



CONECTIVIDAD DE POBLACIONES DE
Plasmodium vivax EN LAS REGIONES DE
AMAZONAS Y LORETO

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

JHON ERIC ZUMAETA MARIN

LIMA – PERÚ

2026

ASESORA

Dra. Dionicia Baziliza Gamboa Vilela

CO ASESORES

Dr. Luis Esteban Cabrera Sosa

Dra. Stella Maris Chenet Carrasco

JURADO DE TESIS

DR. JOSE RONALD ESPINOZA BABILON

PRESIDENTE

DRA. MARIA LISSETH EGUILUZ MOYA

VOCAL

MG. IRENE ROSA MARIA DELGADO DE LA FLOR MONTAUBAN

SECRETARIA

DEDICATORIA

A mi familia, que con su ternura y comprensión me apoyaron incluso cuando yo era difícil de entender. En especial a mi hermana mayor Elizabeth, nada de esto no hubiese sido posible sin ella. Te amo hermana, de aquí al cielo, ida y vuelta, miles de veces.

A mi sobrino Ian. Su inocencia y curiosidad por el mundo me impulsó a recordar mis sueños de niño. Querido sobrino, nunca dejar de soñar.

A mis mejores amigos, Roosevelt y Jenniffer. Si la casualidad existe, soy el humano más afortunado por haberlos conocido. A ustedes, los quiero y los extraño.

A mis amigos del club ZC. Brigitte, Tatiana y Paulito. Atesoro cada momento con ustedes. Gracias por ser mi lugar seguro.

A Molly. Mi hermosa caniche, una curita en el alma.

AGRADECIMIENTOS

A mis asesores, quienes me guiaron y me enseñaron hacer ciencia. La Dra. Stella Chenet por contagiarme su pasión por el estudio de la malaria, por confiar en aquel pequeño investigador que había en mí y enseñarme que existe un mundo más grande que Chachapoyas. Al Dr. Luis Cabrera, por sus valiosos consejos, su paciencia y su amistad, por enseñarme que el camino en ciencia nunca es una línea recta y hacerme entender que tengo lo necesario para seguirla. A la Dra. Dionicia Gamboa, por confiar en mí, acogerme en su grupo de investigación y darme aliento cuando estaba apunto de rendirme; muchas gracias desde el fondo de mi corazón. Ustedes representan mis pilares formativos en esta larga trayectoria.

A mis amigos del grupo Malaria de la Universidad Peruana Cayetano Heredia. Pamela, Gerardo, Daniela, Berónica y todos los demás de integrantes. Su compañía, su confianza y nuestras discusiones científicas me ayudó a crecer como profesional; pero su amistad, a crecer como persona. También, de manera especial, a mi amiga de laboratorio, clases y vida, Diana Cubas. El destino definitivamente nos unió y marcaste un hito importante en mi historia, cuando sentí que todo estaba perdido, me ayudaste a recuperar el sentido.

Al Dr. Jorge Arévalo, por despertar mi sentido crítico e impulsarme a superar los límites. A la Dra. Katherine Torres, por confiar en mi capacidad y darme la oportunidad de aprender de nuevos retos en investigación.

A la Dra. Eline Kattenberg, por estar atenta a todas las preguntas que me surgían a lo largo de la ejecución de este trabajo y su plena disposición por guiarme a resolverlas.

FUENTES DE FINANCIAMIENTO.

La bases de datos no publicadas utilizadas en esta tesis fueron secuenciados bajo el financiamiento del proyecto FA5 “Better polices and less infectious diseases in Peru and Latin America”.

Esta investigación fue financiada con un aporte de las becas Milla - Goñi de la Facultad de Ciencias e Ingeniería.



UNIVERSIDAD PERUANA
CAYETANO HEREDIA

DECLARACIÓN DE ORIGINALIDAD

Los egresados:

N°	APELLIDOS Y NOMBRES
1.	ZUMAETA MARIN JHON ERIC

Pertenecientes al programa de la **MAESTRÍA EN BIOQUÍMICA Y BIOLOGÍA MOLECULAR**, autores del trabajo titulado: **CONECTIVIDAD DE POBLACIONES DE *Plasmodium vivax* EN LAS REGIONES DE AMAZONAS Y LORETO**, el cual ha sido elaborado, sustentado y aprobado, según corresponda, para optar por el grado de **MAESTRO EN BIOQUÍMICA Y BIOLOGÍA MOLECULAR** bajo la modalidad de **TESIS**.

En calidad de docentes asesores de la Universidad Peruana Cayetano Heredia:

N°	APELLIDOS Y NOMBRES DEL DOCENTE	FACULTAD	NIVEL DE ASESORÍA
1.	GAMBOA VILELA DIONICIA BAZILIZA	FACI	ASESOR
2.	CABRERA SOSA LUIS ESTEBAN	FACI	CO ASESOR
3.	CHENET CARRASCO STELLA MARIS	FACI	CO ASESOR

Declaramos que el contenido del presente documento es original y que las citas y referencias a otros autores cumplen con las normas académicas establecidas. En ese sentido, hacemos constar que:

- El documento presenta un porcentaje de similitud de **4%**, según el reporte emitido por el software **Turnitin®** (identificador de entrega: **3014729304**; fecha de entrega: **19-08-2026**).
- Tras una revisión detallada del reporte y del contenido del trabajo en cuestión, no se han identificado indicios de plagio.
- Se certifica que el documento respeta los principios de integridad académica y cumple con los requisitos institucionales de originalidad.

Lugar y fecha: **Lima, 19 de agosto de 2026**

Firma del asesor
N° DNI: 07478767
ORCID: 0000-0002-1420-7729

Firma del Co-asesor
N° DNI: 73089368
ORCID: 0000-0002-8958-9709

Firma del Co-asesor
N° DNI: 41255427
ORCID: 0000-0002-5305-0664

ÍNDICE

RESUMEN

ABSTRACT

I.	INTRODUCCION	1
1.1.	Antecedentes	1
1.2.	Planteamiento del Problema	3
1.3.	Justificación	6
1.4.	Pregunta de investigación	8
II.	OBJETIVOS	9
2.1.	Objetivo general	9
2.2.	Objetivos específicos	9
III.	HIPÓTESIS	9
IV.	MARCO TEÓRICO	10
4.1.	Aspectos generales de la malaria	10
4.2.	Epidemiología de la malaria en el Perú	11
4.3.	Ciclo de vida de <i>Plasmodium</i>	17
4.4.	Características biológicas de <i>Plasmodium vivax</i>	19
4.5.	Factores que contribuyen a la transmisión de malaria en el Perú	21
4.6.	Diagnóstico y herramientas de vigilancia epidemiológica de malaria vivax	24
4.7.	Genética de poblaciones de <i>Plasmodium vivax</i> en el Perú	27
V.	METODOLOGÍA	34
5.1.	Tipo y diseño de Estudio	34
5.2.	Cálculo del número muestral del estudio	34
5.4.	Minería de datos	36
5.5.	Área de Estudio	42
5.6.	Análisis bioinformático	43

5.6.1. Obtención de las lecturas genómicas crudas-----	44
5.6.2. Filtrado y control de calidad de las lecturas genómicas crudas-----	44
5.6.3. Alineamiento y control de calidad del mapeo-----	45
5.6.4. Identificación, filtrado y selección de variantes-----	46
5.6.5. Estadísticas de los archivos de secuenciación-----	48
5.6.6. Análisis de Estructuración Genética-----	48
5.6.7. Análisis de Diversidad Genética-----	52
5.6.8. Análisis de Conectividad Genética-----	53
5.7. Análisis estadísticos -----	55
VI. RESULTADOS -----	57
6.1. Características de calidad y procedencia de las muestras filtradas-----	57
6.2. Estructuración Genética Poblacional -----	60
6.3. Estimaciones de la Diversidad Genética -----	80
6.4. Conectividad Genética-----	86
VII. DISCUSIÓN -----	96
7.1. Calidad de las secuencias de <i>P. vivax</i> -----	97
7.2. Estructuración Poblacional -----	98
7.3. Diversidad Genética-----	107
7.4. Conectividad Genética-----	110
7.5. Relevancia-----	116
7.5. Limitaciones -----	118
VIII. CONCLUSIONES -----	120
IX. RECOMENDACIONES -----	121
X. REFERENCIAS BIBLIOGRÁFICAS -----	123
XI. ANEXOS -----	

RESUMEN

La malaria, principalmente causada por *Plasmodium vivax*, es un problema de salud pública en Perú, especialmente en las regiones de Loreto y Amazonas. Las zonas rurales aisladas presentan mayores riesgos por limitaciones en diagnóstico y tratamiento. Loreto mantiene la mayor prevalencia, mientras que Amazonas presentó un incremento desde 2014, posiblemente por introducción desde Loreto. Actualmente existe limitada información sobre la interacción entre poblaciones de *P. vivax* en ambas regiones; la vigilancia molecular es clave para entender su dinámica y estructura poblacional.

Se analizó la variabilidad y flujo genético de *P. vivax* en las provincias endémicas de las regiones de Loreto (Maynas, Datem del Marañón y Ramón Castilla) y Amazonas (Río Santiago). Utilizando secuencias genómicas de nueva generación (NGS) y genoma completo (WGS) de bases de datos públicas y no publicadas (Loreto=334, Amazonas=68), se identificaron 802 polimorfismos de un solo nucleótido (SNP) bialélicos en los segmentos amplificados por el ensayo *Pv Ampliseq v2 Peru*, un panel NGS para el análisis de genética poblacional de *P. vivax* en Perú.

La población de *P. vivax* en Amazonas exhibió menor diversidad genética ($H_e=0,06$) que Loreto ($H_e=0,3$). Los valores de F_{ST} indicaron alta diferenciación entre las regiones ($F_{st}=0,27$), destacando Condorcanqui y Ramón Castilla con respecto a otras provincias ($F_{st}=0,3-0,73$). Los análisis de estructuración poblacional revelaron ambas provincias como grupos genéticos distintos. La conectividad genética de *P. vivax* es limitada entre Amazonas y Loreto, el bajo parentesco genético y valores reducidos de la cuantificación de la conectividad evidencian un flujo génico asimétrico, posiblemente por un evento de

introducción mediado por movilizaciones humanas desde Loreto hacia Amazonas, seguido de la persistencia de linajes clonales en zonas aisladas de transmisión residual. Estos resultados resaltan la importancia de implementar estrategias de control diferenciadas por región y la necesidad de una vigilancia molecular continua para la detección de brotes y propagación entre regiones.

PALABRAS CLAVE

Conectividad, malaria, *Plasmodium vivax*, SNP, Amazonas, Loreto.

ABSTRACT

Malaria, primarily caused by *Plasmodium vivax*, is a public health problem in Peru, especially in Loreto and Amazonas regions. Hard-to-reach rural areas face greater risks due to limitations in diagnosis and treatment. Loreto maintains the highest prevalence, while prevalence in Amazonas has shown an increase since 2014, possibly due to introduction from Loreto. Currently, there is limited information on the interaction between *P. vivax* populations in both regions; molecular surveillance is key to understand population dynamics and structure.

The variability and gene flow of *P. vivax* were analyzed in the endemic provinces of the Loreto (Maynas, Datem del Marañón, and Ramón Castilla) and Amazonas (Río Santiago) regions. Using next-generation sequencing (NGS) and whole-genome sequencing (WGS) from public and unpublished databases (Loreto=334, Amazonas=68), 802 biallelic single nucleotide polymorphisms (SNPs) were identified in segments amplified by the *Pv Ampliseq* v2 *Peru* assay, an NGS panel for population genetics analysis of *P. vivax* in Peru.

The *P. vivax* population in Amazonas exhibited low genetic diversity ($H_e=0.06$), compared to Loreto ($H_e=0.3$). F_{ST} values indicated high differentiation between regions ($F_{ST}=0.27$), with Condorcanqui and Ramón Castilla being the most differentiated to other provinces ($F_{ST}=0.3-0.73$). Population structure analyses revealed both provinces as distinct genetic groups. The genetic connectivity of *P. vivax* is limited between Amazonas and Loreto. The low genetic relatedness and reduced connectivity values indicated asymmetric gene flow, suggesting an introduction event likely mediated by human migrations from Loreto to Amazonas, followed by the persistence of clonal lineages in isolated areas of residual transmission. These results highlight the importance of implementing region-specific control

strategies and the need for continuous molecular surveillance to detect outbreaks and spread between regions.

KEYWORDS

Connectivity, malaria, *Plasmodium vivax*, SNP, Amazonas, Loreto.

I. INTRODUCCION

1.1. Antecedentes

La malaria es una de las enfermedades con mayor tasa de prevalencia en las regiones más pobres del mundo, tales como en América Latina. Dentro de este contexto, Perú está entre los primeros cinco países con mayor número de casos de malaria a nivel de Latinoamérica (WHO, 2025). Las infecciones están focalizadas en la selva peruana y alrededor del 85% son causadas por *Plasmodium vivax* (MINSA, 2026), el cual representa un problema de salud pública vigente.

En Perú, las regiones de Loreto y Amazonas albergan el mayor número de casos de malaria dentro de los últimos 5 años (MINSA, 2026). Las comunidades nativas localizadas en estas regiones son las poblaciones más vulnerables a la infección por malaria causada por *P. vivax* (malaria vivax) debido a diferentes factores demográficos y socioeconómicos (Montenegro et al., 2021).

Con el objetivo de controlar y eliminar la malaria, el Ministerio de Salud del Perú (MINSA) implementó políticas de contención como el programa “Control de la Malaria en las Zonas Fronterizas de la Región Andina: Un enfoque comunitario - PAMAFRO” (financiado por el Fondo Mundial de lucha contra el SIDA, tuberculosis y malaria) y el Plan Malaria Cero (PMC) con resultados muy exitosos, disminuyendo hasta el 80% y 75% de casos en los años 2010 y 2020, respectivamente (MINSA, 2022). Basados en el éxito de estos planes, el MINSA lanzó el Plan Nacional de Eliminación de la Malaria (PNEM) en el 2022 con el objetivo de bajar la incidencia en 90% para el 2030 (MINSA, 2022). Este plan se ha fortalecido con la nueva norma técnica de salud contra la malaria con el objetivo

de robustecer la atención a la malaria complicada y grave en el Perú a través del mejoramiento de las intervenciones sanitarias y consolidación de la vigilancia epidemiológica y molecular integrada (MINSA, 2025).

Dado que la movilidad en la Amazonía peruana depende principalmente del transporte fluvial (es decir, navegación en los ríos), este podría estar facilitando la dispersión de los casos de malaria hacia zonas rurales y de difícil acceso (Gomez et al., 2025). Así, en determinadas comunidades amazónicas, la transmisión de malaria aún persiste incluso tras la aplicación de intervenciones antimaláricas consideradas eficaces en regiones endémicas; este fenómeno es denominado como malaria residual (Rodriguez, 2021). Además, las diferencias socioculturales y económicas representan un desafío significativo para los programas de control, ya que limitan el diagnóstico temprano, el acceso a medicamentos, y con ello, dificulta la eliminación de la malaria (Rosas-Aguirre et al., 2016).

Entre las principales herramientas utilizadas para realizar la vigilancia epidemiológica tenemos, la microscopía y las pruebas de diagnóstico rápido. No obstante, estas estrategias solo permiten la obtención de datos sobre el número de casos, mas no de información correspondiente a la evolución genética de los parásitos (Lucchi et al., 2013). En contraste, la información molecular ha dilucidado cómo la diversidad genética está ligada a procesos de adaptación y conectividad entre áreas geográficas (Dalmat et al., 2019; Manrique et al., 2019).

Por ello, los nuevos enfoques de vigilancia molecular están orientadas a describir la dinámica de transmisión de malaria (patrones de transmisión de poblaciones parasitarias en tiempo y espacio) tanto en áreas periurbanas como en áreas rurales

de difícil acceso, lo cual permite evaluar, monitorear y afinar las estrategias de control (Delgado-Ratto et al., 2016). No obstante, hasta la fecha no hay información disponible sobre las relaciones entre poblaciones de *P. vivax* de Loreto y Amazonas.

El objetivo de esta tesis es dilucidar los flujos genéticos entre poblaciones parasitarias de *P. vivax* presentes en áreas endémicas de las regiones de Loreto y Amazonas. Para ello, se usó bases de datos genómicos de *P. vivax*, tanto públicos como privados, para profundizar la comprensión sobre la dinámica de transmisión (flujo genético) y la estructura poblacional del parásito entre ambas regiones. Así, se espera que los resultados de este trabajo contribuyan a establecer bases científicas robustas que permitan el desarrollo e implementación de estrategias hacia la eliminación de la malaria tales como la vigilancia epidemiológica y genómica, fortaleciendo su efectividad dentro de las políticas de salud públicas actuales.

1.2. Planteamiento del Problema

El Plan Nacional de Eliminación de la Malaria (PNEM) 2022-2030, actualmente en vigor en el Perú, mantiene un sistema de vigilancia focalizado en todas las regiones de la Amazonía peruana, con especial énfasis en la región de Loreto. Sin embargo, entre los mayores desafíos que enfrenta es el alcance de la cobertura de las estrategias incluidas en este plan. Estos son más notorios en las comunidades nativas de la Amazonía peruana, las cuales están entre las más afectadas por la malaria. Estas poblaciones permanecen vulnerables debido a la limitada accesibilidad geográfica y a la reducción en la implementación de las intervenciones contempladas en las estrategias de eliminación (Carrasco-Escobar et al., 2024).

La mayoría de las comunidades nativas se concentran geográficamente en las regiones de Loreto y Amazonas. La región de Amazonas alberga aproximadamente el 13,4% de comunidades nativas a nivel nacional. Específicamente, en la provincia de Condorcanqui, viven aproximadamente 235 comunidades nativas con poblaciones pertenecientes a las etnias Awajún y Wampis, la mayoría de esta última población ubicadas a lo largo del río Santiago. En la región de Loreto, se censaron el 43,2% de las comunidades nativas, incluyendo comunidades de 28 diferentes pueblos de origen étnico, incluyendo las Awajún y Wampis situadas en la provincia del Datem del Marañón, región fronteriza con Amazonas (INEI, 2018). La mayoría de estas son comunidades remotas de difícil entrada que poseen una limitada accesibilidad a servicios como agua, electricidad y atención sanitaria. Además de las diferencias culturales, los puntos de control de enfermedades en estas áreas son reducidos en comparación con zonas urbanas y periurbanas (Soto-Calle et al., 2017).

Históricamente, Loreto es la región con la mayor tasa de prevalencia de malaria a nivel nacional (MINSA, 2024). Por otro lado, en la región de Amazonas, los casos de malaria han aumentado significativamente desde el 2014, tras un periodo de silencio epidemiológico (periodo sin reportes de casos o con un número muy bajo de casos notificados) entre el 2010 y 2013 (Saavedra-Samillán et al., 2024), lo que sugiere una posible introducción de malaria desde una región más endémica a través de la frontera con el distrito de Morona, en Loreto (Montenegro et al., 2021). El 99% de los casos reportados desde el 2018 fueron registrados en la provincia de Condorcanqui, resaltando una focalización en comunidades nativas del distrito de Río Santiago (MINSA, 2026).

La situación aún se mantiene crítica para estas áreas, dado que las tasas de incidencias acumuladas de *P. vivax* oscilan entre alto y muy alto riesgo en el distrito de Río Santiago entre los años 2020 – 2024. Así mismo, se registra un aumento de riesgo de malaria vivax en los distritos adyacentes como Nieva y Bagua, en Amazonas (MINSA, 2024). En Loreto, las áreas con mayores tasas de incidencia se reportan en los distritos remotos de Yavarí, Andoas, Pastaza, Ramón Castilla y Trompeteros y el distrito periurbano de San Juan Bautista (MINSA, 2024). Esto confirma que las mayores tasas de incidencia de malaria vivax se acentúan en áreas de difícil acceso como comunidades remotas y regiones fronterizas, cuya transmisión y reintroducción de casos a menudo está ligado a procesos de movilidad humana (Carrasco-Escobar et al., 2024; Gomez et al., 2025).

En estas áreas geográficas, los ríos funcionan como los principales medios de transporte, permitiendo redes de conectividad entre comunidades con fines culturales, comerciales y/o laborales, aunque no se ha encontrado reportes de movilidad humana mediante ríos desde Loreto a Amazonas, o viceversa. Asimismo, en las comunidades asentadas a lo largo de una misma cuenca hidrográfica, estas mantienen vínculos que podrían facilitar el flujo genético de parásitos de malaria (Janko et al., 2024). En la Amazonía peruana, las poblaciones endémicas conectadas son conocidas como corredores de malaria, y se ha reportado la propagación de infecciones de malaria vivax en distancias considerablemente lejanas (~250 km) (Manrique et al., 2019).

Sin embargo, el conocimiento sobre la dinámica de expansión de *P. vivax* en la Amazonía peruana aun sigue siendo un panorama poco conocido, especialmente en lo que respecta a las comunidades de las regiones Amazonas y Loreto. Así mismo,

persiste la falta de información sobre los niveles de diversidad genética de las poblaciones de *P. vivax* en Amazonas y sobre la magnitud del flujo genético con las poblaciones de Loreto.

En este sentido surge la pregunta: ¿Cuáles son los niveles y patrones de flujo genético entre las áreas endémicas de malaria vivax situados en Amazonas y Loreto?

1.3. Justificación

El PNEM tienen como meta reducir la incidencia entre un 70% a 90% hasta el 2030 (MINSA, 2022). Pese a los esfuerzos de eliminación de malaria, paradójicamente el número de casos de malaria ha aumentado de manera alarmante dentro de los últimos dos años en la Amazonía peruana, y las incidencias de los casos siguen concentrándose principalmente en comunidades nativas de difícil acceso de las regiones de Amazonas y Loreto, con variaciones espacio-temporales que dificultan las estrategias focalizadas y eliminación (MINSA, 2026).

Este aumento está asociado no solo a la baja cobertura de las intervenciones del PNEM, sino también a la presencia de infecciones submicroscópicas y asintomáticos, las mismas que escapan de los métodos de diagnóstico convencionales y mantienen la transmisión “silenciosa” a través de recurrencias y bajas parasitemias (García Castillo et al., 2024).

La persistencia de estos reservorios de malaria plantea la necesidad de comprender la diversidad y estructura genética de *P. vivax* entre ambas regiones, lo que permitiría identificar el grado de conectividad entre sus poblaciones parasitarias y

aportar evidencia clave para responder la pregunta de investigación planteada.

Dado que el nivel de transmisión de la malaria en el país es heterogéneo (Pinedo-Cancino et al., 2024), las estrategias actuales podrían no ser efectivas en comunidades remotas, resaltando la necesidad de plantear enfoques adaptados a las características y necesidades de cada territorio. Por ello, la diferencia en los niveles de endemidad de Amazonas y Loreto constituyen un factor clave la cual es necesario tener en cuenta para el monitoreo mediante estudios genéticos necesarios para la implementación y fortalecimiento de los puntos vigilancia de la malaria.

La información derivada de la genotipificación y la estructuración genética obtenida de *P. vivax* puede explicar los procesos evolutivos del parásito presente en comunidades remotas de la Amazonía peruana, identificar cepas resistentes a antimaláricos y describir patrones de la dinámica de transmisión, contribuyendo a esclarecer el panorama genético en las regiones de Loreto y Amazonas. Así, este enfoque permitiría dilucidar los niveles de flujo genético y evaluar si los procesos de diversificación de los parásitos están influenciados por la ubicación geográfica y su variación temporal (Auburn et al., 2021; Brown et al., 2021).

La elaboración de redes y mapas de estructura poblacional que representen la relación genética de los parásitos en regiones específicas permite no solo identificar su origen, sino también trazar las rutas de movimiento poblacional. Esto resulta clave para evaluar la probabilidad de una eliminación efectiva en un área determinada, considerando la diversidad parasitaria, la movilidad de las poblaciones y la carga de casos importados (Arnott et al., 2012).

Conocer los niveles de flujo genético entre los distritos endémicos de estas regiones

es importante desde dos perspectivas diferentes y complementarias. Por un lado, desde la perspectiva científica, aportando evidencia fundamental para comprender la dinámica poblacional y evolutiva de *P. vivax*; y desde el ámbito de salud pública, proporcionando insumos para el diseño de estrategias más eficaces de eliminación, considerando contextos epidemiológicos diferentes, las comunidades periurbanas y las rurales de difícil acceso (Fernandez-Miñope et al., 2021).

1.4. Pregunta de investigación

En este sentido, considerando el vacío científico entre la conectividad de las poblaciones parasitarias entre ambas regiones endémicas, surge la pregunta sobre cuáles son los niveles y patrones de flujo genético de *P. vivax* entre las áreas endémicas de las regiones Amazonas y Loreto, a partir de los datos genómicos disponibles.

II. OBJETIVOS

2.1. Objetivo general

- Analizar los niveles y patrones de flujo genético de poblaciones de *P. vivax* entre las áreas endémicas de Amazonas y Loreto en Perú, utilizando datos genómicos disponibles.

2.2. Objetivos específicos

- Analizar la estructura genética interpoblacional de las poblaciones de *P. vivax* de las regiones de Amazonas y Loreto.
- Analizar la estructura genética intrapoblacional de las poblaciones de *P. vivax* de las regiones de Amazonas y Loreto.
- Determinar la diversidad genética de las poblaciones de *P. vivax* de las regiones de Amazonas y Loreto.
- Determinar las redes de conectividad de las poblaciones de *P. vivax* de las regiones de Amazonas y Loreto.

