58		
Pescador en el hogar (Ref = No)						
Sí						
Maderero en el hogar (Ref = No)						
Sí	2.87	0.1	0.82	10.08		
Viaje en el último mes (Ref = No)						
Sí	4.33	0.044	1.04	18.10		
Viaje de algún miembro de la casa (Ref = No)						
Sí	3.35	0.052	0.99	11.31	3.52	0.09 0.81 15.36
Complejidad de Infección (Ref = 1)						
Policlonal	0.70	0.571	0.20	2.44		
Sintomatología (Ref = No)						
Sí	2.78	0.086	0.86	8.96		
Patencia (Ref = Sub-Patente)						
Patente	2.17	0.224	0.62	7.58		
Parasitemia						
	1.02	0.735	1.00	1.00		

ANEXO 20

Tabla suplementaria 7: Modelos bivariados de regresión logística para la presencia del Haplogrupo 3 en la cuenca Mazán.

Factor de exposición	OR	p-valor	[IC 95%]	
Edad (Ref = 7 - 12 años)				
12 - <18	1.89	0.666	0.11	33.89
> 18	2.06	0.532	0.21	19.91
Sexo (Ref = Femenino)				
Masculino	0.53	0.471	0.09	3.03
Ocupación (Ref = Estudiante)				
Estudiante				
Agricultor	4.19	0.216	0.43	40.62
Otros	1.83	0.678	0.11	32.00
Agricultor en el último mes (Ref = No)				
Sí	1.38	0.73	0.22	8.80
Pescador en el último mes (Ref = No)				
Sí	6.42	0.141	0.54	76.34
Maderero en el último mes (Ref = No)				
Sí				
Ocupación fuera de la comunidad (Ref = No)				
Sí	2.06	0.426	0.35	12.27
Agricultor en el hogar (Ref = No)				
Sí				
Pescador en el hogar (Ref = No)				
Sí				
Maderero en el hogar (Ref = No)				
Sí	2.23	0.378	0.37	13.34
Viaje en el último mes (Ref = No)				
Sí	1.97	0.562	0.20	19.64
Viaje de algún miembro de la casa (Ref = No)				
Sí	2.47	0.343	0.38	15.99
Complejidad de Infección (Ref = 1)				
Policlonal	1.93	0.439	0.37	10.18
Sintomatología (Ref = No)				
Sí	0.42	0.443	0.04	3.90
Patencia (Ref = Sub-Patente)				
Patente	1.17	0.87	0.18	7.37
Parasitemia				
	1.00	0.969	1.00	1.00

ANEXO 21

Tabla suplementaria 8: Modelos bivariados de regresión logística para la presencia del Haplogrupo 4 en la cuenca Mazán.

Factor de exposición	OR	p-valor	[IC 95%]	
Edad (Ref = < 7 años)				
7 - <12	1.90	0.443	0.37	9.87
12 - <18				
> 18	0.38	0.312	0.06	2.48
Sexo (Ref = Femenino)				
Masculino	1.42	0.621	0.35	5.68
Ocupación (Ref = Menor de Edad)				
Estudiante	0.86	0.857	0.17	4.33
Agricultor	0.24	0.216	0.02	2.31
Otros	0.48	0.531	0.05	4.78
Agricultor en el último mes (Ref = No)				
Sí	0.26	0.225	0.03	2.26
Pescador en el último mes (Ref = No)				
Sí				
Maderero en el último mes (Ref = No)				
Sí				
Ocupación fuera de la comunidad (Ref = No)				
Sí	0.46	0.473	0.05	3.90
Agricultor en el hogar (Ref = No)				
Sí	0.77	0.706	0.19	3.08
Pescador en el hogar (Ref = No)				
Sí				
Maderero en el hogar (Ref = No)				
Sí	0.49	0.518	0.06	4.23
Viaje en el último mes (Ref = No)				
Sí	1.08	0.946	0.12	9.82
Viaje de algún miembro de la casa (Ref = No)				
Sí	0.47	0.491	0.05	4.06
Complejidad de Infección (Ref = 1)				
Policlonal	0.91	0.9	0.21	3.92
Sintomatología (Ref = No)				
Sí				
Patencia (Ref = Sub-Patente)				
Patente	1.32	0.718	0.29	5.91
Parasitemia				
	1.03	0.528	1.00	1.00

ANEXO 22

Tabla suplementaria 9: Modelos bivariados de regresión logística para la presencia del Haplogrupo 5 en la cuenca Mazán.