III. HIPÓTESIS

H₀: Las poblaciones de *P. vivax* de las áreas endémicas de las regiones de Loreto y Amazonas no muestran estructura genética significativa, indicando alto flujo genético.

H₁: Las poblaciones de *P. vivax* de las áreas endémicas de las regiones de Loreto y Amazonas muestran estructura genética significativa, indicando flujo genético limitado.

IV. MARCO TEÓRICO

4.1. Aspectos generales de la malaria

La malaria es una enfermedad metaxénica causada por parásitos del género *Plasmodium*: *P. vivax*, *P. falciparum*, *P. malarie*, *P. ovale* y *P. knowlesi*, los cuatro primeros de infección natural a humanos y el último zoonótico proveniente de macacos reportados en el sudeste asiático (Sato, 2021). La transmisión de la malaria es llevada a cabo únicamente por picaduras de mosquitos femeninos del género *Anopheles* infectados con parásitos, los cuales actúan como vectores altamente capaces y competentes (Cohuet et al., 2010) .

La malaria es una de las enfermedades con mayor impacto a nivel mundial, principalmente en regiones tropicales y subtropicales, el cual comprende 85 países endémicos de medio y bajos recursos económicos. Se estimaron 282 millones de casos de malaria globalmente en el 2024, el 93.9% de casos está focalizado en países africanos tales como Nigeria, República Democrática del Congo, Uganda, Etiopía y Mozambique. En el contexto de la región de las Américas, se registraron 663 mil casos de malaria en el año 2024, los países como Venezuela, Brasil y Colombia representaron el 75% del total de casos de malaria en dicha región (WHO, 2025).

El mayor número de infecciones de malaria a nivel mundial es causado por *P. falciparum*, representando aproximadamente el 96% del total de casos globales y más del 90% de muertes causadas en el 2024. El continente de africano es el más afectado por esta especie, que notificó aproximadamente el 95% de muertes globales en ese año, de las cuales el 75% correspondieron a menores de 5 años

(WHO, 2025). La importancia en la salud pública de *P. falciparum* en África radica en los diferentes espectros de intensidad de transmisión, altas tasas de mortalidad y morbilidad principalmente en países de bajos recursos económicos que poseen limitaciones en el diagnóstico, vigilancia y tratamiento por los sistemas de salud (Bhatt et al., 2015; Griffin et al., 2014). Aunque los casos de *P. falciparum* se están reduciendo en el transcurso de los años, en determinadas regiones se observa un aparente desplazamiento por *P. vivax* como especie predominante.

P. vivax es considerada como una infección benigna y es la segunda especie que causa con más frecuencia malaria a nivel global, con aproximadamente el 3.5% de casos reportados en el 2024 (WHO, 2025). Las infecciones por malaria vivax se originaron en el Sudeste Asiático y Pacífico Occidental; sin embargo, se reportan aumentos alarmantes de casos en África y Sudamérica, siendo más notorio los impactos económicos y severidad de la enfermedad (Price et al., 2007). En efecto, *P. vivax* posee una prevalencia del 66% del total de casos en el 2024 en la región de las Américas, siendo los países de Brasil, Venezuela, Colombia, Guyana y Perú los más endémicos (WHO, 2025). Aunque se han reportado reducciones significativas para Brasil y Colombia como resultado de las aplicaciones efectivas de políticas de control y eliminación, Perú resaltó por el aumento de los registros de malaria, especialmente en la región de Loreto en los dos últimos años.

4.2. Epidemiología de la malaria en el Perú

A nivel internacional, Perú es considerado como un entorno de baja endemicidad de malaria; sin embargo, se han observado fluctuaciones en el número de casos en los últimos años, con 22.6 mil casos en el 2022, 33.9 mil en 2024 y 33.5 mil en 2025.

Aproximadamente, el 99% de estos casos fueron reportados en comunidades rurales ubicadas en la Amazonía peruana (MINSA, 2026), con acceso limitado a servicios básicos como luz, agua, y salud (Rosas-Aguirre et al., 2016). Históricamente, la región de Loreto presenta el mayor número de casos de malaria, albergando entre el 80 y 90% de los casos notificados cada año. Por otro lado, las regiones de Junín y Amazonas registraron cifras considerablemente menores, con contribuciones promedio en los últimos 6 años de 4.1% y 4.4% respectivamente (Tabla 1).

Tabla 1. Casos anuales de malaria notificados en las regiones del Perú desde el 2020 al 2025. Adaptado de la Sala Situacional de Malaria, Centro Nacional de Epidemiología, Prevención y Control de Enfermedades (CDC-Perú), Ministerio de Salud (https://app7.dge.gob.pe/maps2/shiny_metaxenicas_web/)

Región	Año					
	2020	2021	2022	2023	2024	2025
LORETO	13397	15490	22725	20606	32261	33184
JUNÍN	517	1522	2292	408	748	143
AMAZONAS	1553	659	1657	1095	573	83
Otras regiones	343	404	384	508	322	114
Total	15810	18075	27058	22617	33904	33524

Las infecciones por *P. vivax* representan aproximadamente el 84.4% del total de casos en el país, en comparación con *P. falciparum* que representa alrededor del 15.6% según el reporte epidemiológico de la semana 52 en el 2025 (MINSA, 2026). Pese a las altas tasas de incidencia en distritos endémicos de esas regiones, es necesario apuntar que la transmisión es heterogénea en la Amazonía peruana. En otras palabras, existen regiones geográficas donde hay alta transmisión espacial y temporal que están rodeados por áreas extensas de poca o ninguna transmisión local.

Esta variación en la transmisión posiblemente está determinada por factores ambientales, antropológicos y económicos (Rosado et al., 2021).

En los últimos 24 años, la prevalencia de malaria en el Perú ha variado en intervalos extremos, esto mediado por la implementación, interrupción y continuación de los programas de intervención de malaria (Soto-Calle et al., 2017). Algunos mecanismos de control de la malaria en Latinoamérica fueron implementados con fondos internacionales tales como el programa PAMAFRO (2005 - 2010), el cual consistió en la distribución de insecticidas, reforzamiento de las capacidades locales para el diagnóstico, tratamiento y manejo del medio ambiente en once zonas de fronteras estratégicas (Feo et al., 2009). En consecuencia, se redujo el 78% de casos de malaria en Loreto (Janko et al., 2023). No obstante, la discontinuación de las políticas de control entre los años 2011-2016, cambios ecológicos y climáticos como el Fenómeno de El Niño contribuyeron al aumento de la transmisión de malaria, por el aumento a la exposición por picaduras de mosquitos y factores humanos que limitan la cobertura y eficacia de las intervenciones de control de malaria (Soto-Calle et al., 2017).

En el 2017, dado el aumento significativo del número de casos de malaria después de la discontinuación de las medidas por PAMAFRO, el Ministerio de Salud del Perú implementó el PMC (2017 - 2021), principalmente enfocado hacia el control de la malaria en la región de Loreto con tres fases principales: 1) fase I: control de la malaria, 2) fase II: eliminación de la malaria y 3) fase III: eliminación de la malaria residual (MINSA, 2017). El PMC demostró en Loreto una reducción cercana al 70% de los casos al 2020, evidenciando la efectividad de intervenciones integrales en áreas endémicas. Sobre la base de este antecedente, y ante los brotes

alarmantes de malaria en regiones como Amazonas y Junín, el MINSA implementó el PNEM, cuyo objetivo es disminuir en un 90% el número de casos de malaria en el Perú en el lapso 2022 – 2030 (MINSA, 2022).

Según lo esperado, entre 2022 y 2023 el número de casos de malaria disminuyó hasta un 18% a nivel nacional en comparación con el 2021. Sin embargo, en 2024 se reportó, principalmente en áreas limítrofes, un aumento considerable del 47.3% en comparación con 2023 (MINSA, 2024). Esto evidencia la presencia de “hotspots” de malaria (áreas de elevada prevalencia/incidencia y/o riesgos de transmisión) ubicados en las regiones fronterizas con Ecuador (Gunderson et al., 2020), Colombia (Rodríguez et al., 2024) y Brasil (Palma-Cuero et al., 2022), los cuales contribuyen con el aumento de infecciones.

En efecto, los factores ecológicos de Perú favorecen la alta prevalencia de *P. vivax* dado que los niveles de transmisión de malaria son perennes, es decir todo el año, con etapas de incrementos epidémicos estacionales y esporádicos en distintas zonas geográficas. La mayoría de los casos de *P. vivax* son asintomáticos, por lo que escapan de la vigilancia epidemiológica rutinaria basada en microscopía (Rosado et al., 2021). Además, las etapas latentes en estadios hepáticos (hipnozoitos) de *P. vivax* mantiene las infecciones mediante relapsos, complicando su diagnóstico y tratamiento eficaz (Rosas-Aguirre et al., 2016).

Aunque Loreto ha sido considerada como una región hipoendémica de malaria, los datos epidemiológicos y vigilancia sugieren que la transmisión de malaria es altamente heterogénea (Rosas-Aguirre et al., 2016). Estudios espacio-temporales (2006 - 2018) sobre la transmisión de malaria en la frontera amazónica indican una

mayor transmisión persistente de *P. vivax* en comparación a *P. falciparum*. La transmisión de *P. vivax* no es uniforme en todo el Perú, esto implica que algunas áreas poseen una mayor tasa de transmisión de malaria vivax, mientras que es menor en otras áreas. Esto revela un patrón similar a una gradiente de transmisión desde zonas más endémicas hacia zonas menos endémicas. Por ejemplo, la incidencia aumenta primero en Loreto y posteriormente en los cantones fronterizos de Ecuador (Gunderson et al., 2020), sugiriendo que las variaciones estarían influenciadas por factores ecológicos y ambientales como las temperaturas y lluvias (Rosas-Aguirre et al., 2016).

En Loreto, las tasas de incidencia anual entre 2002 al 2013 tuvieron una evolución en segmentos, se observó una disminución del 80% en la tasa de incidencia entre los años 2006 – 2010, resultados atribuidos a las medidas de control realizadas por PAMAFRO. Posteriormente, en el intervalo 2011 y 2013, la tasa de incidencia aumentó significativamente dado al bajo presupuesto nacional para continuar con los programas de eliminación de la malaria. Más aún, los análisis de conglomerados espaciales indicaron una mayor incidencia de *P. vivax* en las regiones fronterizas (Soto-Calle et al., 2017).

Por otro lado, en esta región se observó una variación significativa en la incidencia acumulada por cada mil habitantes entre los distintos periodos analizados (Tabla 2), evidenciando el impacto de los programas implementados como PAMAFRO y PMC (Díaz-Salazar et al., 2024). Sin embargo, es también probable que los valores de la incidencia acumulada del último periodo sea producto de los subregistros ocasionados por el COVID-19 (Torres et al., 2020).

Tabla 2. Incidencia acumulada de malaria por 1 000 habitantes en Amazonas y Loreto, 2000–2022. Adaptado de Díaz-Salazar et al. (2024).

Región	Periodos			
	2000-2005	2006-2011	2012-2017	2018-2022
Amazonas	0.37	0.18	0.85	3.02
Loreto	6.28	12.45	51.84	22.93

Por otro lado, Amazonas es la segunda región con mayor incidencia de malaria en Perú, con oscilaciones desde 1553 casos notificados en el 2020, 1657 en 2022 seguida de una progresiva disminución de la incidencia en los siguientes años (MINSA, 2026). El distrito Río Santiago, provincia de Condorcanqui, es un área crítica donde se reportó entre el 94.5% y 99.7% de los casos durante el 2023 y 2024 respectivamente. De estos casos, más del 90% fueron infecciones causadas por *P. vivax* (MINSA, 2024).

En los últimos 15 años, las tasas de incidencia de malaria en esta región han mostrado fluctuaciones significativas, alcanzando un pico máximo en el 2019. Curiosamente, entre el 2010 y el 2013, el área experimentó un periodo de silencio epidemiológico, periodo que coincide con la conclusión de PAMAFRO. Sin embargo, a partir del 2014, se observó un incremento notable en los casos de malaria, particularmente en el distrito Río Santiago (Saavedra-Samillán et al., 2024). Además, se reportó un brote inusual del *P. falciparum* en el 2018 (Montenegro et al., 2021). Todo esto sugiere una posible reintroducción de *Plasmodium* desde Loreto (Bustamante-Chauca, 2019).

Los valores de incidencia acumulada de malaria por periodos en Amazonas (Tabla

2) indica un aumento significativo durante el 2018 al 2022, principalmente en la provincia de Condorcanqui (Díaz-Salazar et al., 2024). En específico, las tasas de incidencia anual (AIR) de malaria en la provincia de Condorcanqui son mayores en los años 2019, 2020 y 2022 (52.28, 42.95 y 44.01, respectivamente) (Saavedra-Samillán et al., 2024). No obstante, la disminución de la incidencia en el 2021 (AIR=17.9), no necesariamente se atribuye a la efectividad de los planes de control como el PMC, sino posiblemente a las interrupciones de detección de casos activos debido a la pandemia COVID-19 (Montenegro et al., 2021).

También se destaca un patrón de gradientes de transmisión dentro de Amazonas, pues, se detectó un aumento en las tasas de incidencias y riesgo de malaria de áreas geográficas limítrofes con Río Santiago, tales como los distritos de Nieva, El Cenepa (ambas en la provincia de Condorcanqui) y provincias cercanas como Bagua a lo largo de los últimos 6 años (MINSA, 2024). Así, estos contextos específicos destacan la importancia de reforzar la vigilancia epidemiológica y genómica continua y aplicar estrategias que permitan comprender las dinámicas transmisión de malaria en regiones locales, todo esto con la finalidad de poder controlar la propagación a comunidades alejadas de difícil acceso.

4.3. Ciclo de vida de *Plasmodium*

Las especies de *Plasmodium* poseen un ciclo de vida complejo cuyas características generales son similares entre las especies, el cual incluye un ciclo asexual que se desarrolla en el humano, hospedero intermediario, y un ciclo sexual en los vectores, hospedero definitivo (Fig. 1) (Lima-Junior & Pratt-Riccio, 2016)

El ciclo empieza cuando los esporozoitos de *Plasmodium* producidos en el vector anofelino ingresan al huésped humano después de una picadura por este vector, posteriormente los esporozoitos se depositan en la dermis y migran hacia el hígado. El ciclo pre-eritrocítico ocurre en el hígado, donde los esporozoitos infectan a los hepatocitos y realizan un proceso de fisión múltiple llamado esquizogonia, entre aproximadamente 5 a 7 días. El producto son miles de células hijas llamadas merozoitos, los cuales son liberados a la sangre para la infección a eritrocitos, iniciando el ciclo eritrocítico, por lo general sintomático (Lima-Junior & Pratt-Riccio, 2016).

Cada merozoito infecta un eritrocito y se multiplica mediante esquizogonia generando entre 8 a 64 merozoitos dependiendo de la especie. En este punto, los merozoitos maduran pasando por etapas: anillo, trofozoito y esquizonte. Los nuevos merozoitos son liberados del esquizonte a la sangre mediante lisis del eritrocito infectado, repitiendo así el ciclo de propagación intraeritrocitaria cuya duración es aproximadamente entre 48 horas, aunque depende de la biología intrínseca de cada especie (Roques et al., 2023). En la etapa eritrocitaria, una porción de los merozoitos se diferencia en gametocitos (microgametocito y macrogametocito) para la reproducción sexual del parásito.

La siguiente etapa ocurre en el vector (ciclo esporogónico), cuando un mosquito pica e ingiere sangre de pacientes infectados con parásitos en la etapa de gametocitos. Posteriormente cada gametocito está programado para la diferenciación en gametos masculinos o femeninos (gametogénesis). Los gametos masculinos (microgameto) y femenino (macrogameto) se fusionan en la luz del intestino del mosquito produciendo un cigoto con genoma diploide.

El cigoto entra en etapa de meiosis y se diferencia a ookinete (forma no móvil). El ookinete atraviesa la pared del intestino y forma un ooquiste en el exterior donde entra en varias rondas de mitosis generando numerosos esporozoitos. El ooquiste maduro se rompe y libera los esporozoitos en la hemolinfa, estos migran a las glándulas salivales donde mantienen la capacidad de infectar a humanos, posterior a una picadura de mosquito. Esta etapa ocurre entre aproximadamente 10 a 18 días (Sato, 2021).

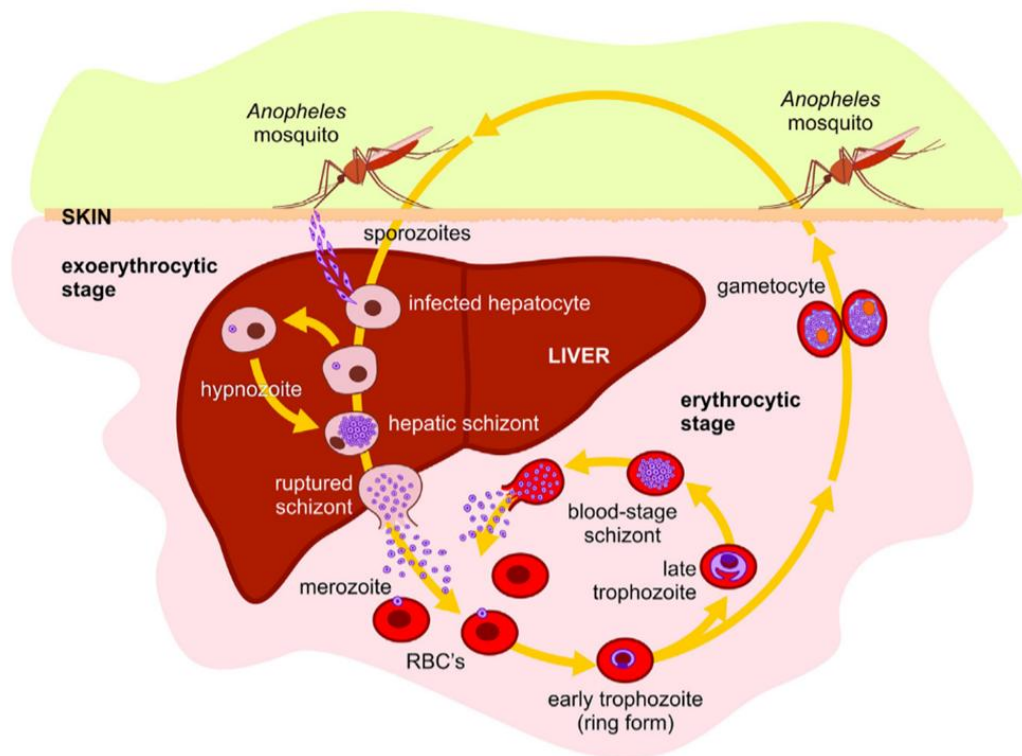


Figura 1. Ciclo biológico del parásito *Plasmodium vivax*. *P. vivax* es transmitido a los humanos mediante la picadura de un mosquito hembra *Anopheles* infectado. Imagen tomada de Lima-Junior & Pratt-Riccio, (2016).

4.4. Características biológicas de *Plasmodium vivax*

P. vivax posee particularidades en su biología en comparación con *P. falciparum*. Morfológicamente, la identificación de *P. vivax* puede distinguirse mediante la

observación de gránulos oscuros (puntos de Schüffner) en los reticulocitos mediante microscopía con láminas de gota gruesa, dado que la abundancia de estos gránulos aumenta durante su maduración en el estadio eritrocitario para *P. vivax*. Por otro lado, todos los estadios de desarrollo del ciclo eritrocitario pueden ser observados en la sangre periférica, siendo las morfologías de sus estadios sexuales similares a los estadios asexuales (Adams & Mueller, 2017).

El periodo de incubación en *P. vivax* es ligeramente más largo (12 días a meses) con respecto a *P. falciparum*, y aunque posee un ciclo eritrocítico semejante (48 h aproximadamente), produce menos merozoitos por esquizonte. Además, los eritrocitos infectados por *P. vivax* no presentan citoaderencia ni apilamientos en la microvasculatura (Price et al., 2007), pero tienen deformaciones en su estructura celular a medida que los parásitos maduran, evitando así el secuestro esplénico (Suwanarusk et al., 2004).

Los tratamientos antimaláricos principalmente disminuyen los niveles de parásitos intraeritrocitarios. Una diferencia importante de esta especie, presente también en *P. ovale*, es la capacidad de presentar recaídas tras los tratamientos antimaláricos de la infección inicial. Esto se debe a que una porción de los esporozoitos permanece latente durante periodos largos en el hígado en estadios llamados hipnozoitos, cuyo desarrollo se realiza después de culminar con los tratamientos causando infecciones recurrentes (Adams & Mueller, 2017).

P. vivax se caracteriza por la capacidad de infectar exclusivamente reticulocitos (eritrocitos inmaduros). La interacción ligando-receptor entre el parásito (merozoito) y los eritrocitos del huésped está mediado por la proteína de unión a

Duffy (DBP) de *P. vivax* (también en *P. knowlesi*) y el receptor del antígeno Duffy para quimiocinas (DARC) de humanos (Ntumngia et al., 2016).

Las poblaciones humanas en África no presentan en gran medida este receptor (grupos sanguíneos Duffy negativos), por lo que explicaría la baja prevalencia de *P. vivax* en este continente; sin embargo, también se reportaron infecciones de malaria vivax en poblaciones Duffy negativas, sugiriendo otras rutas de invasión aún no descritas completamente (Cavasini et al., 2007; Mendes et al., 2011).

La malaria causada por *P. vivax* en general posee baja parasitemia, relacionado a la preferencia selectiva de invadir los reticulocitos, a diferencia de *P. falciparum* que invade a eritrocitos en diferentes estados de maduración. Ello restringe el crecimiento de *P. vivax* tanto en *in vitro* como *in vivo*. Pese a que este factor limita los niveles de parasitemia, *P. vivax* es capaz de inducir fiebre en menores niveles de parasitemia que *P. falciparum* (Price et al., 2007).

4.5. Factores que contribuyen a la transmisión de malaria en el Perú

La transmisión de malaria es un proceso dinámico que involucra diversos factores interrelacionados unos con otros, tanto factores ambientales y antropológicos, es decir, demográficos y socioeconómicos (Kar et al., 2014). Las regiones de la Amazonía peruana y países fronterizos poseen condiciones que favorecen la distribución y la sobrevivencia de vectores de malaria. Ello sumado a la falta de continuidad y financiamiento de los planes de eliminación, disparidades y aplicaciones graduales de los recursos, contribuyen al aumento de la tasa de incidencia y transmisión de malaria (Janko et al., 2023).

En Perú se han caracterizado tres regiones endémicas de malaria: 1) Amazonía norte (Amazonas, San Martín, Huánuco, Loreto y Ucayali) 2) Amazonía sur (Madre de Dios, Junín y Cusco) y 3) Costa Norte (Tumbes, Piura, Lambayeque y La Libertad), cuyas áreas de riesgo de malaria están en zonas urbanas, periurbanas y rurales (Rosas-Aguirre et al., 2016). Las dos primeras regiones abarcan zonas correspondientes a la Amazonía peruana y se caracterizan por presentar condiciones ambientales como bosques tropicales, cuerpos de aguas, altas temperaturas, precipitaciones y humedad (Rosas-Aguirre et al., 2016).

Los bosques tropicales en relación con la dinámica de transmisión de malaria influyen mediante la amortiguación de la temperatura, las precipitaciones, la humedad, su densa vegetación y la presencia de ríos y cuerpos de agua cercanos a las poblaciones. Todos ellos contribuyen en la distribución y bionomía de los vectores, es decir, aspectos de su biología asociados a factores genéticos, ecológicos y ambientales (Kar et al., 2014). Esta ecología compleja y las interacciones de poblaciones humanas con mosquitos ocasionan que la transmisión de la malaria en la Amazonía peruana varíe significativamente en el espacio, tiempo y por especie (Gunderson et al., 2020).

Existen diversos factores antropogénicos que aumentan la vulnerabilidad a la infección por malaria. Estos incluyen la deforestación, reforestación, actividades humanas en los entornos forestales, movilizaciones humanas, migraciones estacionales y la pobre infraestructura de los hogares principalmente en zonas rurales (Gunderson et al., 2020). En conjunto, estos factores contribuyen a la abundancia de vectores y la supervivencia de especies de *Plasmodium* (Kar et al., 2014).

Otro factor que contribuye a la persistencia de la transmisión de malaria en la Amazonía peruana es la alta vulnerabilidad de las comunidades nativas. Las barreras culturales y sociales como las elevadas tasas de analfabetismo, uso de lenguas originarias y prevalencia de creencias tradicionales, dificultan su integración con los servicios de salud urbanos. Esto retrasa el diagnóstico temprano, limita el acceso a los tratamientos antimaláricos, favoreciendo en muchos casos la dependencia de prácticas terapéuticas autóctonas. Así mismo, los factores socioeconómicos de las comunidades nativas tales como el nivel de educación, ocupación, ingresos mensuales e incluso el tipo de vivienda (infraestructura) se asocian a mayores niveles de transmisión de malaria (Montenegro et al., 2021; Rojas-Muro, 2024).

El movimiento humano principalmente por vías fluviales y la migración masiva desde países altamente endémicos también agravan la situación de riesgo de malaria. Como ejemplo, Colombia y Brasil principalmente aumentaron el número de casos de malaria importada debido las migraciones desde Venezuela (Gunderson et al., 2020). Mientras que en Loreto, las cuencas fluviales (algunas principales como los ríos Pastaza, Yavarí y Amazonas) sirven como los principales medios de transporte (corredores de malaria) lo cual permite que reservorios humanos puedan transportar los parásitos en áreas alejadas que tienen las características para la sobrevivencia del parásito (Janko et al., 2024).

Otras características demográficas tales como sexo, edad, origen étnico y ubicación rural/urbana son factores asociados a los aumentos de transmisión de malaria. Por ejemplo, los casos de malaria están principalmente concentradas poblaciones etarias jóvenes (< 18 años) (MINSA, 2024) tanto en Amazonas (Bustamante-Chauca, 2019; Rojas-Muro, 2024), Loreto (Rosas-Aguirre et al., 2015, 2017) y Madre de

Dios (Sanchez et al., 2017), dado que estos constituyen el mayor grupo laboral en la Amazonía peruana relacionado a minería, comercio, agricultura, tala entre otros.

4.6. Diagnóstico y herramientas de vigilancia epidemiológica de malaria vivax

El diagnóstico de *P. vivax* tiene ciertas dificultades debido a que presentan bajos niveles de parasitemia (por debajo de los niveles de detección microscópica) y/o puede presentar formas latentes conocidos como hipnozoitos. Además, debido a que los signos y síntomas clínicos son inespecíficos y comunes frente a otras enfermedades febriles (Tangpukdee et al., 2009), el diagnóstico temprano es un componente vital para garantizar una administración adecuada del tratamiento, minimizar los riesgos de resistencia a antimaláricos, reacciones adversas y enfocar los planes de control y eliminación de la malaria (Amir et al., 2018).

Existen tres métodos de diagnóstico recomendados por la OMS: 1) microscopía, 2) pruebas de diagnóstico rápidas (PDR) y 3) pruebas moleculares. Los dos primeros son métodos rutinarios de diagnóstico parasitológico estrictamente utilizados, independiente o en conjunto, para el tratamiento clínico por los proveedores de atención sanitaria. En tanto, el tercero es utilizado como prueba confirmatoria de especie, reconocimiento de recaídas/recrudescencias y con fines de investigación científica para la contribución a proyectos de eliminación y contención de malaria (WHO, 2011).

Las técnicas moleculares como la reacción en cadena de la polimerasa (PCR) convencional, en tiempo real, múltiple, anidada y la amplificación isotérmica mediada por bucle (LAMP), microarreglos, entre otros (Tangpukdee et al., 2009),

pueden detectar los parásitos de manera precisa, con la capacidad de discernir entre infecciones mixtas, infecciones de baja parasitemia que no son detectables mediante microscopía o PDR (Oyegoke et al., 2022). Estos métodos se basan en la amplificación de ácidos nucleicos del parásito, con una alta sensibilidad de detección hasta 0.004 parásitos/uL en sangre (Tambo et al., 2018). Sin embargo, dado a la complejidad de la infraestructura, costos, necesidad de personal entrenado y de falta de implementación en campo, no son apropiados para el diagnóstico rutinario/clínico de malaria en áreas endémicas. Una excepción es LAMP, una herramienta prometedora para el uso en campo por su simplicidad y rapidez; no obstante, enfrenta limitaciones como el riesgo de falsos positivos y la dependencia de un diseño óptimo de cebadores (Yang et al., 2024).

Entre los últimos años, la epidemiología genómica ha ido ganando importancia significativa debido al poder resolutivo en torno a las respuestas evolutivas en base a los análisis de genomas del parásito. Los datos de genoma completo han podido dilucidar genotipos resistentes a antimaláricos, estructuración de poblaciones específicas y dinámicas de transmisión en contextos geográfico-temporales, conectividad entre regiones geográficas y dispersión de linajes a nivel global (Cowell et al., 2018; Hupalo et al., 2016; Kattenberg et al., 2024a).

Dada su importancia, la epidemiología genómica es utilizada por los programas nacionales de control y vigilancia de la malaria como herramienta para la formulación de políticas y guías normativas (Dalmat et al., 2019; Neafsey & Volkman, 2017). Pero, en vista que la secuenciación de genoma completo (WGS) es costosa, se optan por metodologías alternativas que sean similarmente resolutivas tales como microsatélites, microarreglos, la secuenciación de amplicones

(Ampliseq Custom Assay o AmpliSeq) para la vigilancia genómica de especies de *Plasmodium* (Kattenberg et al., 2022; Kattenberg et al., 2023a; LaVerriere et al., 2022).

La secuenciación de Amplicón (o AmpliSeq), utilizado como herramienta de vigilancia genómica, se basa en la secuenciación multiplexada de amplicones previamente amplificados por PCR, con la finalidad de enriquecer las regiones de interés y obtener una mejor profundidad de la secuenciación, incluso en muestras de baja concentración del ADN (Kattenberg et al., 2022). La vigilancia genómica de patógenos evalúa principalmente marcadores genéticos asociados a resistencia a los medicamentos, paneles de polimorfismos de un solo nucleótido (SNPs) que dilucidan estructuraciones poblacionales, así como genes sometidos a presión de selección dado su interacción con el sistema inmune (Kattenberg et al., 2023a). Así mismo, a diferencia de enfoques como la WGS, el AmpliSeq tiene la singularidad de basarse en los análisis de secuencias o regiones parciales del genoma, particularmente con SNPs, cuya información proporcionada revela estructuración poblacional, conectividad, origen de infecciones importadas en contexto geográfico (Kattenberg et al., 2023a).

En el contexto de la vigilancia molecular de malaria, se han diseñado ensayos AmpliSeq para *P. falciparum*, el cual involucra 233 amplicones que dentro de ellos incluyen fragmentos de 13 genes asociados a resistencia, fragmentos del gen *ama1* y *cpmp*, cuatro microsatélites y un código de barras de 28 SNPs específicos para las poblaciones parasitarias ubicadas en la Amazonía peruana, y finalmente fragmento de los genes *pfhrp2* y *pfhrp3* (Kattenberg et al., 2023a). Otro ensayo Ampliseq que involucra un primer panel “4CAST” que abarca fragmentos de cuatro loci altamente

polimórficos y otro panel que involucra 129 locus de marcadores asociados a resistencia a antimaláricos para *P. falciparum* presentes en América del Sur (LaVerriere et al., 2022).

Para *P. vivax* se han validado relativamente pocos polimorfismos en genes asociados a resistencia, principalmente debido a la ausencia de un cultivo *in-vitro* continuo. Por ello, los ensayos publicados para esta especie se han basado en genes ortólogos asociados a resistencia de *P. falciparum*. Un panel actualmente disponible (*Pv Ampliseq v2 Peru assay*) incluye de 227 amplicones, que incluyen fragmentos 11 genes ortólogos asociados a resistencia a fármacos, regiones que flanquean 41 SNPs específicos para el análisis poblacional en la Amazonía peruana, 33 SNPs del código de barras GEO33 para predicción de origen geográfico, y el gen *ama1*, un gen inmunogénico sometido a presión selectiva (Kattenberg et al., 2024b).

4.7. Genética de poblaciones de *Plasmodium vivax* en el Perú

La genética de poblaciones estudia la variación de la diversidad genética y los procesos evolutivos que la moldean en el espacio-tiempo, tanto dentro y entre poblaciones determinadas (Hartl & Clark, 2007). Algunos de los parámetros de medición de la diversidad genética están basados en la variabilidad alélica que posee la población, tales como la riqueza alélica, diversidad nucleotídica (π), diversidad haplotípica (h), la heterocigocidad esperada (H_e) y otros según el tipo de marcador utilizado (Kanaka et al., 2023).

En áreas endémicas de malaria es común que distintos clones de parásitos co-infecten a un mismo hospedero. Además, los parásitos de *P. vivax* muestran una notable capacidad de adaptación frente a estímulos o cambios ambientales,

desarrollando fenotipos únicos que favorecen su sobrevivencia, como la resistencia a antimaláricos, cambio de especificidad por el hospedero y la evasión inmunológica (Arnott et al., 2012; Sato, 2021). La diversidad genética de parásitos de *Plasmodium* está determinada por una variedad de factores evolutivos, los cuales incluye selección natural, deriva génica, flujo genético, recombinación entre otros (Barry et al., 2013).

Por ello, el uso de la genética de poblaciones permite cuantificar esta diversidad genética y asociarlas con procesos evolutivos específicos, contextualizando cómo estos influyen en los patrones de transmisión y persistencia de la malaria (Akoniyon et al., 2024; Hendry et al., 2021). En el caso de la Amazonía peruana, estudios previos sobre genética de poblaciones de *P. vivax* han documentado diferentes grados de diversidad y varios niveles de estructuración genética usando marcadores de genes específicos, microsatélites, polimorfismos de nucleótido único (SNPs) e incluso enfoques genómicos de mayor escala (Tabla 3).

Tabla 3. Resumen de los estudios previos en genética de poblaciones de *P. vivax* en la Amazonía peruana.

Años de muestreo	Marcadores / Plataforma	Diversidad Genética	Estructura Poblacional	Referencia
2003-2004	Gen <i>Pvmsp3a</i>	He: 0.6 - 0.8 (He = 0.845)	No reporta	(Sutton et al., 2009)
2006-2007	Genes <i>PvCSP</i> , <i>PvTRAP</i> , <i>PvDBP</i> , <i>PvMSP-1</i> , <i>PvAMA-1</i>	Hd: 0.85 - 0.97, (Hd _{MSP-1} = 0.36) π : 0.006 - 0.016	Evidencia de 3 subpoblaciones	(Chenet et al., 2012)
2006-2008	16 microsatélites neutrales	He: 0.44 - 0.69	Evidencia de 6 subpoblaciones	(Van den Eede et al., 2010)

2006-2008	Genoma completo	No reporta	Evidencia 5 subpoblaciones	(Cowell et al., 2018)
2019-2020	16 microsatélites neutrales	He: 0.68 - 0.78	Evidencia de 3 y 4 subpoblaciones	(Cabrera-Sosa et al., 2024b)
	SNPs*	He: 0.36 - 0.38		
2007-2019	Genoma completo	π : 0.001	Evidencia 7 subpoblaciones	(Kattenberg et al., 2024a)
2007-2020	SNPs*	He: 0.012 - 0.4	No reporta	(Kattenberg et al., 2024b)
2008-2010	15 microsatélites neutrales	He: 0.1 - 0.67 (He= 0.49)	Evidencia 4 subpoblaciones	(Delgado-Ratto et al., 2014)
2008	14 microsatélites neutrales	He: 0.66 - 0.76 (He=0.74)	Evidencia de 2-3 subpoblaciones.	(Delgado-Ratto et al., 2016)
2012-2015	16 microsatélites neutrales	He: 0.54 – 0.64 (He= 0.64)	Evidencia 6 subpoblaciones	(Manrique et al., 2019)
2013-2016	16 microsatélites neutrales	He: 0.60 – 0.64 (He=0.71)	Evidencia de 4 subpoblaciones	(Ramirez et al., 2025)
2015	16 microsatélites neutrales	He: 0.39 – 0.71 (He= 0.62)	Evidencia de 7 subpoblaciones	(Manrique, 2018)
2015-2017	16 microsatélites neutrales	He: 0.65 – 0.80	No reporta	(Ramirez, 2020)
2015-2019	SNPs**	π : 0.309 - 0.381 (π =0.34)	Evidencia de 3 subpoblaciones	(Villena et al., 2022)
2019-2020	SNPs*	He: 0.01 - 0.35	No reporta	(Cabrera-Sosa et al., 2024a)
2021-2022	Genoma completo	He: 0.15 - 0.2	Evidencia de 5 subpoblaciones	(Ramirez, 2024)

*panel Pv AmpliSeq v2 Peru

**panel SNPs de genes asociados a resistencia

Los estudios comparativos han revelado que *P. vivax* mantiene un patrón diferente al de *P. falciparum* en cuanto a diversidad genética. En *P. falciparum*, la variabilidad depende estrechamente de las intensidades de transmisión; en

cambio, *P. vivax* mantiene niveles altos de diversidad independientemente de la intensidad de transmisión (Sutton et al., 2009). Los altos niveles de diversidad genética de *P. vivax* en la Perú resultan inusuales considerando la relativamente baja intensidad de transmisión en la Amazonía peruana. Esta aparente contradicción podrían explicarse por los atributos únicos del parásito, entre ellas entre ellas la acumulación de hipnozoítos genéticamente heterogéneos a lo largo del tiempo (Rosas-Aguirre et al., 2016), su tendencia a mantener bajas densidades de parasitemia y una elevada producción de gametocitos (Arnott et al., 2012; Barry et al., 2015). También la intensidad gradual de las intervenciones de control de la malaria puede estar desempeñando un papel importante en la determinación de la heterogeneidad geográfica y cronológica de la transmisión de la malaria y su diversidad genética (Soto-Calle et al., 2017).

La diversidad intermedia a alta observada entre poblaciones de *P. vivax* están acompañados de una considerable proporción de infecciones policlonales y alta multiplicidad de infección (MOI). Además, las poblaciones regionales presentes en la Amazonía peruana son genéticamente diferentes al panorama global, con algunas excepciones de casos importados desde Brasil y/o Colombia en el punto triple frontera en la provincia de Mariscal Ramón Castilla, región de Loreto (Hupalo et al., 2016).

Los microsatélites han sido ampliamente utilizados para caracterizar los patrones de transmisión de *P.vivax* en la Amazonía peruana. Sin embargo, los estudios realizados no muestran un consenso respecto al número ni al tipo de microsatélites utilizados para definir la estructura genética poblacional. Aun así, los diferentes paneles aplicados coinciden en demostrar una alta diversidad genética de *P. vivax*

($H_e > 0.1$). De esta manera, se ha reportado una baja diferenciación genética entre comunidades que se encuentran en las mismas cuencas hidrográficas (Delgado-Ratto et al., 2016), así como genotipos similares de *P. vivax* presentes en comunidades considerablemente lejanas dentro de la Amazonía peruana (Manrique et al., 2019).

No obstante, el uso de microsatélites presenta limitaciones importantes debido a posibles artefactos de amplificación, la falta de estandarización entre plataformas y laboratorios, y su escasa viabilidad como herramientas sostenibles a largo plazo en programas de eliminación de malaria. Los barcodes de SNP, compuestos por paneles de 20 a 50 SNPs bialélicos, pueden ofrecer resultados más consistentes y representan una alternativa más eficaz (Auburn & Barry, 2017).

El *Pv AmpliSeq v2 Peru assay* posee un panel de 41 SNPs diseñado para analizar la estructura genética de *P. vivax* en la Amazonía peruana cuyos resultados son comparables a los microsatélites clásicos (Cabrera-Sosa et al., 2024b). Esta herramienta permitió detectar diferencias genéticas entre parásitos procedentes de zonas remotas y aquellos de áreas periurbanas de Iquitos, así como niveles moderados de diversidad genética ($H_e > 0.3$) en la mayoría de distritos, con excepción de Yavarí, ubicado en la región triple frontera. En este último caso, los hallazgos resultan consistentes con los análisis de genoma completo, que muestran una marcada similitud genética con cepas brasileñas (Kattenberg et al., 2024b). Por otro lado, mediante el mismo panel de SNPs, se ha reportado una moderados niveles

de diversidad genética ($H_e=0.35$) y diferentes linajes de *P. vivax* en comunidades nativas (Nueva Jerusalén) alejadas de lugares urbanos (Cabrera-Sosa et al., 2024a)

La mayoría de los estudios reportan un patrón consistente: las poblaciones de *P. vivax* en Iquitos y áreas adyacentes exhiben alta diversidad genética, acompañada de una marcada diferenciación de las localidades geográficamente lejanas. Esta estructura genética evidencia transmisión predominantemente local y espacialmente focalizada (Rosas-Aguirre et al., 2016), lo que refleja dinámicas complejas de conectividad y flujo génico entre regiones distantes.

La conectividad de la malaria se entiende como el flujo de parásitos entre las zonas de transmisión en un momento determinado, principalmente asociado al movimiento humano (Guerra et al., 2019). La información genética de los parásitos de la malaria ofrecen el potencial de proporcionar una medida más directa a dicha conectividad, además de permitir la identificación de fuentes y sumideros relevantes del movimiento del parásito (Auburn & Barry, 2017; Tessema et al., 2019).

En la Amazonía peruana, los altos niveles de flujo genético de malaria entre áreas geográficamente distantes están asociados a los niveles de movilidad humana entre localidades (Janko et al., 2023; Villena et al., 2022), tales como las migraciones estacionales, las cuales aumentan en épocas lluviosas en comparación a las épocas secas (Gunderson et al., 2023). En sitios de transmisión focalizada, se ha reportado altos niveles de similitud genética entre poblaciones parasitarias procedentes de

comunidades del río Mazán, sugiriendo una fuerte movilidad entre comunidades ubicadas en la misma cuenca (Carrasco-Escobar et al., 2020, 2022).

V. METODOLOGÍA

5.1. Tipo y diseño de Estudio

Este trabajo es un estudio de tipo retrospectivo, observacional, longitudinal, analítico. Comprende de análisis comparativos haciendo uso de genética de poblaciones para evaluar la estructuración genética, los niveles de diversidad genética basados en la estructuración, y flujo genético entre poblaciones parasitarias de *P. vivax* de las regiones de Amazonas y Loreto. Los análisis se basaron en secuencias genómicas (completas y parciales) obtenidos de repositorios públicos y privados. La agrupación de las muestras se realizó de acuerdo a su lugar de procedencia a nivel de distrito, provincias y región para evaluar los patrones de estructuración genética. La dimensión temporal no fue incluida en las interpretaciones debido a la baja representación muestral por ubicación geográfica. Las interpretaciones y conclusiones se fundamentan estrictamente en la estructura genética espacial entre regiones y provincias.

5.2. Cálculo del número muestral del estudio

Es fundamental considerar un número adecuado de muestras para lograr una representación confiable de la población. De acuerdo con los cálculos propuestos por Kattenberg, et al. (2024b), el tamaño muestral mínimo para cada región se estimó considerando una frecuencia alélica mínima de 1% (p), un margen de error de 2% (E) y un nivel de confianza del 95% ($Z=1.96$) en la fórmula de estimación del tamaño de muestra (Hajian-Tilaki, 2014):

$$n_0 = \frac{Z^2 \times p \times (1 - p)}{E^2} = \frac{1.96^2 \times 0.01 \times (1 - 0.01)}{0.02^2} = 95.0796 \cong 96$$

Según los cálculos, el número mínimo de muestras para cada región es de 96. Con el fin de ajustar este valor a la población real de infecciones, se aplicó la corrección por población finita utilizando como referencia el número promedio de casos confirmados malaria reportados (Tabla 4). Considerando la limitada disponibilidad de series anuales completas para 2007–2023 para ambas regiones, se adoptó el tamaño de la población como el promedio anual de casos confirmados en 2024-2022 y 2024-2019, con el fin de suavizar la variabilidad interanual.

Tabla 4. Casos confirmados de malaria en las regiones de Loreto y Amazonas entre los años 2019 – 2024. Información recopilada del reportes situacional de malaria del MINSA (2024).

Año	Loreto	Amazonas
2019	22,028	1,844
2020	13,397	1,553
2021	15,490	659
2022	22,723	1,657
2023	20,606	1,095
2024	31,678	570
Promedio 3 años (2021–2023)	19,606.33	1,137.00
Promedio 6 años (2019–2024)	20,987.00	1,229.67

Apartir del tamaño de la muestra inicial calculado con la formula anterior (n_0), se obtuvo los valores definitivos aplicando la fórmula de corrección infinita (Thrusfield et al., 2017):

$$n = \frac{N \times n_0}{N + n_0 - 1}$$

Para el periodo 2021–2023, el promedio anual (N) fue de aproximadamente 19,606 casos en Loreto y 1,137 en Amazonas, obteniéndose tamaños muestrales corregidos

de 94.6 y 87.8, respectivamente. De manera similar, al considerar el periodo 2019–2024, los promedios fueron de 20,987 casos en Loreto y 1,230 en Amazonas, con tamaños muestrales corregidos de 94.7 y 88.3, respectivamente. Dado que estas diferencias no son significativas, se adoptó un tamaño uniforme de 96 muestras por región, garantizando tanto la comparabilidad entre regiones como un margen adicional de seguridad estadística.

5.4. Minería de datos

En una primera etapa, se generó una base de metadatos a partir de la información disponible en los repositorios públicos internacionales. La búsqueda de accesiones se efectuó en GenBank (NCBI) en enero del 2024 utilizando operadores booleanos, con la siguiente consulta:

Plasmodium vivax[Organism] AND Peru[All Fields]

Esta búsqueda recuperó un total de 1,320 archivos de lectura de secuencias. Posteriormente, se implementó un script en Python (v.3.10) empleando la librería Biopython (Cock et al., 2009), que automatizó la recuperación de información mediante las funciones *Esearch* y *Efetch* de las *Entrez Programming Utilities (Eutilities)* (Sayers et al., 2025). De cada acceso se extrajeron metadatos relevantes, entre los que se incluyeron: ubicación geográfica (*country*), fecha de colección (*collection_date*), hospedero (*host*), nombre y tipo de muestra (*isolate*, *isolation_source*), número de BioProject, BioSample, así como identificadores asociados a SRA y RUN. El archivo fue enriquecido con la metadata pertinente correspondiente a secuencias que aún no se encuentran publicadas en repositorios públicos.

Una vez consolidada la metadata, se aplicaron criterios de inclusión y exclusión para la selección de las muestras a analizar:

- **Criterios de inclusión:**

- Muestras positivas para *Plasmodium vivax* con datos genómicos correspondientes a genomas completos (WGS) y/o parciales obtenidos mediante el panel *Pv AmpliSeq v2 Peru*.
- Datos genómicos de *P. vivax* procedentes de muestras recolectadas en las regiones de Loreto y Amazonas
- Datos genómicos secuenciados con la plataforma Illumina.
- Datos genómicos que incluyan información completa sobre año de colecta, ubicación geográfica, BioSample y BioProject.

- **Criterios de exclusión:**

- Muestras de *P. vivax* procedentes de regiones distintas a Amazonas y Loreto.
- Muestras *P. vivax* sin datos oficiales de año o lugar de colecta.

Tras aplicar los criterios de inclusión y exclusión, se realizó una búsqueda manual de información complementaria en los artículos científicos asociados a las secuencias. Esto con el fin de obtener datos adicionales sobre la ubicación geográfica (comunidades, distritos, provincias, regiones), fechas de colecta y nombres de las muestras y así obtener una metadata más completa e informativa, la cual esta detallada en el Anexo 1.

Una segunda etapa consistió en incorporar a la metadata y a la datos crudos aquellas secuencias genómicas no públicas, generadas por el Laboratorio de Malaria:

Parásitos y Vectores de la Universidad Peruana Cayetano Heredia, que cumplen con los criterios de inclusión y exclusión establecidos.

A partir de la minería de datos se identificó un total de 478 muestras de *P. vivax*, de las cuales 303 correspondieron a secuencias generadas mediante el panel *Pv AmpliSeq v2 Peru* y 175 a datos de genoma completo. Las muestras abarcan el período comprendido entre 2007 y 2023. En cuanto a la distribución geográfica, se registraron 73 muestras provenientes de Amazonas y 405 de Loreto, tal como se detalla en la Tabla 5.

Tabla 5. Número de muestras preliminares según tipo de data y ubicación geográfica contempladas para el estudio.

Ubicación geográfica		Tipo de data		Total
Región	Provincia	AmpliSeq*	WGS**	
Amazonas	Condorcanqui	73	---	73
	Datem del Marañón	---	4	4
Loreto	Loreto	68	---	68
	Maynas	139	170	309
	Ramón Castilla	23	1	24
Total		303	175	478

**Pv AmpliSeq v2 Peru*

**Genoma completo

5.4.1. Descripción de la data de Genoma Completo

El genoma de *Plasmodium vivax* está compuesto por 14 cromosomas con un tamaño aproximado de 26.8 Mb (Dharia et al., 2010). De la búsqueda de secuencias de genoma completo se identificaron 175 muestras en total, de las cuales 170 procedían de la provincia de Maynas, 4 del Datem del Marañón y

1 de Ramón Castilla, todos en la región de Loreto.

Se observó que las principales tecnologías de secuenciamiento empleadas fueron la amplificación selectiva del genoma (SWGA, por sus siglas en inglés) (Cowell et al., 2017) y la hibridación selectiva (Hupalo et al., 2016), las cuales han permitido incrementar la cobertura y profundidad de los datos. En consecuencia, los genomas publicados en bases de datos públicas presentan parámetros de alta calidad, con coberturas superiores a 20X en el genoma central según las literaturas consultadas.

Por otro lado, para los datos obtenidos *in house* por el Laboratorio de Malaria: Parásitos y Vectores (Ramírez, 2024), se emplearon todas las muestras disponibles, las cuales fueron previamente generadas y procesadas por dicho laboratorio (Ramírez, 2024). Cabe señalar que no se tuvo acceso directo a los índices de calidad del secuenciamiento (Q30 y cobertura), por lo que las secuencias se analizaron tal como fueron proporcionadas.