Factor de exposición	OR	p-valor	[IC 95%]	
Edad (Ref = < 7 años)				
7 - <12	0.60	0.449	0.16	2.26
12 - <18	1.56	0.563	0.35	6.94
> 18	0.95	0.92	0.33	2.76
Sexo (Ref = Femenino)				
Masculino	1.05	0.912	0.44	2.49
Ocupación (Ref = Menor de Edad)				
Estudiante	1.09	0.879	0.35	3.43
Agricultor	1.13	0.826	0.37	3.47
Otros	0.76	0.697	0.18	3.10
Agricultor en el último mes (Ref = No)				
Sí	1.73	0.244	0.69	4.37
Pescador en el último mes (Ref = No)				
Sí	1.70	0.605	0.23	12.70
Maderero en el último mes (Ref = No)				
Sí	1.68	0.718	0.10	27.79
Ocupación fuera de la comunidad (Ref = No)				
Sí	0.58	0.343	0.19	1.80
Agricultor en el hogar (Ref = No)				
Sí	1.17	0.735	0.48	2.84
Pescador en el hogar (Ref = No)				
Sí				
Maderero en el hogar (Ref = No)				
Sí	0.45	0.193	0.13	1.50
Viaje en el último mes (Ref = No)				
Sí	0.38	0.234	0.07	1.89
Viaje de algún miembro de la casa (Ref = No)				
Sí	0.52	0.252	0.17	1.60
Complejidad de Infección (Ref = 1)				
Policlonal	1.34	0.527	0.54	3.28
Sintomatología (Ref = No)				
Sí	0.86	0.755	0.35	2.16
Patencia (Ref = Sub-Patente)				
Patente	0.53	0.168	0.22	1.30
Parasitemia				
	1.01	0.851	1.00	1.00

ANEXO 23

Tabla suplementaria 10: Modelos bivariados y Multivariado de regresión logística para la presencia de infecciones importadas en la cuenca Mazán.

Factor de exposición	Modelo Bivariado				Modelo Multivariado			
	OR	p-valor	[IC 95%]		OR	p-valor	[IC 95%]	
Edad (Ref = < 7 años)								
7 - <12	0.13	0.074	0.01	1.22	0.15	0.09	0.02	1.36
12 - <18	0.25	0.232	0.03	2.41	0.28	0.28	0.03	2.74
> 18	0.44	0.2	0.13	1.54	0.52	0.31	0.14	1.86
Sexo (Ref = Femenino)								
Masculino	0.33	0.082	0.10	1.15	0.37	0.12	0.10	1.30
Ocupación (Ref = Menor de Edad)								
Estudiante	0.27	0.13	0.05	1.47				
Agricultor	0.71	0.613	0.19	2.63				
Otros	0.24	0.204	0.03	2.18				
Agricultor en el último mes (Ref = No)								
Sí	1.45	0.523	0.46	4.59				
Pescador en el último mes (Ref = No)								
Sí	1.60	0.695	0.15	16.48				
Maderero en el último mes (Ref = No)								
Sí								
Ocupación fuera de la comunidad (Ref = No)								
Sí	1.53	0.515	0.42	5.53				
Agricultor en el hogar (Ref = No)								
Sí	0.67	0.485	0.22	2.06				
Pescador en el hogar (Ref = No)								
Sí								
Maderero en el hogar (Ref = No)								
Sí	1.68	0.432	0.46	6.11				
Viaje en el último mes (Ref = No)								
Sí	0.48	0.508	0.06	4.14				
Viaje de algún miembro de la casa (Ref = No)								
Sí	0.84	0.81	0.21	3.36				
Complejidad de Infección (Ref = 1)								
Policlonal	1.79	0.312	0.58	5.51				
Sintomatología (Ref = No)								
Sí	0.58	0.388	0.17	2.00				
Patencia (Ref = Sub-Patente)								
Patente	1.19	0.761	0.38	3.71				
Parasitemia								
	0.95	0.474	1.00	1.00				

ANEXO 24

Tabla suplementaria 11: Modelos bivariados de regresión logística para la presencia del Haplogrupo 6 en la cuenca Napo.