5.4.2. Descripción de la data de AmpliSeq Custom Assay

Dentro de la data disponible proveniente del panel *Pv AmpliSeq v2 Peru*, se identificaron 73 muestras de la provincia de Condorcanqui (región Amazonas) y 23, 68 y 139 correspondientes a las provincias de Mariscal Ramón Castilla, Loreto y Maynas (región Loreto), respectivamente. Este ensayo cubre aproximadamente 59.8 kb, lo que representa alrededor del 0.22% del genoma total de *P. vivax* (Kattenberg et al., 2024b).

El panel está diseñado para la amplificación de 227 amplicones en total. De estos, 149 corresponden a regiones asociadas a resistencia a antimaláricos distribuidas 11

genes putativos (*pvcrt*, *pvmrp1*, *pvdmt2*, *pvdhfr*, *pvmr1*, *pvp13k*, *pvacbc-e1*, *pvk13*, *pvmr2*, *pvdhps*, *pvmrp2*). Adicionalmente, 73 regiones que abarcan SNPs utilizados como código de barras (*barcode*) ubicados en regiones neutrales del genoma y sin desequilibrio de ligamiento. Estos incluyen 41 SNPs específicos para el análisis de genética de poblaciones en Perú (Tabla 6) y otro complementario de 33 SNPs globales (GEO33) orientado a la clasificación de muestras según su país de origen (Tabla 7). Finalmente, el panel incorpora 6 fragmentos del gen *pama1*, un gen inmunogénico sujeto a presión de selección por su interacción con el sistema inmune del hospedero.

Tabla 6. Lista de los 41 SNPs de *P. vivax* pertenecientes al código de barras del panel *Pv AmpliSeq v2 Peru*. Información adaptada de Kattenberg et al. (2024b)

SNP	Cromosoma	Posición	Alelo de referencia	Alelo alternativo
PvPAM_01_v1_852015	PvPAM_01_v1	852015	G	C
PvPAM_01_v1_1002217	PvPAM_01_v1	1002217	C	T
PvPAM_01_v1_1215815	PvPAM_01_v1	1215815	G	A
PvPAM_02_v1_568237	PvPAM_02_v1	568237	C	G
PvPAM_02_v1_683149	PvPAM_02_v1	683149	T	C
PvPAM_03_v1_80858	PvPAM_03_v1	80858	A	G
PvPAM_03_v1_523358	PvPAM_03_v1	523358	G	A
PvPAM_04_v1_527689	PvPAM_04_v1	527689	C	T
PvPAM_04_v1_570105	PvPAM_04_v1	570105	G	T
PvPAM_04_v1_1021791	PvPAM_04_v1	1021791	T	G
PvPAM_05_v1_478165	PvPAM_05_v1	478165	T	G
PvPAM_05_v1_656195	PvPAM_05_v1	656195	C	T
PvPAM_05_v1_1508069	PvPAM_05_v1	1508069	T	A
PvPAM_06_v1_699539	PvPAM_06_v1	699539	T	G
PvPAM_06_v1_841394	PvPAM_06_v1	841394	C	T
PvPAM_06_v1_1079495	PvPAM_06_v1	1079495	G	T
PvPAM_07_v1_193199	PvPAM_07_v1	193199	A	G
PvPAM_07_v1_370850	PvPAM_07_v1	370850	T	V
PvPAM_07_v1_1056743	PvPAM_07_v1	1056743	A	G

PvPAM_07_v1_1518811	PvPAM_07_v1	1518811	G	A
PvPAM_08_v1_558173	PvPAM_08_v1	558173	C	G
PvPAM_08_v1_1419614	PvPAM_08_v1	1419614	A	C
PvPAM_08_v1_1942368	PvPAM_08_v1	1942368	C	G
PvPAM_09_v1_299609	PvPAM_09_v1	299609	G	A
PvPAM_09_v1_783225	PvPAM_09_v1	783225	C	T
PvPAM_09_v1_819122	PvPAM_09_v1	819122	A	T
PvPAM_10_v1_237217	PvPAM_10_v1	237217	T	C
PvPAM_10_v1_623149	PvPAM_10_v1	623149	G	A
PvPAM_10_v1_1475696	PvPAM_10_v1	1475696	G	A
PvPAM_11_v1_202711	PvPAM_11_v1	202711	T	C
PvPAM_11_v1_277752	PvPAM_11_v1	277752	A	C
PvPAM_11_v1_1852190	PvPAM_11_v1	1852190	A	T
PvPAM_12_v1_366381	PvPAM_12_v1	366381	C	T
PvPAM_12_v1_962294	PvPAM_12_v1	962294	C	A
PvPAM_12_v1_1417132	PvPAM_12_v1	1417132	A	G
PvPAM_13_v1_494676	PvPAM_13_v1	494676	--	--
PvPAM_13_v1_1816561	PvPAM_13_v1	1816561	C	T
PvPAM_13_v1_2201010	PvPAM_13_v1	2201010	G	T
PvPAM_14_v1_933409	PvPAM_14_v1	933409	C	T
PvPAM_14_v1_1501301	PvPAM_14_v1	1501301	G	A
PvPAM_14_v1_2962609	PvPAM_14_v1	2962609	T	G

Tabla 7. Lista de los 33 SNPs de *P. vivax* pertenecientes al código de barras del para la predicción de origen (GEO33). Información adaptada de Trimarsanto et al., (2022)

SNP	Cromosoma	Posición	Alelo de referenci a	Alelo alternativ o
PvPAM_03_v1_186855	PvPAM_03_v1	186855	C	C
PvPAM_05_v1_823331	PvPAM_05_v1	823331	T	G
PvPAM_05_v1_1197701	PvPAM_05_v1	1197701	G	G
PvPAM_05_v1_1403522	PvPAM_05_v1	1403522	T	T
PvPAM_06_v1_896980	PvPAM_06_v1	896980	T	T
PvPAM_07_v1_358111	PvPAM_07_v1	358111	A	A
PvPAM_07_v1_692580	PvPAM_07_v1	692580	G	G
PvPAM_08_v1_797303	PvPAM_08_v1	797303	G	T
PvPAM_08_v1_1235207	PvPAM_08_v1	1235207	T	T
PvPAM_08_v1_1778852	PvPAM_08_v1	1778852	T	T

PvPAM_09_v1_307765	PvPAM_09_v1	307765	C	C
PvPAM_09_v1_453201	PvPAM_09_v1	453201	C	C
PvPAM_09_v1_1109407	PvPAM_09_v1	1109407	G	G
PvPAM_09_v1_1893592	PvPAM_09_v1	1893592	T	G
PvPAM_10_v1_534573	PvPAM_10_v1	534573	G	G
PvPAM_10_v1_537540	PvPAM_10_v1	537540	G	G
PvPAM_10_v1_1187835	PvPAM_10_v1	1187835	G	G
PvPAM_11_v1_334807	PvPAM_11_v1	334807	C	C
PvPAM_11_v1_455099	PvPAM_11_v1	455099	A	A
PvPAM_11_v1_1089463	PvPAM_11_v1	1089463	C	C
PvPAM_11_v1_1178693	PvPAM_11_v1	1178693	G	T
PvPAM_11_v1_1400961	PvPAM_11_v1	1400961	A	G
PvPAM_11_v1_1822062	PvPAM_11_v1	1822062	C	C
PvPAM_12_v1_1102187	PvPAM_12_v1	1102187	C	C
PvPAM_12_v1_1385935	PvPAM_12_v1	1385935	A	T
PvPAM_12_v1_1448959	PvPAM_12_v1	1448959	G	G
PvPAM_12_v1_1974573	PvPAM_12_v1	1974573	C	C
PvPAM_13_v1_528291	PvPAM_13_v1	528291	T	A
PvPAM_14_v1_461256	PvPAM_14_v1	461256	T	G
PvPAM_14_v1_1277006	PvPAM_14_v1	1277006	C	C
PvPAM_14_v1_1345741	PvPAM_14_v1	1345741	G	G
PvPAM_14_v1_1374641	PvPAM_14_v1	1374641	C	C
PvPAM_14_v1_1382422	PvPAM_14_v1	1382422	A	A

5.5. Área de Estudio

Los datos genómicos disponibles corresponden a centros poblados ubicados en doce distritos endémicos de las regiones de Amazonas y Loreto. En Amazonas, el área de estudio se concentra en comunidades de la provincia de Condorcanqui, mientras que en Loreto incluye las provincias de Datem del Marañón, Mariscal Ramón Castilla, Loreto y Maynas (Fig. 2).

El mapa del área de estudio fue elaborado en QGIS (versión 3.44.2) a partir de capas oficiales de límites político-administrativos del Instituto Nacional de Estadística e Informática – INEI (2017). Esta representación cartográfica permitió visualizar la

distribución espacial de los distritos endémicos incluidos cuyos datos genómicos de *P. vivax* están disponibles y delimitar las zonas consideradas en el análisis.

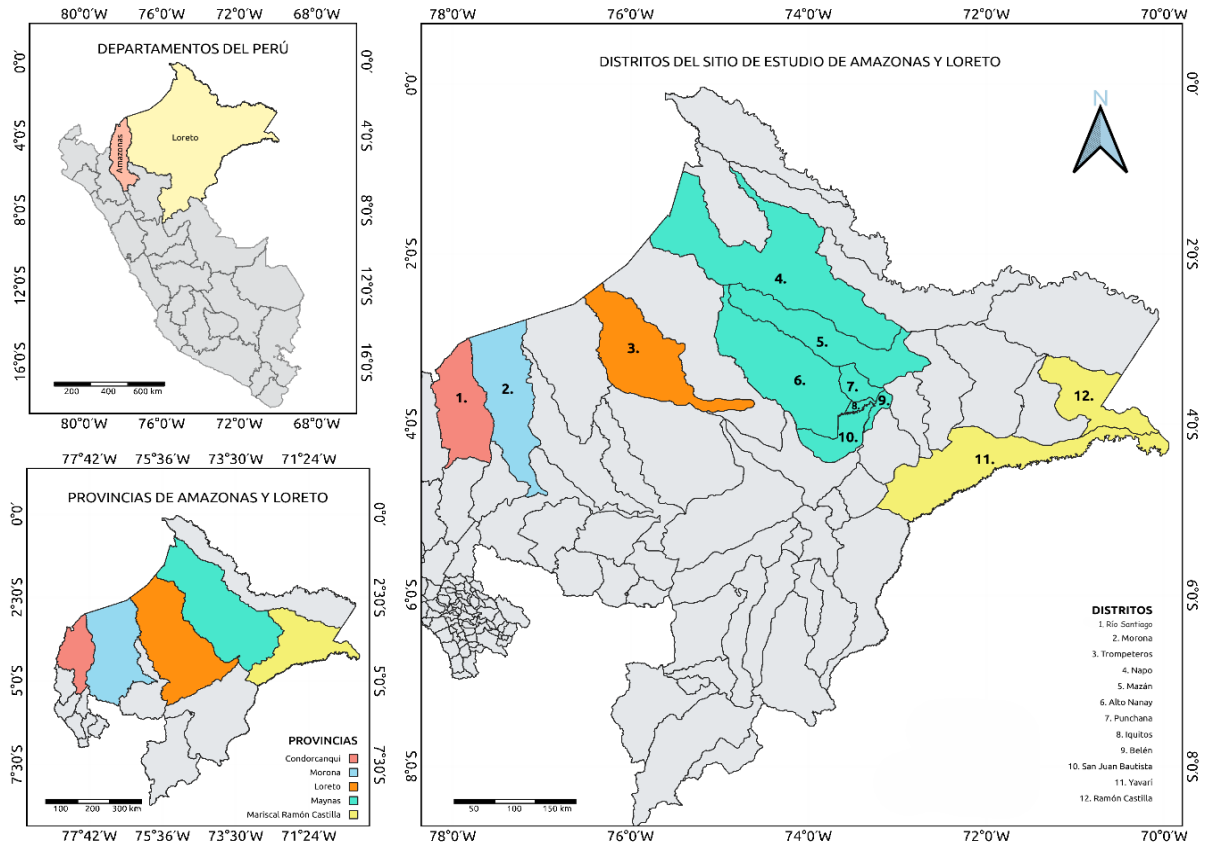


Figura 2. Mapa del área de estudio. Muestra la ubicación de los distritos endémicos incluidos de las muestras recolectadas en Amazonas y Loreto.

5.6. Análisis bioinformático

Todos los scripts desarrollados en Python y R, junto con los archivos de entrada utilizados para la visualización de los resultados se encuentran disponibles en el repositorio privado de GitHub con el siguiente link: “<https://github.com/JhonZumaeta/Master-thesis-UPCH-JZ-guide>”. El acceso al repositorio puede solicitarse mediante el correo electrónico.

5.6.1. Obtención de las lecturas genómicas crudas

Los datos genómicos se descargaron desde el repositorio público “Sequence Read Archive (SRA)” del NCBI (Kodama et al., 2012). Para ello se utilizó el paquete SRA Toolkit (NCBI, 2024), el cual permite descargar en lotes de los archivos de corridas de secuenciación depositados. A partir de una lista de 383 identificadores de acceso (SRR), previamente definida tras el proceso de minería de datos, se descargaron los archivos públicos en formato comprimido “.sra”. Posteriormente, estos archivos fueron transformados al formato FASTQ mediante la utilidad fasterq-dump del mismo paquete. Los 95 archivos restantes fueron obtenidos directamente de la base de datos internas del Laboratorio de Malaria: Parásitos y Vectores.

5.6.2. Filtrado y control de calidad de las lecturas genómicas crudas

Los 478 datos de secuenciación de *P. vivax* obtenidos en formato FASTQ fueron procesados en un entorno Linux, utilizando micromamba para la gestión de dependencias y la instalación de herramientas. Se procuró mantener la reproducibilidad documentando las versiones de cada programa y los parámetros aplicados en cada etapa.

Para el procesamiento de la data genómica se siguió el flujograma descrito por (Kattenberg et al., 2023b), con algunas modificaciones menores que se describen a continuación. Para evaluar la calidad de las lecturas de los archivos FASTQ se usó el programa FastQC v0.12.1 (Andrews, 2010). Este control incluyó la inspección de la calidad promedio por posición, la detección de contenido GC anómalo, la identificación de adaptadores remanentes y la presencia de duplicados. Adicionalmente, se empleó FastQ-Screen v.0.16.0 (Wingett & Andrews, 2018) con

el alineador BWA v0.7.18 (Li & Durbin, 2009) para descartar contaminaciones o lecturas que pudieran mapear a genomas distintos al de interés. Para ello, se consideraron genomas de referencia descargados de la base de datos NCBI Assembly (GenBank assemblies, accesiones GCA), incluyendo *P. falciparum* (3D7, GCA_000002765.3), *P. malariae* (PmG01, GCA_900090045.1), *P. ovale* (Poc2, GCA_900088565.1), *P. knowlesi* (strain H, GCA_000006355.2) y el genoma humano (GRCh38, GCA_000001405.29).

Posteriormente, se aplicó un proceso de filtrado y recorte de lecturas con Trimmomatic v0.39 (Bolger et al., 2014), eliminando secuencias de adaptadores y bases con baja calidad en los extremos. Se definió un umbral de calidad Phred ≥ 15 en ventanas deslizantes de cuatro bases, se descartaron lecturas con longitud inferior a 36 nucleótidos y se conservaron únicamente aquellas parejas de lecturas que superaban los criterios establecidos. En los conjuntos de datos de genoma completo, se utilizó la herramienta Trim Galore v0.6.10 (Krueger, 2015), que combina Cutadapt y FastQC, con parámetros equivalentes para garantizar la homogeneidad del procesamiento. Una vez depuradas, las lecturas fueron nuevamente sometidas a FastQC para confirmar la mejora en los perfiles de calidad.

5.6.3. Alineamiento y control de calidad del mapeo

El alineamiento de las lecturas filtradas se realizó contra el genoma de referencia PvPAM (GCA_949152365.1) (De Meulenaere et al., 2023) empleando BWA v0.7.18 con el algoritmo BWA-MEM (Li & Durbin, 2009). Durante este proceso se asignaron etiquetas de grupo de lectura (read groups) necesarias para el análisis posterior con GATK, y se habilitó la opción -M para marcar alineamientos secundarios. Los alineamientos generados en formato SAM se convirtieron a BAM,

se ordenaron por coordenadas genómicas con respecto al genoma de referencia y se indexaron utilizando Samtools v0.1.7 (Li et al., 2009). Asimismo, se utilizó Picard v2.20.4 (Broad Institute, 2019) a través del módulo CollectAlignmentSummaryMetrics. Estas estadísticas resumen la proporción de lecturas mapeadas, duplicados, y parámetros generales de la eficiencia de alineamiento, siendo almacenadas en archivos de texto para cada tipo de secuenciación (WGS y Ampliseq).

5.6.4. Identificación, filtrado y selección de variantes

Los alineamientos procesados se sometieron al flujo de trabajo recomendado en las “*Best Practices*” de Genome Analysis Toolkit (GATK) v4.3.0.0 (McKenna et al., 2010). Para cada muestra, los archivos BAM alineados fueron analizados con el módulo HaplotypeCaller para la identificación de SNPs y mutaciones, generando archivos gVCF para cada muestra. Para el análisis conjunto, los archivos gVCF individuales se integraron mediante el módulo GenomicsDBImport y posteriormente se ejecutó GenotypeGVCFs, produciendo un archivo VCF combinado. Las variantes se restringieron a las regiones específicas cubiertas por el panel *Pv AmpliSeq v2 Peru* usando SelectVariants, luego se filtraron y descartaron aquellos archivos que cumplieron los siguientes criterios:

- Baja calidad de mapeo ($MQ < 50$), indicativas de lecturas mal alineadas o en regiones repetitivas.
- Bajos puntajes de calidad de variante “Phred Score” ($QUAL < 30$), lo que corresponde a menos del 99.9% de confianza.
- Posiciones con cobertura insuficiente ($DP < 100$), asegurando que solo se retuvieran aquellas con respaldo adecuado en profundidad de secuenciación.

- Baja calidad relativa a la profundidad ($QD < 1.0$), señala variantes con baja robustez pese a estar bien cubiertas.
- Alto sesgo de hebra ($SOR > 4.0$), donde las lecturas que soportan la variante provienen casi exclusivamente de una sola orientación
- Baja distribución sesgada dentro de las lecturas ($ReadPosRankSum < -10.0$), lo que ocurre típicamente cuando los cambios aparecen solo en los extremos de las lecturas.

De manera adicional, se aplicó un filtrado por profundidad mínima de genotipado ($DP \geq 6$), marcando como no confiables los genotipos que no alcanzaban este umbral. Los genotipos extraídos mediante *VariantsToTable* incluyeron información de cromosoma, posición, tipo de variante, calidad y métricas de soporte de lectura. Posteriormente se conservaron aquellos archivos VCF que superaron todos estos filtros.

Asímismo, a partir de los archivos VCF del panel *Pv Ampliseq v2 Peru* que cumplieron con los filtros, se generaron subconjuntos de variantes correspondientes a tres archivos de interés. Primero, se obtuvo el barcode *ITM-Perú*, conformado por 41 SNPs bialélicos adaptados para genética de poblaciones de aislados peruanos (Tabla 6); luego, se generó el barcode GEO33, integrado por 33 SNPs bialélicos para la predicción de origen geográfico (Tabla 7). Ambos fueron extraídos de las coordenadas genómicas específicas. Finalmente, se obtuvo un archivo que incluyó todos los SNPs bialélicos identificados en este estudio ($n = 802$) para los análisis de genética de poblaciones.

5.6.5. Estadísticas de los archivos de secuenciación

Se integraron los archivos de estadísticas de alineamiento y metadatos de cada muestra, incluyendo información de las muestras, lugar de colecta y tipo de plataforma (AmpliSeq o WGS). A partir de estos datos se calcularon métricas como número total de lecturas, porcentaje de lecturas alineadas, longitud media de lectura y cobertura sobre el panel de referencia (~59,8 kb). Para la visualización se emplearon paquetes de R: ggplot2 (Wickham, 2016) y ggpubr (Kassambara, 2020), generando diagramas comparativos por regiones y tipos de datos. Adicionalmente, los reportes de control de calidad de FastQC fueron agregados mediante el paquete fastqcr (Kassambara, 2018), permitiendo comparar métricas antes y después del recorte de lecturas.

Además, se establecieron criterios de inclusión para garantizar la calidad de todos los datos, reteniéndose únicamente las muestras con una cobertura media de alineamiento mayor 10× (reads/posición) y con un porcentaje de genotipos faltantes menor al 35%. Estos criterios definieron el conjunto final de muestras utilizado en los análisis genéticos.

Para contextualizar el conjunto final de muestras, se elaboró un gráfico circular con el número de muestras por distrito y año mediante el paquete ggplot2 y un mapa temático con el paquete sf (Pebesma, 2018) para mostrar la distribución espacial de las muestras por provincia.

5.6.6. Análisis de Estructuración Genética

Para los análisis comparativos y la evaluación de la resolución de la estructuración genética, se utilizaron dos conjuntos de datos: (i) el conjunto completo de SNPs

bialélicos identificados en este estudio ($n = 802$), en adelante denominado “conjunto completo de SNPs”, y (ii) el “barcode *ITM-Peru*” ($n = 41$).

Los archivos VCF del conjunto completo de SNPs y el barcode *ITM-Peru* se transformaron en una matriz de secuencias y posteriormente en formato DNABin con el paquete ape (Paradis & Schliep, 2019). A partir de esta matriz se calcularon distancias genéticas bajo el modelo de diferencias sin corrección (raw distances), aplicando eliminación por pares para datos faltantes. Con estas distancias se construyó un árbol filogenético mediante el algoritmo Neighbor-Joining (NJ). El árbol resultante fue exportado en formato Newick (.nwk) y graficado utilizando los paquetes ggtree (Yu et al., 2017) y ggtreeExtra (Xu et al., 2021) con la incorporación de metadatos asociados a cada muestra, es decir, ubicación geográfica y periodos de muestreo.

La estructuración genética se exploró mediante el Análisis de Componentes Principales (PCA) con el paquete stats. Este enfoque permitió resumir la variabilidad genética de los SNPs bialélicos en ejes ortogonales, visualizando agrupamientos entre muestras sin suponer una estructura poblacional previa. Primero los genotipos se convirtieron en una matriz de frecuencias alélicas con el paquete adegenet (Jombart & Ahmed, 2011) imputando valores faltantes con la media por locus. Sobre esa matriz se aplicó un PCA con los parámetros por defecto. La proyección de los individuos se realizó con los paquetes ggplot2 y ggfortify (Tang et al., 2016), diferenciando por ubicación geográfica; además se realizó un sondeo por periodos de muestreo. En el caso de cluster solapamiento, se procedió a realizar el PCA solo incluyendo las muestras de dichos clústeres.

De manera complementaria, se aplicó un Análisis Discriminante de Componentes Principales (DAPC), también con el paquete *adegenet*, para maximizar la variación genética entre grupos definidos a priori (provincias y periodos de muestreo). El número óptimo de componentes principales retenidos se determinó mediante validación cruzada. Los resultados se representaron gráficamente de acuerdo a la distribución por ubicación geográfica y periodos de muestro. De igual manera, en el caso de solapamiento de clusters, la visualización se realizó incluyendo únicamente las muestras involucradas.

Adicionalmente se realizó un Análisis de Varianza Molecular (AMOVA) empleando los genotipos del barcode *ITM-Peru* y el conjunto completo de SNPs bialelicos del panel. Con base en los resultados del análisis de componentes principales y la visualización de una estructuración genética consistente con la distribución por provincias, se aplicaron dos enfoques complementarios: 1) AMOVA jerárquico geográfico: Regiones (nivel superior) y Provincias (nivel inferior); las provincias fueron asignadas a su respectiva región. El modelo jerárquico se especificó en el paquete *poppr* (Kamvar et al., 2014). Apartir de este análisis se estimaron los componentes de varianza, índices de diferenciación y la significancia estadística de los coeficientes se evaluó mediante permutaciones aleatorias ($n=999$), estableciendo un nivel de significancia de $p < 0.05$. 2) AMOVA no jerárquico temporal: se consideró un único nivel de agrupamiento según el rango temporal de muestreo, con el mismo procedimiento para las estimaciones de los coeficientes, índices de diferenciación y significancia estadística.

El análisis de estructura poblacional también se realizó con *ADMIXTURE* v1.3.0 (Alexander et al., 2009), a partir de datos en formato PED/MAP generados con

PLINK v1.9 (Purcell et al., 2007) usando solo la data del conjunto completo de SNPs bialélicos. El modelo no supervisado de ADMIXTURE se ejecutó para $K = 1-10$ y 20 iteraciones para cada K . Se incluyó bootstrapping (10000) para estimar la incertidumbre de las proporciones de ancestría. Para cada K se redirigió la salida a un registro (logK.out). El error de validación cruzada reportado por ADMIXTURE se extrajo de los archivos de registro y se compiló en una tabla resumen para seleccionar el K óptimo (mínimo *CV error*), utilizando los paquetes tidy (Wickham et al., 2023) y tibble (Müller & Wickham, 2023). El ordenamiento de individuos por cluster dominante se realizó mediante el paquete dplyr (Wickham et al., 2014), y las matrices Q resultantes (proporciones de ancestría por individuo), organizadas según la metadata disponible (provincia, periodo de muestro), fueron utilizadas en la fase de visualización.

Consistente con los patrones de estructuración genética encontrados, se procede a analizar los niveles de diferencia poblacional. Para el análisis a nivel intrapoblacional, se estimó coeficiente de endogamia dentro de poblaciones ($F_{IS} = 1 - \frac{H_O}{H_E}$) utilizando el paquete hierfstat (Goudet, 2005). Los valores obtenidos de F_{IS} por locus y agrupados por provincia fueron empleados para resumir la variación del coeficiente y graficar su distribución mediante diagramas de caja. Los valores de F_{IS} se interpretaron de manera descriptiva, como un indicador indirecto de la variabilidad genética intra-poblacional mediante la ausencia o presencia relativa de heterocigotos, sin asumir equilibrio de Hardy–Weinberg ni endogamia real entre parásitos. Así, con los valores de F_{IS} obtenidos del conjunto completo de SNPs bialélicos se compararon con los derivados de los valores obtenidos por el barcode *ITM-Peru* a fin de evaluar la consistencia de los patrones de diversidad intra-

poblacional detectados entre ambos conjuntos de datos.

La diferenciación genética interpoblacional se evaluó mediante estimaciones del estadístico F_{ST} (Weir & Cockerham, 1984) usando también los paquetes hierfstat (Goudet, 2005) y poppr (Kamvar et al., 2014). El análisis se realizó a nivel de regiones, provincias y distritos, considerando ambos conjuntos de datos de SNPs bialélicos. Los valores negativos de F_{ST} , que carecen de interpretación biológica, fueron ajustados a cero. Los resultados se representaron como matrices de diferenciación genética y se visualizaron en forma de mapas de calor utilizando el paquete ggcorrplot (Kassambara, 2023b). Con esa matriz, se contruyó un matriz de distancias y se realizó un agrupamiento jerárquico por promedio no ponderado – UPGMA (Segura-Alabart et al., 2022).

5.6.7. Análisis de Diversidad Genética

Los archivos VCF fueron procesados con el paquete vcfR (Knaus & Grünwald, 2017) y convertidos a objetos genind utilizando el paquete adegenet (Jombart & Ahmed, 2011), tanto para el conjunto completo de SNPs bialélicos como para el barcode *ITM-Peru*. La diversidad genética se evaluó a partir de dos parámetros principales, considerando las agrupaciones definidas a partir de los patrones de estructuración genética. La heterocigosidad esperada (H_e) se calculó únicamente a partir del barcode *ITM-Peru*, siguiendo la formula: $H_e = \frac{1}{K} \sum_{k=1}^K (1 - \sum_{i=1}^{m(k)} f_i^2)$, donde $m(k)$ corresponde al número de alelos en el locus k , K al número total de loci y f_i a la frecuencia del alelo i en la población considerada (Jombart & Ahmed, 2011). La diversidad nucleotídica (π) se estimó con el paquete VCFtools v0.1.17 (Danecek et al., 2011), utilizando ventanas deslizantes de 500 pb a lo largo de las regiones cubiertas por el panel *Pv AmpliSeq v2 Peru*.

La complejidad de infección (COI) de muestras se estimó a partir del índice de fijación dentro del hospedador (F_{WS}) (Auburn et al., 2012), calculado sobre los genotipos derivados del archivo VCF del barcode *ITM-Peru*. El archivo fue convertido al formato GDS y procesado con el paquete SeqArray (Zheng et al., 2017), excluyendo las muestras con más del 50% de datos faltantes. El índice F_{WS} se obtuvo mediante la función `getFws` del paquete `moimix` (Lee & Bahlo, 2016), el cual cuantifica la proporción de diversidad genética dentro de cada infección en relación con la población total. Los valores de F_{WS} se interpretaron como una medida indirecta de COI: $F_{WS} \geq 0.95$ indica infecciones monoclonales, mientras que $F_{WS} < 0.95$ refleja infecciones policlonales.

5.6.8. Análisis de Conectividad Genética

La detección de segmentos de identidad por descendencia (IBD) se realizó a partir de los genotipos en formato PED/MAP, generados a partir del conjunto completo de SNPs bialélicos. Las posiciones físicas fueron convertidas a unidades de centiMorgans para incorporar información de distancia genética. Los datos se filtraron considerando una frecuencia alélica mínima de 0.001, un umbral máximo de 30% de datos faltantes por individuo y 60% por SNP. Los parámetros de IBD se estimaron con el paquete `isoRelate` (Browning & Browning, 2013; Henden et al., 2018). Los segmentos IBD se identificaron aplicando como criterios un mínimo de 15 SNPs consecutivos, una longitud ≥ 3 kb y una tasa de error de 0.001.

Con el fin de caracterizar patrones de relación entre muestras, los segmentos IBD detectados se transformaron en grafos de conectividad mediante el paquete `igraph` (Csardi & Nepusz, 2006). Se construyeron redes con distintos umbrales de similitud genética (10%, 50% y 90%), lo que permitió explorar conexiones desde relaciones

más laxas hasta vínculos altamente compartidos. En cada red, los vértices representaron muestras individuales y las aristas los segmentos IBD compartidos según el umbral definido. Se asignaron atributos de metadatos (ubicaciones geográficas que coinciden con la estructuración genética) a los vértices, lo que permitió visualizar la distribución espacial de los clústeres genéticos. Los subgrafos correspondientes a Loreto y Amazonas se analizaron por separado para evaluar patrones locales de conectividad.

Las redes de transmisión se construyeron a partir del árbol filogenético previamente inferido del conjunto completo de SNPs bialélicos. El árbol (formato Newick) se importó en StrainHub (<https://strainhub.io/>), donde se asignaron como metadatos a las unidades geográficas. StrainHub estimó la direccionalidad y flujo entre localidades mediante algoritmos de reconstrucción de estados ancestrales y generó un grafo dirigido con métricas de centralidad (Source/Hub Ratio, betweenness y closeness) (De Bernardi Schneider et al., 2020). Las redes resultantes se exportaron en formato RDS y junto con las métricas de centralidad se importaron en R para su visualización geográfica de fuente/receptor de infecciones y distribución de las métricas por provincia usando el paquete ggplot2.

Otro enfoque para evaluar la conectividad genética es a través del número de migrantes efectivos por generación (N_m), calculado en `divMigrate` del paquete `diveRsity` en R (Keenan et al., 2013; Sundqvist et al., 2016). Este estadístico se derivó de la diferenciación genética entre poblaciones con el estadístico G_{st} (Nei, 1973), lo que permitió inferir la magnitud y la direccionalidad del flujo génico. Los valores de N_m se representaron como redes dirigidas ($N_m > 0.1$), donde los nodos correspondieron a provincias, y las aristas indicaron el sentido y magnitud relativa

del flujo. Finalmente, estos resultados se integraron en un mapa geográfico, para la visualización tanto la dirección como la intensidad del intercambio genético entre provincias de Amazonas y Loreto.

Por último, la asignación de origen geográfico de las muestras disponibles se realizó utilizando el barcode GEO33 (Trimarsanto et al., 2022). A partir del conjunto filtrado de los SNPs bialélicos, se extrajeron las coordenadas correspondientes a estos marcadores y se prepararon en formato compatible para su análisis en la plataforma vivaxGEN GEO (<https://geo.vivaxgen.org/>). En esta plataforma se implementa un algoritmo de clasificación supervisada que compara los genotipos del barcode con una base de datos global de referencia, asignando la muestra a la región más probable de origen. Los resultados incluyeron probabilidades de asignación para cada país de referencia, los cuales fueron exportados para la visualización en R usando el paquete ggplot2.

5.7. Análisis estadísticos

Todos los análisis estadísticos se realizaron en R Studio (v2024.4.2.764) utilizando R (v4.2.1). Para la heterocigosidad esperada (H_e) entre unidades de análisis se aplicó la prueba de Kruskal–Wallis (bilateral). Cuando el contraste global resultó significativo ($p < 0.05$), se efectuaron comparaciones por pares mediante Wilcoxon rank-sum, con ajuste del p -value mediante el método de Benjamini–Hochberg, el cual controla la tasa de falsos descubrimientos, reportando valores p ajustados para cada comparación.

Para la diversidad nucleotídica (π), se verificó previamente la homogeneidad de varianzas con la prueba de Levene; dado que no se asumió normalidad, los

contrastes entre unidades de análisis se realizaron con Kruskal–Wallis, y las comparaciones post-hoc se llevaron a cabo con el test de Dunn aplicando corrección de Bonferroni. En todas las pruebas inferenciales se adoptó un nivel de significancia de $\alpha = 0.05$ (bilateral). Las implementaciones se realizaron con rstatix (Kassambara, 2023a) y las figuras incluyeron anotaciones de significancia sobre box/violin plots.

VI. RESULTADOS

6.1. Características de calidad y procedencia de las muestras filtradas

Los parámetros completos de secuenciación y filtrado para las 478 muestras se detallan en los Anexos 2, 3, 4 y 5, diferenciados por tipo de secuenciamiento y región de procedencia.

Del total de secuencias genómicas analizadas, el 87.4% (402/478) cumplieron con los criterios de calidad establecidos, dentro del cual se consideraron algunos parámetros como la profundidad de cobertura ($>10\times$) y el porcentaje de genotipos faltantes ($<35\%$).

El número de secuencias se representó gráficamente en base a la ubicación geográfica (provincia / distrito) y por periodos de muestreo definidos en base a las estrategias de intervención contra la malaria: 2006 – 2010 (PAMAFRO), 2011 – 2016 (periodo sin estrategias de intervención), 2017 – 2022 (PMC) y 2022 – 2023 (PNEM) (Fig. 3).

Se observa que el mayor número de secuencias genómicas se concentró en la provincia de Maynas ($n = 240$), provenientes de siete distritos endémicos, entre los cuales Mazán ($n = 93$) y San Juan Bautista ($n = 89$) fueron los más representativos. Mazán posee secuencias en los cuatro periodos definidos ($n_{2006-2010} = 11$, $n_{2011-2016} = 32$, $n_{2017-2021} = 35$ y $n_{2022-2023} = 15$), mientras que San Juan Bautista solo en los dos primeros periodos ($n_{2006-2010} = 57$, $n_{2011-2016} = 32$).

Las siguientes provincias con mayor representación fueron Loreto y Condorcanqui. En Loreto, la data disponible proviene exclusivamente del distrito de Trompeteros en un solo periodo ($n_{2017-2021} = 68$). En Condorcanqui (región Amazonas), las

secuencias provienen de dos distritos endémicos ($n = 68$), el distrito de Nieva con registros únicamente en el último periodo ($n_{2022-2023} = 13$) y Río Santiago en los dos últimos periodos ($n_{2017-2021} = 7$, $n_{2022-2023} = 48$). Esto evidencia la ausencia de datos genómicos provenientes de la región de Amazonas en los periodos previos.

Por otro lado, se debe señalar que las provincias de Ramón Castilla y Datem del Marañón estuvieron subrepresentadas. En Ramón Castilla, las secuencias provinieron de dos distritos endémicos, con en mayor aporte del distrito de Yavarí ($n_{2017-2021} = 19$), mientras que el distrito de Ramón Castilla presentó solo 3 secuencias ($n_{2006-2010} = 1$, $n_{2017-2021} = 2$). Finalmente, la provincia del Datem del Marañón registró únicamente cuatro secuencias en el distrito de Morona, distribuidas en los dos primeros periodos ($n_{2006-2010} = 2$, $n_{2011-2016} = 2$).

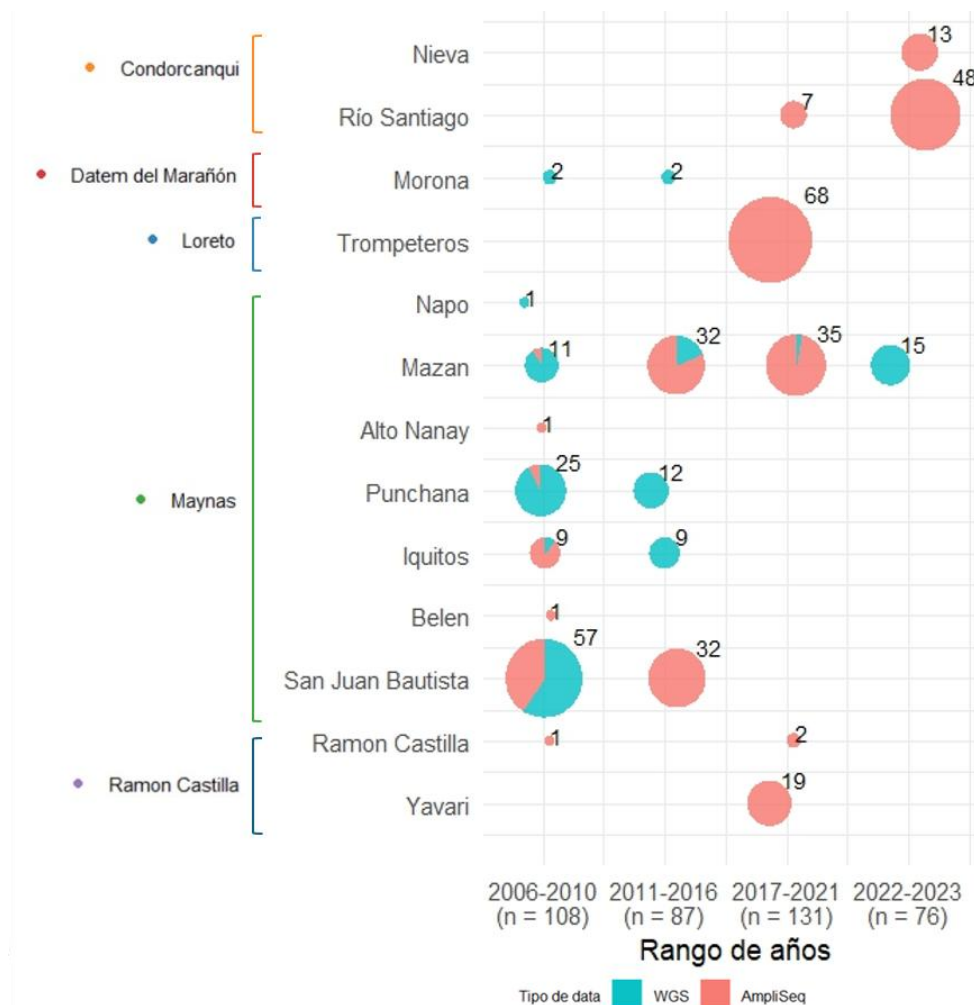


Figura 3. Distribución espacio-temporal de las muestras incluidas en el análisis genómicos. Según el tipo de data, ubicación geográfica (provincia y distrito) y los periodos definidos en base a las estrategias de intervención contra la malaria (2006-2023).

En general, la mayor proporción de muestras corresponde a la provincia de Maynas (60%), seguida de las provincias de Condorcanqui (17%) y Loreto (17%), mientras que Ramon Castilla (5%) y Datem del Marañón (1%) poseen porcentajes menores con respecto al total de secuencias analizadas (Fig. 4).

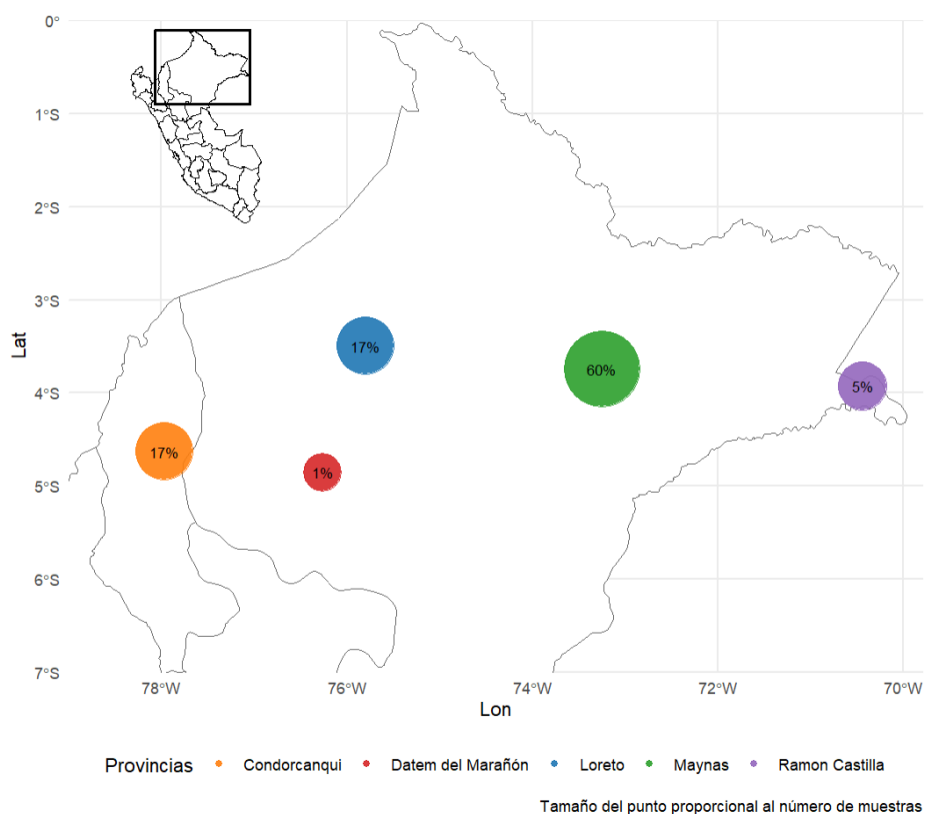


Figura 4. Distribución geográfica de la data filtrada. Distribución en porcentajes de las muestras genómicas analizadas en la región de Loreto y Amazonas.

6.2. Estructuración Genética Poblacional

Con el fin de evaluar la estructuración genética poblacional de todas las muestras analizadas, así como la capacidad resolutoria de los distintos conjuntos de SNPs del panel *Pv AmpliSeq v2 Peru*, a continuación se compararon los resultados obtenidos a partir del conjunto completo de SNPs bialélicos (que incluye marcadores neutrales y bajo selección) con aquellos derivados del barcode *ITM-Peru*.

Primero, analizando la estructuración en base a distancias genéticas, los árboles filogenéticos construidos con el conjunto completo de SNPs bialélicos (Fig. 5A) y el barcode *ITM-Peru* (Fig. 5B) presentaron una topología general congruente, es decir, una distribución similar de los clados principales. Como evidencia de ello, en

ambos árboles, en términos de patrones geográficos, se observó un clado bien soportado correspondiente a las muestras de Condorcanqui, así como otro clado diferenciado que agrupa principalmente a los aislados de Ramón Castilla. Además, los niveles de soporte bootstrap de los nodos principales de cada clado son comparables (Bootstrap > 85).

Por otro lado, las muestras de Loreto y Maynas estuvieron ampliamente dispersas en diferentes clados en ambos árboles, indicando la primera evidencia de la alta diversidad genética y mezcla entre estas poblaciones procedientes de ambas provincias. Los aislados del Datem del Marañón, con menor representación numérica, aparecieron intercalado dentro de claros representados por la provincia de Maynas, indicando proximidad genética entre ambas provincias.

En términos de patrones temporales, no se observa un agrupamiento por periodos, mas que en algunos clados específicos, principalmente los de Condorcanqui y Ramón Castilla, dado que los aislados de dichas provincias fueron principalmente muestrados en los periodos de 2022-2023 y 2017-2021, respectivamente. En general, estos patrones indican que la estructura genética, particulamente la agrupación de poblaciones por provincia, es consistente independiente del número de SNPs utilizados.

Sin embargo, se notaron algunas diferencias en la resolución filogenética y en las distancias genéticas. Aunque las longitudes de rama difirieron entre ambos árboles, esta comparación debe interpretarse con cautela, ya que las escalas de distancia son independientes y no directamente comparables.

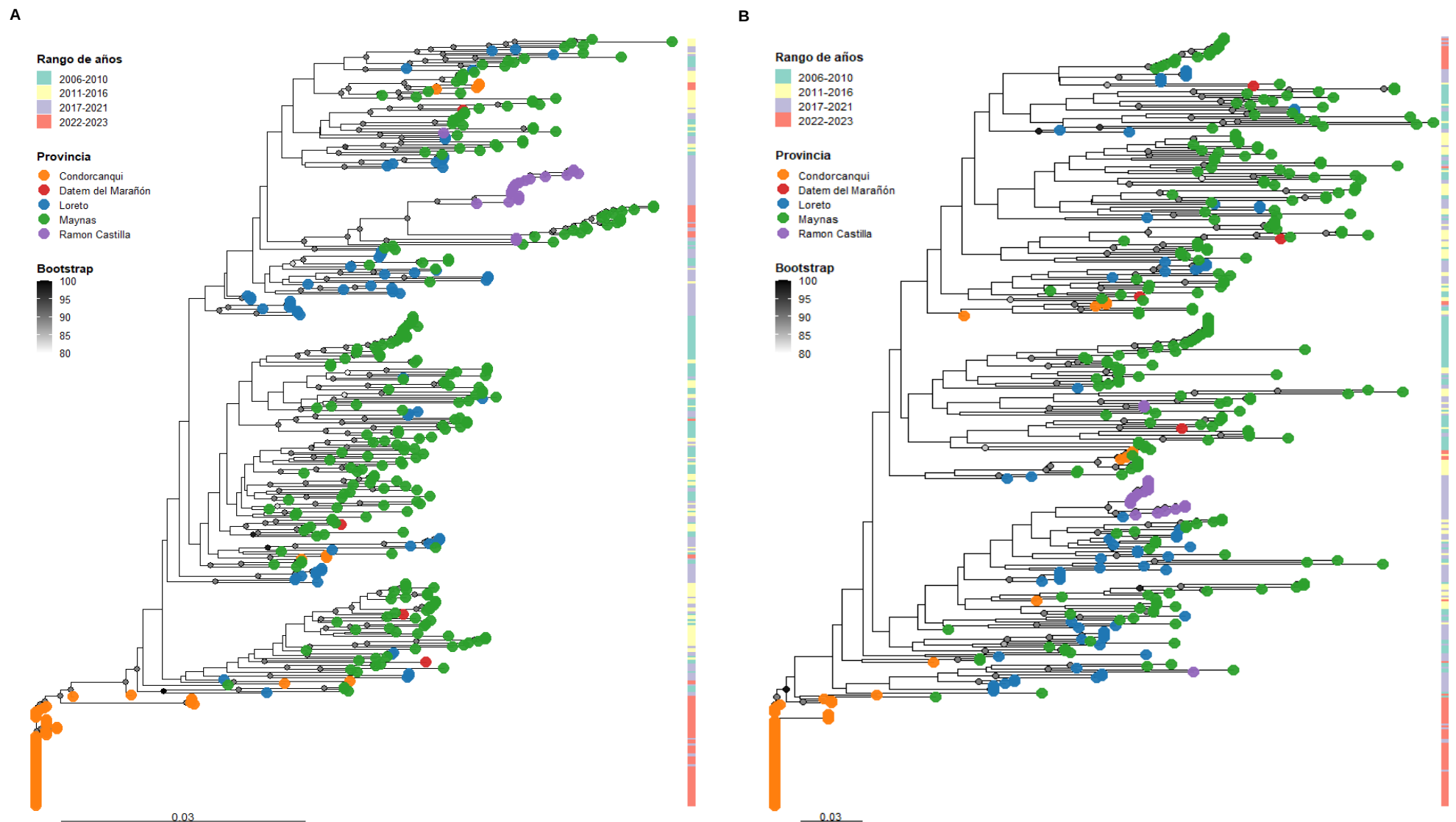


Figura 5. Árboles filogenéticos de Neighbour-Joining (NJ). Mostrando los aislamiento de acuerdo por provincia, rango de años y nivel de soporte bootstrap contruidos partir de los 802 SNPs bialélicos (A) y el barcode ITM (B).

Los resultados de PCA basados en el conjunto completo de SNPs bialélicos mostraron una clara separación de las muestras de Condorcanqui, las cuales formaron un cluster aislado (color naranja) en los extremos positivos de ambos componentes principales (PC1 y PC2). Además, se observó una agrupación parcial de las muestras de Ramon Castilla en un cluster aislado (color morado) en el extremo negativo de PC1 y positivo de PC2 (Fig. 6A). En cambio, las muestras restantes mostraron un grado de solapamiento en la zona central del gráfico, evidenciando una menor diferenciación genética entre las demás provincias y los distintos periodos de muestreo.

Con el objetivo de analizar las muestras sobrelapadas, se realizó un PCA únicamente con aquellas muestras sobrelapadas (Fig. 6B). Este análisis permitió identificar solo un clúster aislado (color verde), correspondiente principalmente a las muestras de Maynas del periodo 2022-2023, agrupado en los extremos negativos de los componentes principales. Fuera de este agrupamiento, no se observaron patrones de estructuración claramente definidos que coincidiera con la ubicación o periodo.

Por otro lado, los PCA utilizando el barcode *ITM-Peru* evidenciaron una menor estructuración genética. En el análisis que incluyó todas las muestras, solo se observó una separación parcial correspondiente a Condorcanqui (cluster naranja), es decir, no incluye a todos los aislados de dicha provincia; además, no se observó clusterización de los aislados de Ramón Castilla tal cual se visualizó en los análisis usando el conjunto completo de SNPs (Fig. 6C).