Factor de exposición	OR	p-valor	[IC 95%]	
Edad (Ref = < 7 años)				
7 - <12	0.55	0.384	0.15	2.09
12 - <18	0.42	0.179	0.12	1.48
> 18	0.57	0.395	0.16	2.07
Sexo (Ref = Femenino)				
Masculino	0.93	0.875	0.38	2.27
Ocupación (Ref = Menor de Edad)				
Estudiante	0.50	0.222	0.16	1.53
Agricultor	0.35	0.164	0.08	1.53
Otros	1.06	0.952	0.17	6.74
Agricultor en el último mes (Ref = No)				
Sí	1.38	0.619	0.39	4.92
Pescador en el último mes (Ref = No)				
Sí				
Maderero en el último mes (Ref = No)				
Sí				
Ocupación fuera de la comunidad (Ref = No)				
Sí	3.23	0.295	0.36	28.97
Agricultor en el hogar (Ref = No)				
Sí	1.67	0.285	0.65	4.28
Pescador en el hogar (Ref = No)				
Sí	3.23	0.295	0.36	28.97
Maderero en el hogar (Ref = No)				
Sí				
Viaje en el último mes (Ref = No)				
Sí	0.58	0.701	0.03	9.56
Viaje de algún miembro de la casa (Ref = No)				
Sí	1.09	0.878	0.36	3.32
Complejidad de Infección (Ref = 1)				
Policlonal	0.73	0.577	0.24	2.20
Sintomatología (Ref = No)				
Sí	0.86	0.744	0.35	2.11
Patencia (Ref = Sub-Patente)				
Patente	1.90	0.169	0.76	4.77
Parasitemia				
	1.11	0.133	1.00	1.00

ANEXO 25

Tabla suplementaria 12: Modelos bivariados de regresión logística para la presencia de la del Haplogrupo 7 en la cuenca Napo.

Factor de exposición	OR	p-valor	[IC 95%]	
Edad (Ref = < 7 años)				
7 - <12	3.02	0.131	0.72	12.70
12 - <18	3.96	0.048	1.01	15.52
> 18	2.37	0.229	0.58	9.72
Sexo (Ref = Femenino)				
Masculino	1.16	0.756	0.46	2.88
Ocupación (Ref = Menor de Edad)				
Estudiante	3.36	0.056	0.97	11.68
Agricultor	3.39	0.128	0.70	16.38
Otros	1.58	0.641	0.23	10.90
Agricultor en el último mes (Ref = No)				
Sí	0.55	0.398	0.14	2.19
Pescador en el último mes (Ref = No)				
Sí				
Maderero en el último mes (Ref = No)				
Sí				
Ocupación fuera de la comunidad (Ref = No)				
Sí	0.36	0.367	0.04	3.27
Agricultor en el hogar (Ref = No)				
Sí	0.60	0.292	0.23	1.56
Pescador en el hogar (Ref = No)				
Sí	0.36	0.367	0.04	3.27
Maderero en el hogar (Ref = No)				
Sí				
Viaje en el último mes (Ref = No)				
Sí	2.04	0.62	0.12	33.83
Viaje de algún miembro de la casa (Ref = No)				
Sí	1.12	0.848	0.36	3.41
Complejidad de Infección (Ref = 1)				
Policlonal	0.85	0.788	0.27	2.74
Sintomatología (Ref = No)				
Sí	1.24	0.64	0.50	3.10
Patencia (Ref = Sub-Patente)				
Patente	0.63	0.336	0.25	1.61
Parasitemia				
	0.89	0.141	1.00	1.00

ANEXO 26

Tabla suplementaria 13: Modelos bivariados y multivariado de regresión logística para la presencia de Sintomatología en el distrito de Mazán.