Analizando exclusivamente las muestras sobrelapadas, el cual incluye las muestras

de las provincias de Loreto, Maynas, Ramón Castilla y Datem del Marañón, no se identificó una estructuración definida que coincidiera con las ubicaciones geográficas o periodo debido a la dispersión entre las muestras (Fig. 6D).

Los análisis de componentes principales evidencian una mayor resolución para detectar estructuración genética usando el conjunto completo de SNPs bialélicos en comparación con el barcode *ITM-Peru*. Además, los clústeres identificados con el conjunto completo de SNPs mostraron una concordancia con la ubicación geográfica de los aislados, principalmente agrupados por provincias tales como Condorcanqui y Ramón Castilla, y en algunos casos, aparentemente con los periodos de muestreo, sin embargo, no se puede realizar una afirmación concluyente de dicha separación temporal dado la heterogeneidad del muestreo por tiempo y ubicación.

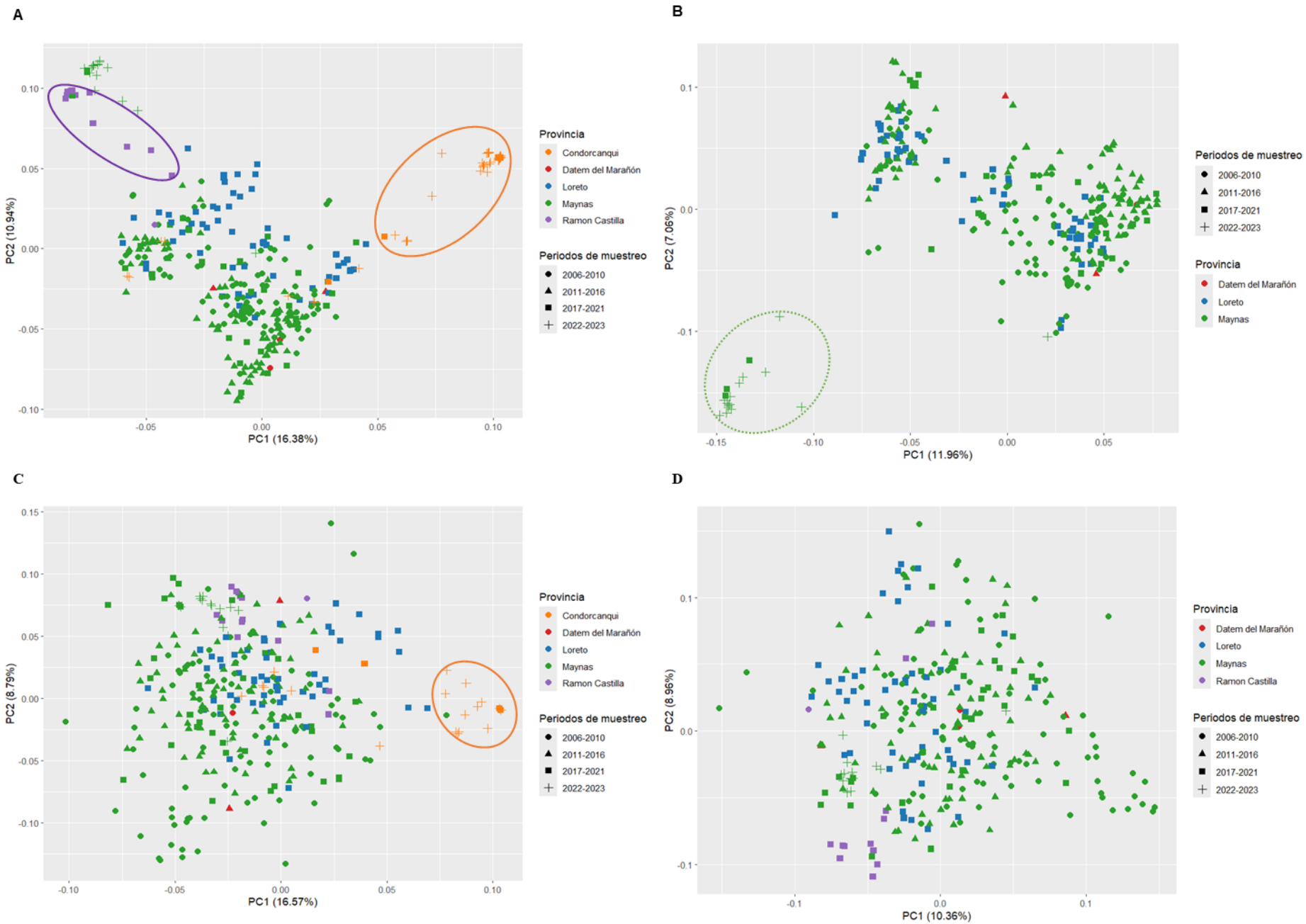


Figura 6. Análisis de Componentes Principales (PCA) por provincia y periodos de muestreo definidos. Patrones de estructuración genética de todas las muestras analizadas (A) y de las muestras provenientes de las provincias del Datem del Marañón, Loreto y Maynas (B) utilizando el conjunto completo de SNPs bialélicos. Asimismo, los patrones de estructuración genética de todas las muestras (C) y de las procedentes de las provincias de la región de Loreto (D) utilizando el barcode *ITM-Peru*. En todos los paneles, cada muestra está representada por provincia y periodos de muestreo, mediante color y por forma, respectivamente.

Dado los resultados del PCA derivados de los SNPs bialélicos y la concordancia de la clusterización por ubicaciones geográficas, se realizó los análisis de discriminación de componentes principales (DAPC) basados en una agrupación *a priori* por provincias y/o distrito, con el fin de maximizar sus diferencias genéticas y sondear la estructuración de los clústeres.

Los primeros análisis mediante DAPC se realizaron utilizando el conjunto completo de SNPs bialélicos. Primero se utilizaron todos los aislados con una agrupación *a priori* por provincias. Se evidenciaron la presencia de 4 clústeres principales, los cuales indican un estructuración genética coherente a la distribución a nivel de provincia, más no a nivel temporal (Fig. 7A).

Dentro de los clústeres formados se tiene de las poblaciones de Ramón Castilla, un clúster morado agrupado en los extremos negativos de los componente LD1 y LD2; Condorcanqui, clúster naranja agrupado en los extremos positivos de ambos componentes; Maynas, clúster verde, y Loreto, clúster azul, ambos clústeres muy cercanos agrupados en los extremos positivos de los componentes, mientras que los aislados de Datem del Marañón están distribuidos en dentro del clúster de Maynas.

Además, se evidencia que tanto los clústeres correspondientes a Condorcanqui y Ramón Castilla presentan una mayor separación genética en comparación con

Maynas y Loreto, ambos detallados por el 51.4% y 32.18% de diferenciación entre poblaciones, lo cual indica una buena representación para integrar los datos acorde a su procedencia, en este caso, específicamente a nivel de provincia (Fig. 7A).

La visualización por distrito de este DAPC demuestra que no hay estructuración genética entre los distritos de Condorcanqui, pero si dentro de los distritos de Ramón Castilla (Anexo 6A). Dado que los aislados de Yavarí están clusterizados en los extremos negativos del LD1 en comparación con los aislados del distrito Ramón Castilla, ambos pertenecientes a la provincia de Ramón Castilla.

También, se realizó el DAPC de los aislados de la provincia de Maynas con una agrupación *a priori* a nivel de distritos. Se optó por obviar aquellos que tienen solo aislado por distrito. Así la distribución de los aislados de la provincia por Maynas no evidencian una estructuración genética a nivel de distritos, sin embargo, es importante notar que los aislados de los distritos de San Juan Bautista y Punchana son los más separados en el primer componente LD1 (56.38). El solapamiento con el distrito de Mazán sugiere que no hay una estructuración genética completa entre distritos (Fig. 7B).

Los análisis posteriores de DAPC se elaboraron mediante el uso del barcode *ITM-Peru*. En el DAPC utilizando todos los aislados y con agrupaciones *a priori* por provincia, se observa que no hay una mayor resolución para detectar estructuración genética a nivel de ubicación o tiempo. Sin embargo, cabe resaltar que, se detectaron la clusterización de los aislados de la provincia Ramón Castilla en el primer componente LD1, mientras que Condorcanqui presenta un agrupamiento parcial en el segundo componente LD2, parcialmente sobrelapado con los aislados

de Maynas y Loreto (Fig. 7C). Estos últimos patrones descritos coinciden con una clusterización a nivel de distrito, puesto que los aislados del distrito de Yavarí difieren genéticamente del resto de aislados al usar el barcode *ITM-Peru* (Anexo 6C).

Posteriormente, se realizó el DAPC de los aislados de las provincias con dispersiones similares, el cual incluye a Loreto, Maynas y Datem del Marañón, se agruparon *a priori* por distrito. Aunque se observa una separación de algunos aislados del distrito de Trompeteros, provincia Loreto, se mantiene un solapamiento con las muestras de Maynas. Por ello se concluye que bajo este enfoque no se encontró patrones de estructuración genética que este delimitada por tiempo y/o ubicación debido al nivel resolutivo propio del panel barcode *ITM-Peru* (Fig. 7D).

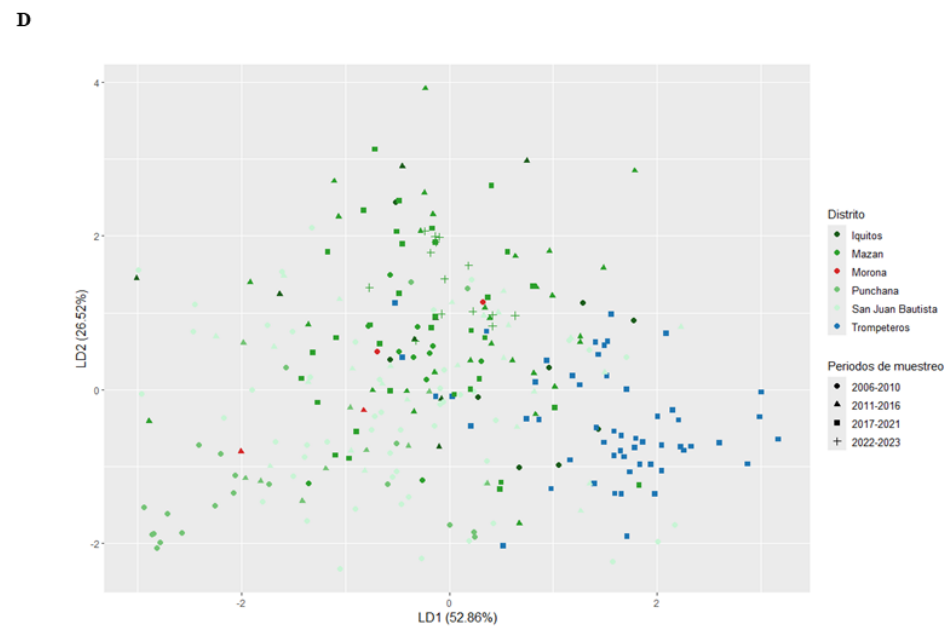
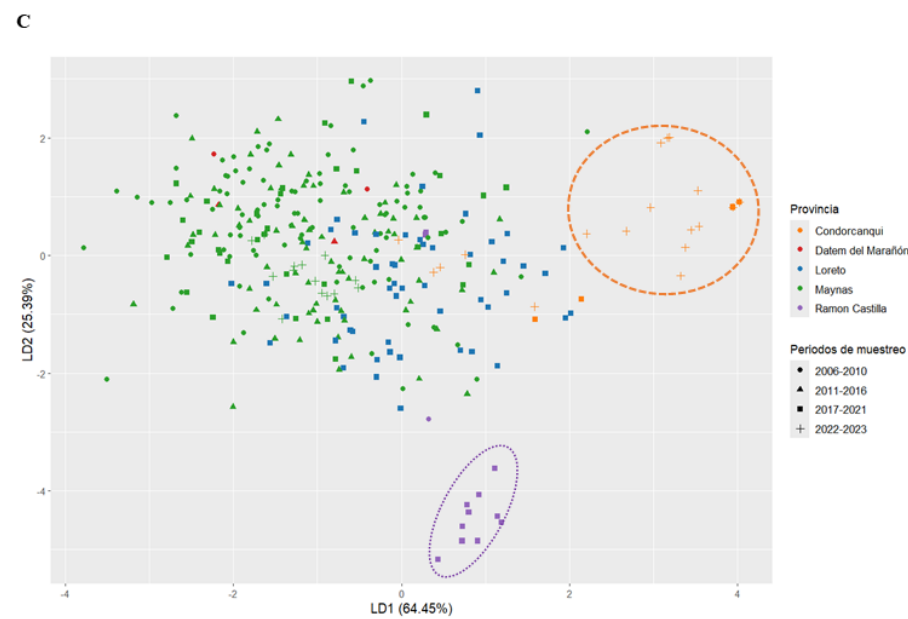
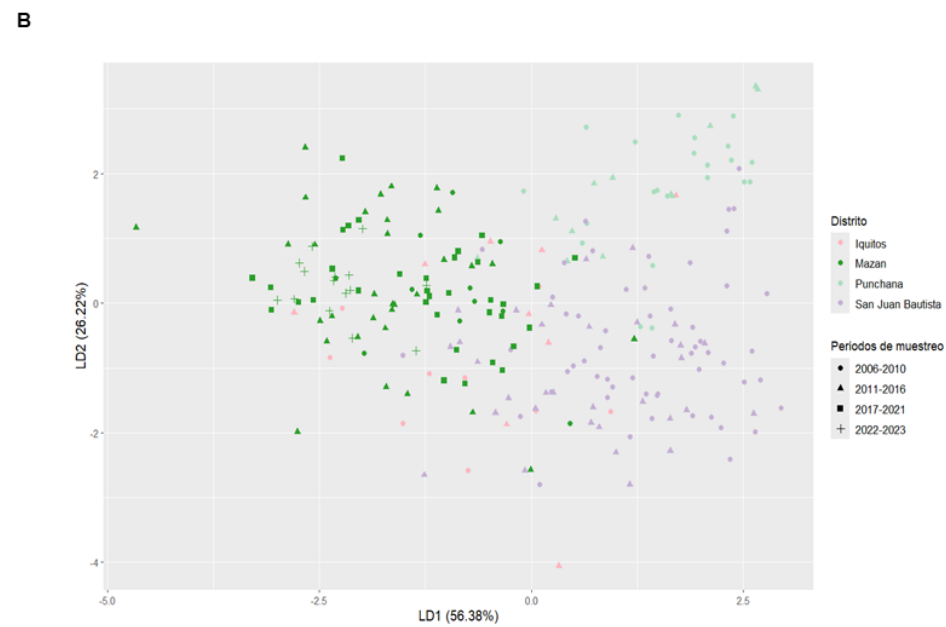
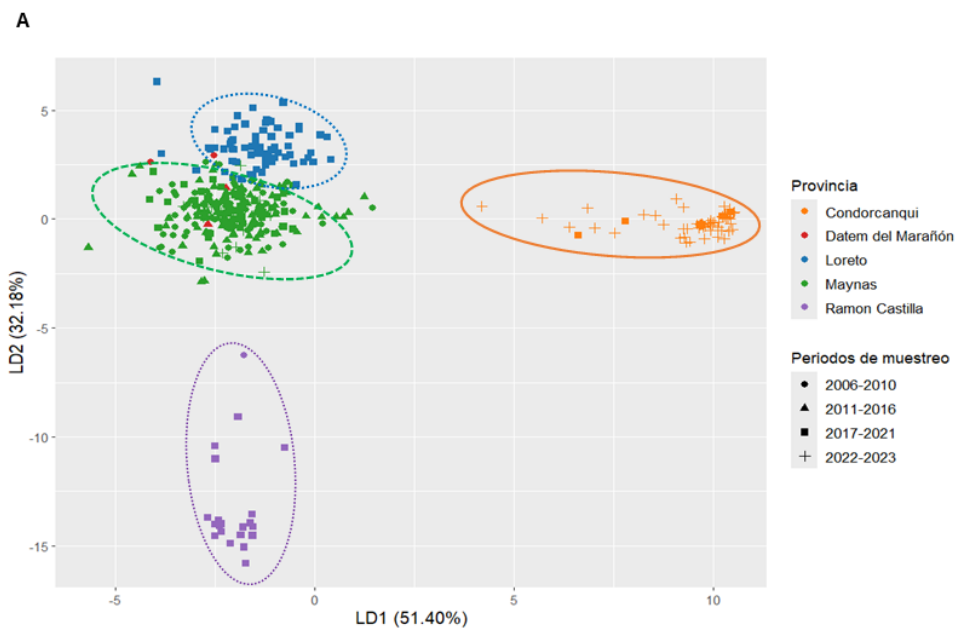


Figura 7. Análisis Discriminante de Componentes Principales (DAPC) agrupados a priori por ubicación geográfica. Patrones de estructuración genética de todas las muestras agrupadas *a priori* por provincias (A) y de las muestras de la provincia de Maynas agrupadas *a priori* por distrito (B) ambos análisis utilizando el conjunto completo de SNPs. Patrones de estructuración genética de todas las muestras agrupadas *a priori* por provincias (C) y de las muestras de las provincias de Maynas, Loreto y Datem del Marañón agrupadas *a priori* por distritos (D), ambos análisis utilizando el barcode del *ITM-Peru*. En todos los paneles, cada muestra está representada por provincia y periodos de muestreo, mediante color y por forma, respectivamente.

Además se realizaron sondeos adicionales para verificar si existe estructuración genética a nivel temporal en base a los periodos definidos en este estudio usando solo el conjunto completo de SNPs dado a su mayor resolución genética. Con los análisis de DAPC realizados con agrupaciones *a priori* en base a los periodos de muestreo se evidenció que no hay una estructuración genética parcial, dado la formación de un cluster que representa a las muestras del 2022-2023, agrupados en los extremos negativos de los componentes; y la formación de otro clúster que representa a las muestras 2006-2010, ambas agrupadas en el extremo positivo de LD1 y negativo de LD2 (Anexo 7A).

Aunque ambos clústeres están diferenciados dentro del primer componente, se debe señalar que el mayor porcentaje de aislados procedentes del periodo 2022-2023 provienen de la provincia de Condorcanqui y que el mayor porcentaje (aproximadamente el 90%) provienen de la provincia del Maynas, indicando que, las separaciones de ambos clústeres sea producto principalmente de la ubicación más que del periodo de muestreo.

Adicional, a partir del PCA construido solo con las muestras de los periodos 2011-2016 y 2017-2021, se realizó el análisis DAPC con agrupaciones *a priori* por

periodos de muestreo (Anexo 7B). Los resultados indicaron que no existe estructuración genética entre ambos periodos, por lo que se sugiere que no hay evidencia suficiente para obtener inferencias significativamente biológicas a partir de las agrupaciones por periodos de muestreo; sin embargo, con la finalidad de poder describir la data en su totalidad, se incluyeron en los análisis sin derivar conclusiones poblacionales a partir de ellas.

El distrito de Mazán presenta aislados secuenciados en los 4 periodos definidos. Este subgrupo muestral representa el único ejemplo para realizar análisis de cambios temporales en un lugar geográfico-específico. Los análisis de PCA evidenciaron que los aislados muestrados en el periodo 2022-2023 mantienen una diferenciación genética agrupados en los ejes negativos de ambos componentes; más no se encontraron estructuración genética dentro de los otros periodos (Anexo 8A). Al maximizar las diferencias en grupos a priori por periodo de muestreo, se observó que las muestras del distrito de Mazán se estructuran genéticamente de acuerdo a los periodos temporales (Anexo 8B). Dentro de estos agrupamientos se observó que tanto los periodos de 2006-2010 y 2022-2023 mantienen una estructuración más separada en comparación a los otros periodos, mientras que los periodos 2011-2016 y 2017-2021 mantienen clústeres parcialmente separados.

Luego de validar la estructura genética nivel de provincia, resultados respaldados por los algoritmos de agrupamiento (PCA y DAPC), se realizaron análisis complementarios para caracterizar la dinámica de dichas poblaciones.

El análisis de varianza molecular jerárquico (AMOVA), es decir a nivel de provincia y regiones, reveló patrones de estructuración genética consistentes entre

los dos conjuntos de SNPs analizados (Tabla 8). Sin embargo, la magnitud de la variación explicada por los distintos niveles jerárquicos difirió entre el conjunto completo de SNPs bialélicos y el barcode *ITM-Peru*.

En el análisis basado en el conjunto completo de SNPs, se observó que la mayor proporción de la variación genética se encuentra dentro de las provincias (68.4%), seguida por una fracción considerable entre provincias dentro de las regiones (16.2%) y en menor medida, entre regiones (15.4%). Los índices de fijación jerárquicos fueron significativos para los niveles regional ($\Phi_{CT} = 0.3157$, $p = 0.001$) y provincial ($\Phi_{SC} = 0.1914$, $p = 0.001$), indicando una diferenciación genética moderada tanto entre regiones como entre provincias, mientras que el valor total ($\Phi_{ST} = 0.1538$) no fue estadísticamente significativo ($p = 0.423$).

Por otro lado, el análisis realizado con el barcode *ITM-Peru* mostró una tendencia similar, aunque con una menor magnitud de variación. La variación dentro de provincias representó el 71.7% del total, mientras que las fracciones entre provincias y entre regiones fueron de 8.8% y 19.5%, respectivamente. De manera análoga, los índices de fijación Φ_{CT} (0.2835) y Φ_{SC} (0.1094) resultaron significativos ($p = 0.001$), en tanto que el valor total Φ_{ST} (0.1955) no mostró significancia ($p = 0.208$).

En base a la capacidad del conjunto completo de SNPs bialélicos, se aplicó el modelo no jerárquico de AMOVA para evaluar la estructura temporal de las muestras agrupadas por periodo de muestreo, estrictamente con finalidades descriptivas más que interpretativas (Anexo 9). Aquí se observó que la mayor parte de la variación genética se encuentra dentro de los rangos de años (84.6%), mientras

que un 15.4% de la variación se atribuye a diferencias entre los rangos de años. El componente de varianza correspondiente a la variación temporal tiende a ser significativo ($\Phi = 0.1541$, $p = 0.001$), lo que se podría indicar como una diferenciación genética moderada a lo largo del tiempo; sin embargo, en base a lo descrito en los análisis de PCA y DAPC, esta diferenciación puede estar sesgada debido a que en cada agrupación temporal no incluye muestras de los mismos sitios de estudio.

Adicionalmente, basándonos en la alta resolución por el conjunto de los SNPs bialélicos se realizó el análisis de ancestría genética mediante ADMIXTURE (Fig. 8). Los resultados de *cross-validation* (Anexo 10) no permitieron identificar el valor óptimo K, dado que el error de validación disminuyó progresivamente con el incremento del K, mostrando una mayor reducción hasta K=5, tras lo cual las disminuciones fueron menores. Esto sugiere que las muestras analizadas poseen una estructura compleja y continua, sin límites genéticos estrictos a las ubicaciones geográficas. Ante la ausencia de un K óptimo definido, se seleccionaron los valores de K igual a 5 (Fig. 8A) y 6 (Fig. 8C) por presentar una mejor correspondencia biológica y menores errores de validación significativos.

Para K=5, se observaron clústeres genéticos diferenciados asociados principalmente a provincias específicas, especialmente entre Condorcanqui y Ramón Castilla, mientras que Loreto, Maynas y Datem del Marañón mostraron una composición mixta de ancestrías (Fig. 8A). En K=6, la aparición de un clúster adicional permitió discriminar subestructuras dentro de Loreto y Maynas, evidenciando una mayor resolución genética (Fig. 8C).

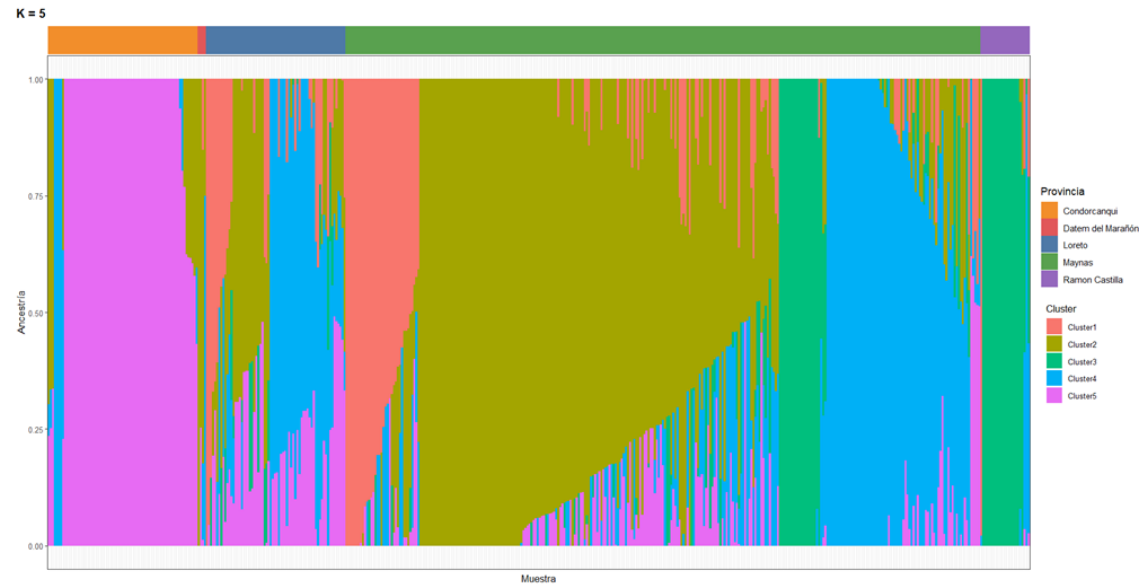
Al evaluar las proporciones de los clústeres genéticos a lo largo del tiempo (Fig. 8B, 8D), se observó una notable persistencia en la composición de ancestrías dentro de las mismas provincias, aunque con variaciones en su frecuencia relativa. Entre ellos, destaca la presencia de un clúster predominante en la provincia de Condorcanqui durante ambos periodos de muestreo (2017-2021 y 2022-2023). Este mismo cluster ancestral, también se observó en periodos anteriores en Datem del Marañón, Loreto y Maynas, aunque en menor proporción, sugiriendo una posible introducción o dispersión de un linaje ancestral desde Loreto a Amazonas.

Así mismo, en Ramón Castilla se observó que un clúster se mantiene y aumenta en proporción después de un intervalo de tiempo sin muestreos, lo que podría reflejar la persistencia y/o expansión local de un linaje dominante. En el caso de Maynas, independientemente del valor de K, los clústeres de ancestría se mantienen estables desde el 2006 hasta el 2021; sin embargo, uno de ellos aumenta en proporción en el último periodo, lo cual sugiere una posible expansión clonal de un linaje específico dentro de Maynas, el cual comparte con Ramón Castilla.

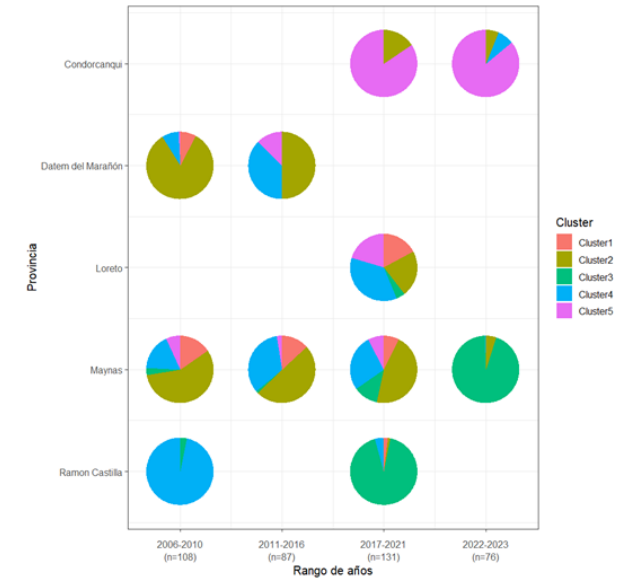
Tabla 8. Análisis de Varianza Molecular jerárquico (AMOVA) por agrupaciones geográficas usando todos los SNPs bialélicos y el barcode ITM.

Por agrupaciones geográficas (AMOVA jerárquico) - Usando todos los SNPs bialélicos							
Fuente de Variación	Df	Suma de Cuadrados	Media de Cuadrados	Componente de varianza	Porcentaje de variación	Indices de fijación	P-value
Entre Regiones	1	882.0125095	882.0125095	4.192056	15.37682	$\Phi_{CT} = 0.3157$	0.001
Entre Provincias dentro de las regiones	3	701.4379016	233.8126339	4.414886	16.19418	$\Phi_{SC} = 0.1914$	0.001
Dentro de Provincias	397	7406.130988	18.65524178	18.655242	68.42901	$\Phi_{ST} = 0.1538$	0.423
Total	401	8989.581399	22.41790872	27.262184	100		
Por agrupaciones geográficas (AMOVA jerárquico) - Usando el barcode ITM							
Fuente de Variación	Df	Suma de Cuadrados	Media de Cuadrados	Componente de varianza	Porcentaje de variación	Indices de fijación	P-value
Entre Regiones	1	248.9847	248.984737	1.5920343	19.548981	$\Phi_{CT} = 0.2835$	0.001
Entre Provincias dentro de las regiones	3	122.2875	40.762487	0.7166866	8.800371	$\Phi_{SC} = 0.1094$	0.001
Dentro de Provincias	397	2316.5354	5.835102	5.8351019	71.650649	$\Phi_{ST} = 0.1955$	0.208
Total	401	2687.8076	6.702762	8.1438228	100		

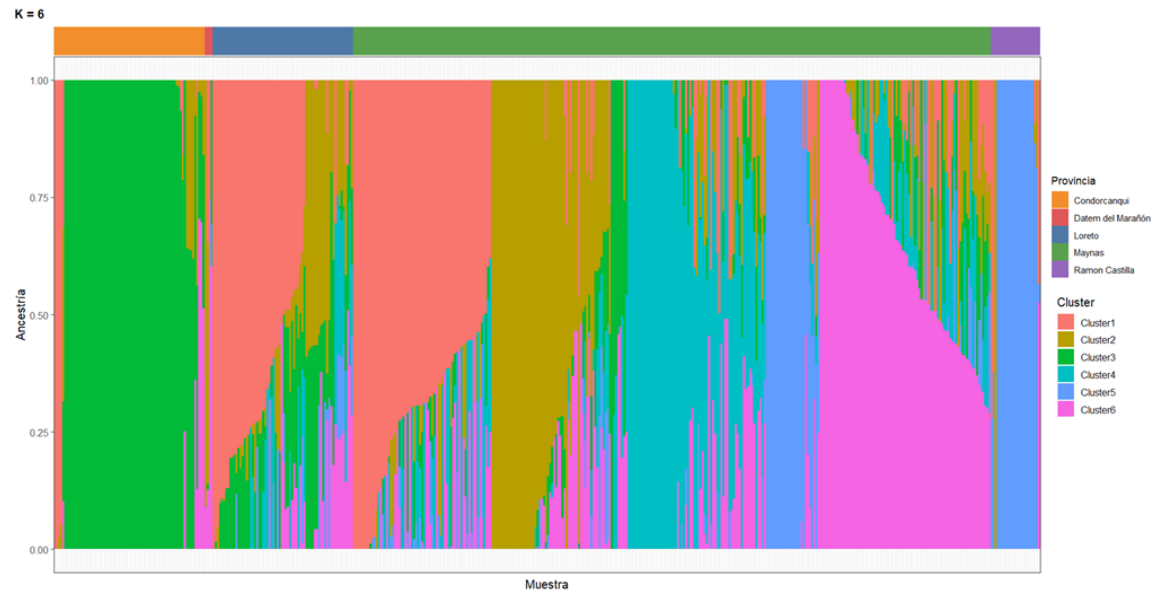
A



B



C



D

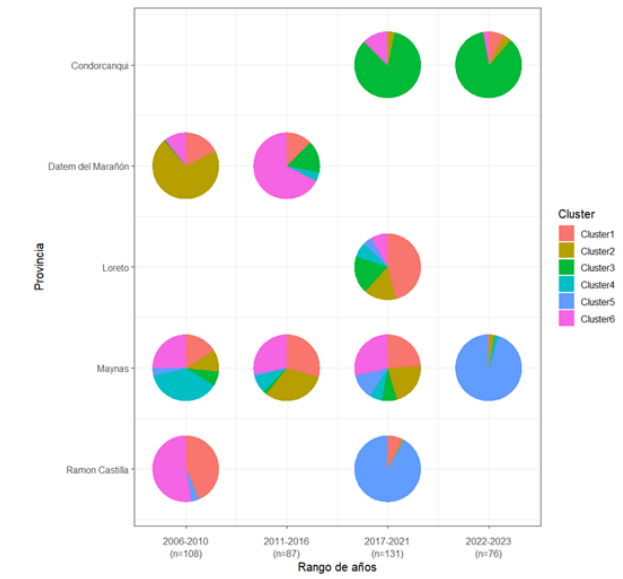


Figura 8. Estructura genética de las muestras de Amazonas y Loreto inferido mediante análisis de ADMIXTURE. Plots de asignación individual de ancestría para $K = 5$ (A) Y $K = 6$ (C). Distribución de las proporciones de los clústeres por provincia y rango de años correspondientes a $K = 5$ (B) y $K = 6$ (D).

Los valores de F_{IS} obtenidos a partir del conjunto completo de SNPs bialélicos (Fig. 9A) evidenciaron una mayor dispersión entre provincias, variabilidad que puede atribuirse a que este conjunto incluye SNPs neutrales y bajo presión selectiva. Las provincias de Condorcanqui y Datem del Marañón presentaron los valores cercanos a 1, evidenciando la baja diversidad clonal, Maynas y Ramón Castilla tienden hacia valores cercanos a cero, indicando exceso relativo de infecciones policlonales, estado que no coincide con los valores encontrados en el F_{WS} para la provincia de Ramón Castilla.

En contraste, los valores estimados con el barcode *ITM-Peru* (Fig. 9B) mostraron menor variabilidad entre provincias, con distribuciones más homogéneas y centradas hacia valores positivos, indicando la baja diversidad clonal. En particular, Condorcanqui y Ramón Castilla poseen valores medios de 0.95, sugiriendo una mayor proporción de muestras monoclonales en estas provincias.

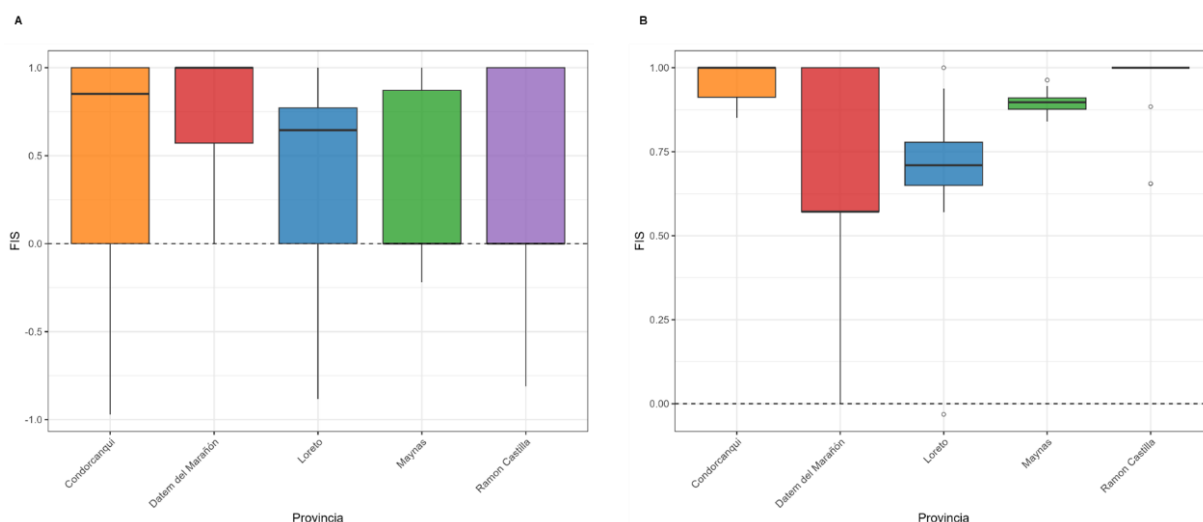
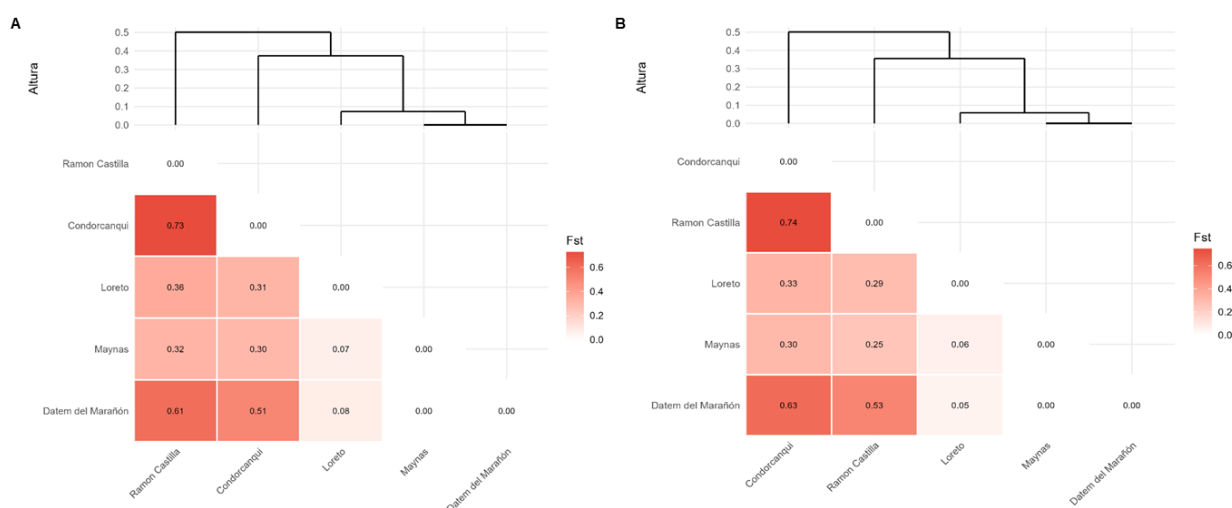


Figura 9. Distribución del índice de endogamia (F_{IS}) distribuido por provincia. Los valores calculados a partir del conjunto completo de SNPs bialélicos (A) y los SNPs del barcode *ITM-Peru* (B).

Los valores de F_{ST} entre los grupos definidos por provincias, estimados a partir del conjunto completo de SNPs bialélicos (Fig. 10A) y el barcode *ITM-Peru* (Fig. 10B), no mostraron diferencias significativas en las comparación por pares, y se mantuvieron en un rango estrecho de variación (0.01 a 0.05). En ambos conjuntos de SNPs bialélicos se observa un patrón consistente en los dendogramas; donde las provincias de Maynas, Datem del Marañón y Loreto forman un clado. Sin embargo, existen diferencias en la posición de Condorcanqui entre ambos conjuntos de datos. Cuando se utiliza el conjunto completo de SNPs bialélicos (Fig. 10A), Condorcanqui aparece más próximo al clado formado por Loreto-Maynas-Datem del Marañón. En contraste, al emplear el conjunto reducido del barcode *ITM-Peru* (Fig. 10B), Condorcanqui se ubica como la provincia más divergente, mientras que todas las provincias de la región de Loreto se agrupan dentro de un mismo clado.

Según los valores de F_{ST} , para ambos conjuntos de datos (Fig. 10), las provincias de Datem del Marañón y Maynas no presentan diferenciación genética entre sí ($F_{ST} = 0$). La provincia de Loreto mantiene una baja diferenciación con respecto a las

dos provincias antes mencionadas ($F_{ST} = 0.05 - 0.08$). La provincia de Ramón Castilla exhibe una alta diferenciación en relación a Maynas y Loreto ($F_{ST} = 0.30 - 0.36$), y muy alta diferenciación respecto a Datem del Marañón ($F_{ST} = 0.61 - 0.63$). Por último, la provincia de Condorcanqui presenta diferenciación muy alta con el resto de provincias, destacando su marcada divergencia con Datem del Marañón ($F_{ST} = 0.51 - 0.53$) y a Ramón Castilla ($F_{ST} = 0.73 - 0.74$). Particularmente, se observa que las provincias de Ramón Castilla y Condorcanqui son poblaciones extremadamente diferenciadas entre sí.



Así mismo, se observa que, usando ambos paneles de SNPs, los distritos más divergentes son Nieva y Río Santiago (provincia de Condorcanqui), y Yavarí (provincia Ramón Castilla) en comparación con el resto de distritos. Los valores de F_{ST} estimados en ambos paneles tampoco se observa diferencias significativas (Anexo 6B y 6D).

De manera descriptiva, se estimaron los valores de F_{ST} en las agrupaciones temporales utilizando el conjunto completo de SNPs (Anexo 7C). Las muestras de

los periodos 2006-2010, 2011-2016 y 2017-2021 mantienen una baja diferenciación genética ($F_{ST} = 0.06 - 0.08$) y se agruparon dentro del mismo clado en el dendograma. Mientras que las muestras del periodo 2022-2023 es un grupo muy diferenciado con el resto de periodos ($F_{ST} = 0.20 - 0.31$).

Por último, en el análisis temporal de las muestras de Mazán se observó que las muestras del 2017-2021 y 2011-2016 mostraron una baja diferenciación genética ($F_{ST} = 0.01$), en cambio las comparaciones de ambos periodos con el rango 2006-2010 demuestra un mayor nivel de diferenciación genética ($F_{ST} = 0.03 - 0.04$). En cambio las comparaciones con los muestreos del 2022-2023 representan una alta diferenciación genética ($F_{ST} = 0.56 - 0.39$) (Anexo 8C).

6.3. Estimaciones de la Diversidad Genética

Con finalidad de realizar interpretaciones poblacionales, los siguientes análisis también se realizaron principalmente con los grupos a nivel de provincias, agrupaciones respaldadas por los análisis de estructuración genética. Los niveles de diversidad genética se estimaron mediante dos parámetros: 1) La heterocigocidad esperada (H_e) utilizando los SNPs neutrales del barcode *ITM-Peru* y hacer comparable los valores con otros estudios; y 2) la diversidad nucleotídica (π) a partir de todas las regiones del panel *Pv Ampliseq v2 Peru* para mejorar la estimación de la diversidad aprovechando un mayor número de sitios. Cada violín representa la densidad de probabilidad de los datos, mientras que el diagrama de cajas interna indica la mediana, cuartiles y valores extremos.

El análisis de la H_e a nivel regional/provincial mostró diferencias significativas entre varias de las comparaciones. A nivel de regiones, Loreto presentó mayor

diversidad genética ($H_e = 0.339$), mientras que Amazonas mostró un valor considerablemente menor ($H_e = 0.065$), con una diferencia altamente significativa ($p_{\text{value}} < 0.001$) (Fig. 11A).

Al analizar por provincias, los mayores valores de H_e se observaron en Maynas ($H_e = 0.351$), seguida de Datem del Marañón ($H_e = 0.295$) y Loreto ($H_e = 0.277$). En contraste, las provincias de Ramón Castilla ($H_e = 0.090$) y Condorcanqui ($H_e = 0.065$) evidenciaron niveles notablemente bajos de heterocigosidad ($p_{\text{value}} < 0.001$ en todas las combinaciones pareadas). En particular, la provincia de Maynas presentó valores significativamente más altos de H_e en comparación con Ramón Castilla y Condorcanqui, al igual que Datem del Marañón y Loreto, cuyos niveles de diversidad genética fueron también superiores respecto a Ramón Castilla y Condorcanqui (Fig. 11B).

Al analizar la H_e por rango de años (Anexo 7D), se observa que los valores son relativamente mayor en los periodos 2006-2010 ($H_e = 0.374$) y 2011-2016 ($H_e = 0.366$), seguidos del periodo 2017-2021 ($H_e = 0.285$). Posteriormente, se evidencia una reducción marcada y estadísticamente significativa en el periodo 2022-2023 ($H_e = 0.056$), lo que sugiere una aparente disminución en la variabilidad genética. Sin embargo, es importante señalar que las muestras provienen de diferentes provincias para cada periodo, y el 19.7% de las muestras del periodo 2022-2023 provienen de la región de Loreto, por lo que la disminución de la H_e no necesariamente refleja con exactitud la dinámica histórica de la variabilidad genética de *P. vivax* en el tiempo.

En el caso particular del análisis temporal de las muestras de Mazán se observó que los valores de la diversidad genética (H_e) se mantiene entre los primeros 3 periodos

de muestro ($H_e=0.35 - 0.36$) hasta el periodo 2022-2023 donde se observó una disminución de la diversidad genética ($H_e=0.05$). Las diferencias significativas entre el periodo 2022-2023 con respecto al resto de periodos concuerdan con los patrones de estructuración genética encontrados (Anexo 8D).

El análisis de la diversidad nucleotídica (π) a nivel regional (Fig. 11C) evidencia diferencias significativas, siendo la región de Loreto ($\pi = 8 \times 10^{-4}$) con mayor diversidad en comparación con Amazonas ($\pi = 4 \times 10^{-4}$) ($p_{\text{value}} < 0.001$). A nivel provincial (Fig. 11D), la provincia del Datem del Marañón ($\pi = 14 \times 10^{-4}$; todas comparaciones por pares con $p_{\text{value}} < 0.001$) y Loreto ($\pi = 10 \times 10^{-4}$; comparaciones por pares con Condorcanqui, Maynas y Ramón Castilla y los $p_{\text{value}} < 0.001$, $p_{\text{value}} < 0.01$ y $p_{\text{value}} < 0.05$, respectivamente) presentan los valores más altos de diversidad; mientras que Maynas ($\pi = 6 \times 10^{-4}$) presenta valores de diversidad intermedia. En contraste, Ramón Castilla ($\pi = 4 \times 10^{-4}$) y Condorcanqui ($\pi = 4 \times 10^{-4}$) registran los valores más bajos. Las comparaciones múltiples indican que Datem del Marañón y Loreto poseen diferencias significativas frente a las otras provincias, principalmente con las de menor diversidad.

Con respecto al análisis de π por rango de años (Anexo 7E), se observa que el periodo 2006-2010 ($\pi = 5 \times 10^{-4}$) tiene los valores más bajos, mostrando diferencias significativas con los periodos 2011-2016 ($\pi = 8 \times 10^{-4}$) y 2017-2021 ($\pi = 9 \times 10^{-4}$). Este patrón de incremento de diversidad aparentemente parece mantenerse en el siguiente periodo 2022-2023 ($\pi = 7 \times 10^{-4}$), aunque con una ligera disminución. Sin embargo, este patrón no refleja la variación histórica de los parásitos, debido a la heterogeneidad en el número de muestras entre regiones/provincias y periodos.

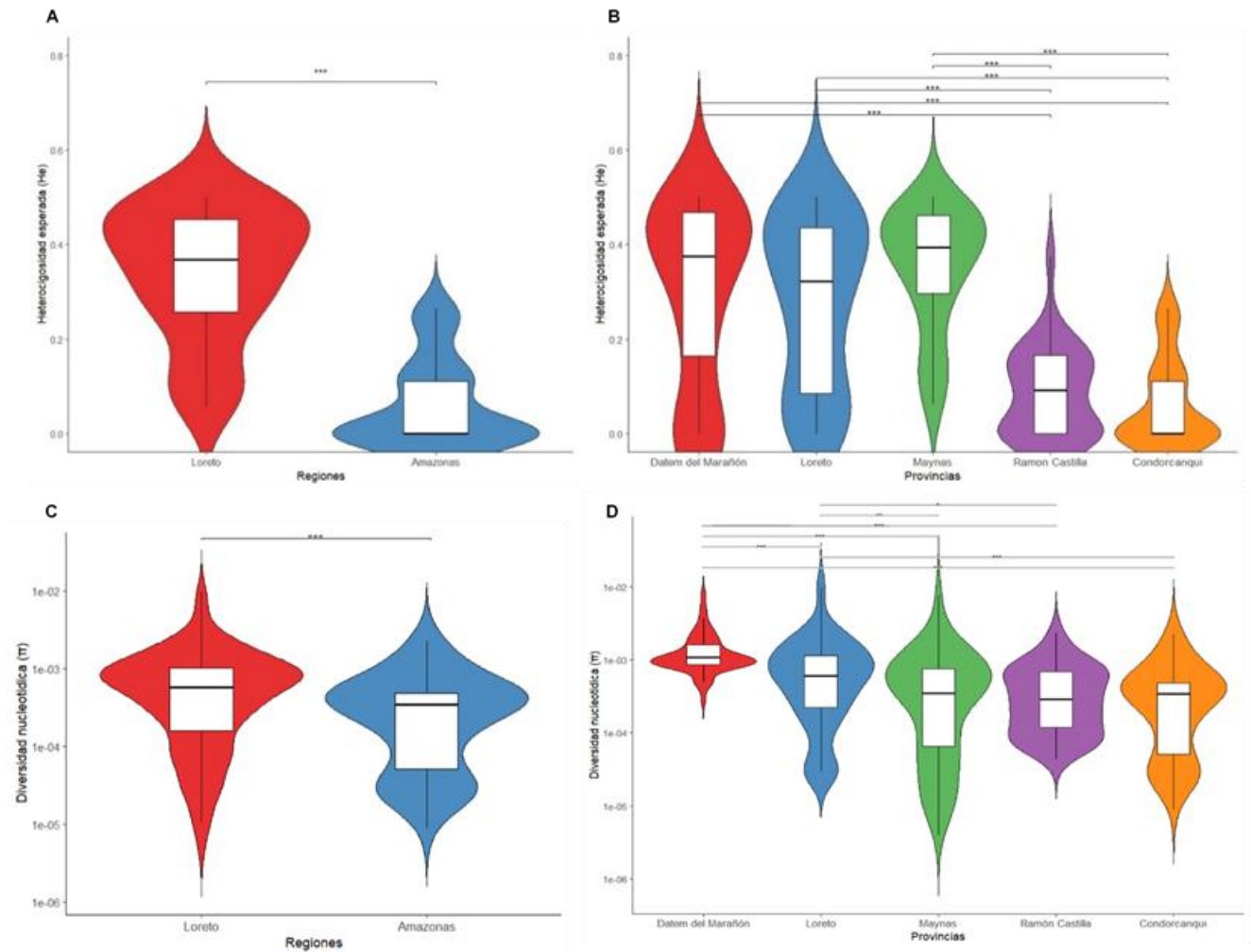


Figura 11. Medidas de la diversidad genética de las muestras *P. vivax* de Amazonas y Loreto. Se muestran las comparaciones de la heterocigosidad esperada (H_e) entre regiones (A) y provincias (B), así como de la diversidad nucleotídica (π) entre regiones (D) y provincias (E). Las barras horizontales indican las comparaciones significativas entre grupos, donde * $p < 0.05$; ** $p < 0.01$ y *** $p < 0.001$.