Factor de exposición	Modelo Bivariado				Modelo Multivariado			
	OR	p-valor	[IC 95%]		OR	p-valor	[IC 95%]	
Patencia (Ref = Sub-Patente)	2.01	0.04	1.05	3.86	2.35	0.01	1.19	4.66
Parasitemia (1/1000)	1.01	0.55	0.98	1.04				
Clonalidad de la infección (Ref = Monoclonal)	0.60	0.17	0.29	1.25				
Complejidad de infección (Ref = 1)								
2	0.62	0.21	0.29	1.31				
3	0.67	0.75	0.06	7.58				
Haplogrupo 1 (Ref = No)	2.86	0.11	0.80	10.18				
Haplogrupo 2 (Ref = No)	2.17	0.17	0.72	6.56				
Haplogrupo 3 (Ref = No)	0.37	0.38	0.04	3.40				
Haplogrupo 4 (Ref = No)								
Haplogrupo 5 (Ref = No)	0.76	0.50	0.34	1.69				
Haplogrupo 6 (Ref = No)	1.12	0.74	0.58	2.17				
Haplogrupo 7 (Ref = No)	1.40	0.42	0.62	3.16				
Infecciones Importadas (Ref = No)	0.55	0.28	0.19	1.63				
Edad (Ref = < 7 años)								
7 - <12	0.55	0.20	0.22	1.38				
12 - <18	0.69	0.44	0.27	1.76				
> 18	0.98	0.95	0.44	2.15				
Sexo (Ref = Femenino)	0.95	0.87	0.51	1.76				
Ocupación (Ref = Menor de Edad)								
Estudiante	0.64	0.26	0.30	1.38				
Agricultor	0.69	0.42	0.29	1.68				
Otros	1.23	0.70	0.43	3.48				
Agricultor en el último mes (Ref = No)	1.11	0.78	0.54	2.26				
Pescador en el último mes (Ref = No)	1.54	0.67	0.21	11.20				
Maderero en el último mes (Ref = No)								
Ocupación fuera de la comunidad (Ref = No)	1.80	0.19	0.75	4.37				
Agricultor en el hogar (Ref = No)	0.71	0.29	0.37	1.34				
Pescador en el hogar (Ref = No)	1.15	0.86	0.25	5.30				
Maderero en el hogar (Ref = No)	2.11	0.16	0.74	5.96				
Viaje en el último mes (Ref = No)	2.26	0.18	0.69	7.45				
Viaje de algún miembro de la casa (Ref = No)	2.29	0.03	1.08	4.83	2.74	0.01	1.25	6.00
Malaria en el último año (Ref = No)	1.32	0.39	0.70	2.49				
Malaria en su vida (Ref = No)	0.91	0.91	0.20	4.26				

ANEXO 28

Tabla suplementaria 15: Modelos bivariados de regresión logística para la presencia de Sintomatología en la cuenca Napo.

Factor de exposición	Modelo Bivariado			
	OR	p-valor	[IC 95%]	
Patencia (Ref = Sub-Patente)	5.00	0.002	1.76	14.19
Parasitemia (1/1000)	1.00	0.92	0.97	1.03
Clonalidad de la infección (Ref = Monoclonal)	1.05	0.94	0.35	3.14
Complejidad de infección (Ref = 1)				
2	1.05	0.94	0.35	3.14
3				
Haplogrupo 1 (Ref = No)				
Haplogrupo 2 (Ref = No)				
Haplogrupo 3 (Ref = No)				
Haplogrupo 4 (Ref = No)				
Haplogrupo 5 (Ref = No)				
Haplogrupo 6 (Ref = No)	0.86	0.74	0.35	2.11
Haplogrupo 7 (Ref = No)	1.24	0.64	0.50	3.10
Infecciones Importadas (Ref = No)				
Edad (Ref = < 7 años)				
7 - <12	0.46	0.23	0.13	1.64
12 - <18	0.52	0.29	0.16	1.73
> 18	0.83	0.76	0.25	2.72
Sexo (Ref = Femenino)	0.57	0.21	0.23	1.37
Ocupación (Ref = Menor de Edad)				
Estudiante	0.53	0.23	0.19	1.49
Agricultor	0.65	0.56	0.16	2.68
Otros	0.92	0.92	0.18	4.58
Agricultor en el último mes (Ref = No)	0.81	0.73	0.24	2.71
Pescador en el último mes (Ref = No)				
Maderero en el último mes (Ref = No)				
Ocupación fuera de la comunidad (Ref = No)	1.36	0.72	0.26	7.19
Agricultor en el hogar (Ref = No)	0.52	0.18	0.21	1.34
Pescador en el hogar (Ref = No)	1.36	0.72	0.26	7.19
Maderero en el hogar (Ref = No)				
Viaje en el último mes (Ref = No)				
Viaje de algún miembro de la casa (Ref = No)	1.67	0.35	0.57	4.86
Malaria en el último año (Ref = No)	1.30	0.55	0.55	3.12
Malaria en su vida (Ref = No)	1.79	0.52	0.30	10.60