La complejidad de infección (COI) se estimó a partir del índice F_{WS} , considerando como infecciones monoclonales aquellas con valores de $F_{WS} \geq 0.95$ y como infecciones policlonales aquellas con $F_{WS} < 0.95$. En total, el 16.17% de las muestras analizadas correspondieron a infecciones policlonales. Entre las provincias evaluadas, Datem del Marañón y Loreto presentaron las mayores proporciones de infecciones policlonales (0.25 y 0.35, respectivamente), mientras

que Condorcanqui y Ramón Castilla mostraron proporciones bajas (0.03 y 0.14, respectivamente), lo cual concuerda con los menores niveles de diversidad genética observados en estas zonas (Fig. 12A). Por su parte, Maynas registró una proporción intermedia (0.15) de muestras policlonales en comparación a las otras proporciones encontradas. Sin embargo, dado a que la provincia de Maynas concentra el mayor número de muestras, representa el grupo con más infecciones policlonales.

Estos resultados indican que la mayoría de las muestras presentan valores altos de F_{WS} , lo que evidencia un predominio de infecciones monoclonales. No obstante, en Loreto y Maynas se identificaron múltiples casos con valores de F_{WS} menores a 0.95, reflejando la presencia de infecciones policlonales (Fig. 12B).

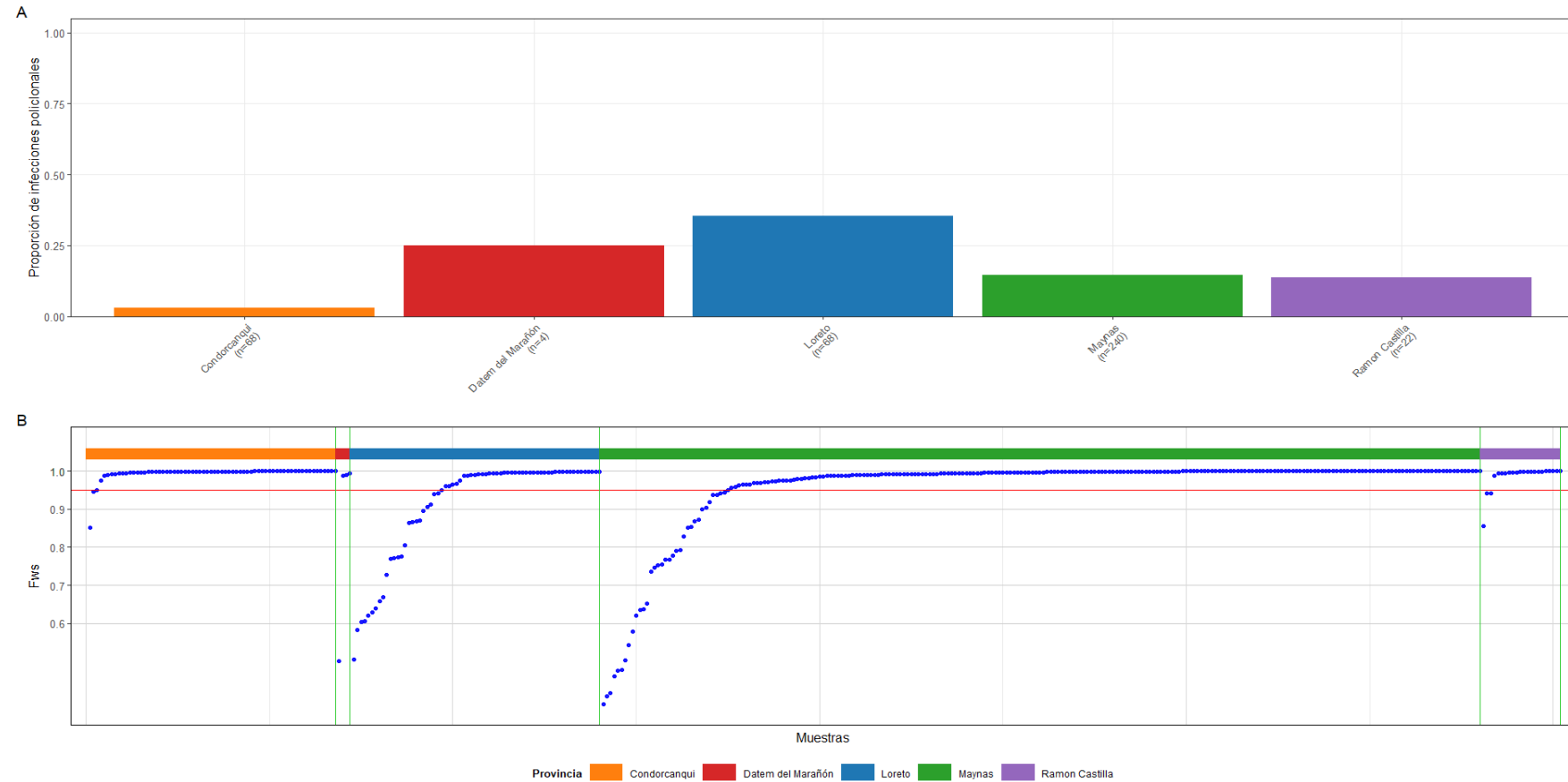


Figura 12. Complejidad de infección en muestras de *P. vivax* de las regiones de Amazonas y Loreto. Proporción de infecciones policlonales por provincia (A) y distribución del coeficiente de fijación F_{ws} , las líneas verdes flanquean las agrupaciones por provincia, la línea roja representa el umbral $F_{ws}=0.95$, cada punto azul representa a un aislado ordenados de menos a mayor valor de F_{ws} (B).

6.4. Conectividad Genética

Los análisis de estructuración genética revelaron las unidades de análisis se pueden agrupar a nivel de provincia, además se sugiere una posible introducción de linajes parasitarios desde Loreto a Amazonas, seguido de expansiones clonales geográficamente aisladas. Para confirmar esta hipótesis y determinar si existe flujo genético entre estas provincias endémicas, se usaron el conjunto completo de SNPs bialélicos para análisis de conectividad en cuatro enfoques complementarios: 1) la conectividad genética basados en la identidad por descendencia (IBD); 2) la inferencia de relaciones direccionales de conectividad a partir de los patrones de árboles filogenéticos; 3) la cuantificación de la conectividad genética de los aislados agrupados por provincias mediante la estimación del número de migrantes por generación; y finalmente 4) la predicción del origen geográfico de las muestras de *P. vivax* agrupadas en provincias.

Primero, la identidad por descendencia (IBD) estima el porcentaje del genoma que es compartido entre dos individuos como resultado de la herencia de un ancestro común reciente. Así, se usa este parámetro para medir el grado de parentesco entre aislamientos y como primer enfoque para detectar los niveles de conectividad genética entre poblaciones. Con un umbral de IBD al 90% (Fig. 13A), se observó que las redes se fragmentan en pequeños clústeres de muestras genéticamente idénticas. Estos agrupamientos reflejan la presencia de linajes clonales dentro de provincias específicas. Se observa que hay diferentes clústeres de linajes clonales en las provincias de Loreto y Maynas. En contraste, en las provincias de Condorcanqui y Ramón Castilla se observa linajes clonales únicos para cada provincia, lo que sugiere la persistencia local de clones dominantes. Aquí es

importante destacar que solo un individuo de Condorcanqui fue casi indentico a cinco individuos de Maynas.

Al reducir el umbral a IBD al 50% (Fig. 13B), se mide la conectividad poblacional intermedia debido al parentesco parcial compatible con el intercambio genético entre subpoblaciones y/o persistencia de linajes comunes. Se observó que incrementa la conectividad entre provincias, observándose grupos interprovinciales y relaciones genéticas más amplias, indicativas de intercambio genético moderado. Así, se observa que individuos de las provincias de Loreto y Maynas se conectan principalmente. Además, se observa que las provincias de Condorcanqui posee una conectividad en menor medida con las provincias de Loreto y Maynas. Además, se observó una baja conectividad entre Ramón Castilla y Maynas. Aún flexibilizando los umbrales de IBD, se mantienen los clústeres dominantes correspondientes a Condorcanqui y Ramón Castilla.

Por último, al aplicar un umbral del IBD = 10% (Fig. 13C), la red se volvió más densa, integrando la mayoría de las muestras en un gran cluster. Este patrón sugiere la existencia de una base genética compartida a nivel regional, donde las poblaciones de *P. vivax* mantienen linajes ancestrales comunes entre provincias. Siguiendo los mismos procedimientos, se realizaron los análisis a nivel de distritos (Anexo 11). Aquí se observó que los distritos de Río Santiago y Nieva (Condorcanqui) forman un solo cluster predominante, al igual que el distrito de Yavarí (Ramón Castilla).

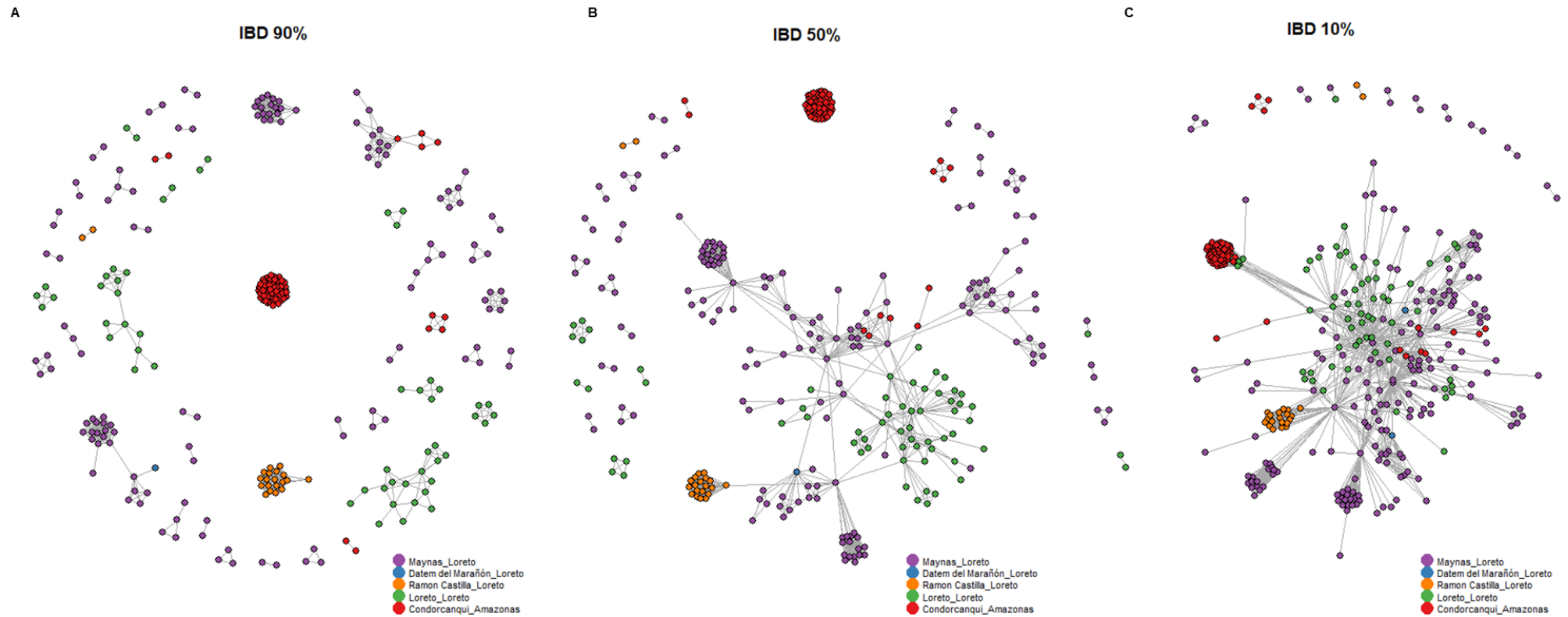


Figura 13. Redes de conectividad genética de *P. vivax* basadas en la Identidad por Descendencia (IBD) identificados por provincia/región. Redes con umbral de IBD igual a 90% (A), 50% (B) y 10% (C). Cada nodo representa un aislamiento y las conexiones representa el par de muestras que comparten la proporción umbral.

El segundo enfoque basado en análisis de redes de transmisión potencial estimados en StrainHub permitió identificar los patrones de conectividad entre provincias amazónicas.

En el gráfico de burbujas (Fig. 14A), construido a partir de las métricas de conectividad (Anexo 12), revelaron diferencias marcadas en los roles de cada provincia dentro de la red. El grado de salida (outdegree, eje X), que refleja el número de conexiones desde una provincia hacia otras, fue mayor en Maynas y Loreto, indicando su función predominante como fuentes de transmisión. El grado de entrada (indegree, eje Y), que representa la cantidad de conexiones recibidas, fue más alto en Condorcanqui y Loreto, lo que sugiere sus papeles principalmente receptores. Es notable que, según este enfoque, la provincia de Loreto tendría un papel simultáneo como fuente y receptora de transmisión con Maynas. El Source Hub Ratio ($SHR = \frac{outdegree}{indegree}$) mostró valores elevados (0.8) en Maynas, confirmando su predominancia como nodo emisor, mientras que en las otras provincias de Ramón Castilla y Dátem del Marañón presentaron valores de 0, confirmando su función como nodos receptores. En cuanto a la centralidad de intermediación (betweenness), que mide cuántas veces un nodo conecta con otros nodos en la red, Maynas destacó con los valores más altos (4), reflejando su rol como punto de conexión entre múltiples rutas de transmisión.

En el mapa de transmisiones potenciales basados en el Source Hub Ratio (Fig. 14B), adaptado de los redes propias de StrainHub (Anexo 13), los patrones espaciales coincidieron con las métricas de conectividad. Se evidenciaron flujos bidireccionales intensos desde Maynas y Loreto, así como flujos unidireccionales

desde Maynas al Datem del Marañón. Además, se identificaron conexiones secundarias entre Loreto y Condorcanqui en ambas direcciones, junto con una conexión unidireccional de Maynas hacia Condorcanqui. Por otro lado, las provincias de Ramón Castilla y Datem del Marañón mostraron un número reducido de eventos de intercambio y un menor peso de conexión (según el SHR), evidenciando su función como nodos extremos terminales de la red.

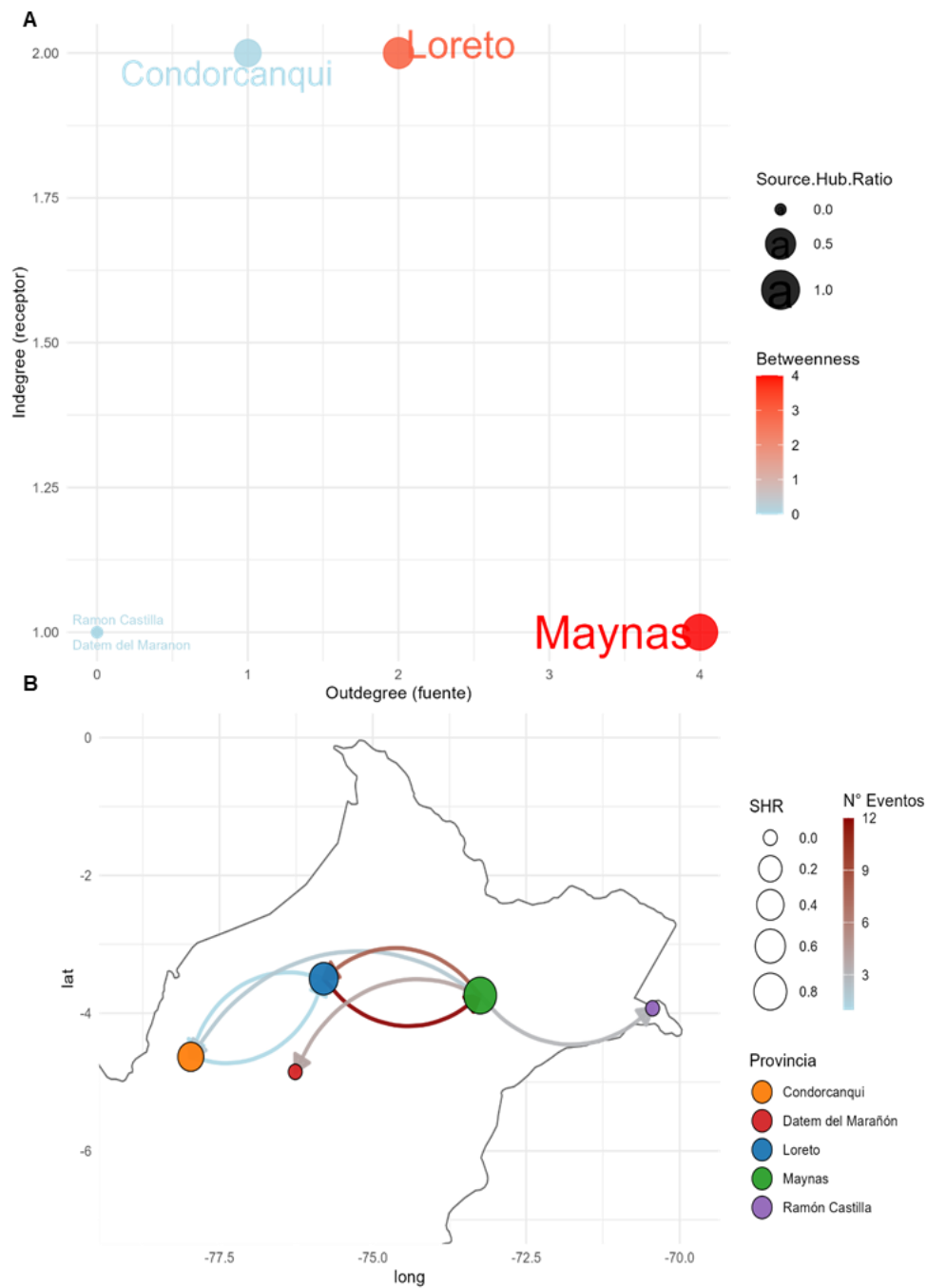


Figura 14. Redes de transmisión de *P. vivax* basados en las métricas de conectividad inferidas por StrainHub. Diagrama de dispersión de métricas de conectividad entre provincias (A). Mapa de transmisiones potenciales entre provincias basados en la métrica Source Hub Ratio (B).

Ambos resultados revelan que los patrones de transmisión potencial están dominados por flujos que emergen principalmente desde Maynas, actuando como foco de redistribución hacia provincias aledañas, generando así un patrón de transmisión provincial jerarquizado.

El tercer enfoque basado en el número de migrantes efectivos por generación (N_m), mide el grado de flujo genético entre poblaciones. Este parámetro (N_m) mide el número promedio de individuos que migran y se reproducen con éxito en una población diferente en cada generación. Valores mayores a 1 indican un flujo genético elevado lo suficiente para contrarrestar la deriva génica; menores a 1, flujo genético restringido y diferenciación entre poblaciones (Slatkin, 1987).

El análisis del flujo génico entre agrupaciones reveló una baja conectividad genética entre provincias. Solo se mostraron los valores de N_m mayores a 0.1, lo que sugiere diferentes grados de intercambio genético entre provincias, incluso en niveles limitados de conectividad (Fig. 15). En general, el mapa sugiere que la migración entre provincias es insuficiente para homogenizar las poblaciones, pero sí lo suficiente para evitar una diferenciación genética completa.

El mayor flujo genético se observó un flujo génico entre Maynas y Condorcanqui, con valores de $N_m \approx 1$ en dirección a Condorcanqui. Esto sugiere la posibilidad de un evento de introducción desde Maynas a Condorcanqui. No obstante, Maynas mostró una conectividad extremadamente baja con Datem del Mararón ($N_m \approx 0.20$). Este patrón respalda que Maynas actúa como nodo central de conectividad genética; sin embargo, los bajos valores de N_m entre provincias evidencian que la conectividad genética es muy restringida entre áreas separadas por más de 300 km

de distancia, más aun en provincias que no están conectadas por las mismas cuencas fluviales.

Por otro lado, Condorcanqui mostró valores menores ($N_m \approx 0.20$) en su conexión con Ramón Castilla, lo que refleja un aislamiento genético relativo de las poblaciones del norte de Amazonas respecto a las del oriente de Loreto, las provincias más alejadas en este estudio. Asimismo, Ramón Castilla presentó flujos débiles y direccionales principalmente hacia Maynas ($N_m \approx 0.20$), lo que implica que probablemente Ramón Castilla tenga flujo génico asimétrico, es decir, entrada de linajes externos desde zonas más conectadas, como las fronteras con Colombia y Brasil.

La provincia del Datem del Marañón presentó conexiones bidireccionales con Condorcanqui y Maynas, sin embargo, el tamaño muestral de Datem del Marañón podría estar generando sesgos en este análisis. Por último, la provincia de Loreto destacó principalmente por tener solo una conexión muy baja con Condorcanqui ($N_m \approx 0.20$), indicando la ausencia de flujo genético entre ambas provincias.

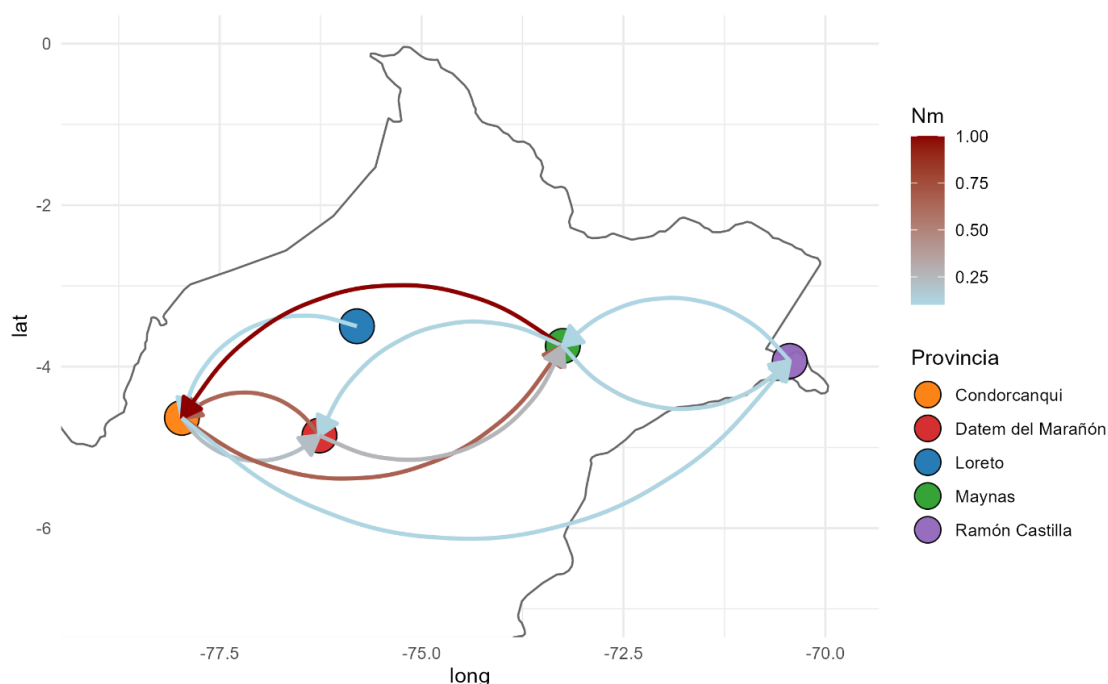


Figura 15. Conectividad genética entre provincias de Loreto y Amazonas basados en el número de migrantes por generación (Nm). Los valores de Nm fluctúan entre 0.1 a 1.

El último enfoque, se basa en la predicción del origen geográfico basado en los 33 SNPs validados internacionalmente del barcode GEO33 (Trimarsanto et al., 2022). La predicción de origen de las infecciones de *P. vivax* resultó en 3 países: Brasil, Colombia y Perú (Fig. 16). Se observa que la mayoría de las muestras de Condorcanqui fueron asignadas con alta probabilidad a Colombia (0.809). Por otro lado, las muestras de Datem del Marañón muestran la mayor probabilidad de origen asociada a Brasil (0.75). Las provincias de Loreto y Maynas exhiben una mayor proporción de asignaciones internas hacia Perú (0.529–0.537), aunque con contribuciones mixtas de Brasil y Colombia. Finalmente, las muestras de Ramón Castilla se asignan en su totalidad a Brasil (1.0), reflejando una fuerte similitud genética con poblaciones brasileñas adyacentes.

La complementariedad de estos resultados evidencian que las poblaciones parasitarias de Condorcanqui y Ramón Castilla están genéticamente aisladas de las otras provincias de la región de Loreto. Además del bajo parentesco parasitario y las redes potenciales de transmisión, los valores de migración genética (N_m) sugiere un evento de introducción único desde la región de Loreto (provincia Maynas) a la región de Amazonas (provincia Condorcanqui). Por otro lado, en base a los resultados de predicción de origen, los aislados de Ramón Castilla podrían estar influenciada por flujos genéticos de poblaciones parasitarias provenientes de países como Brasil.

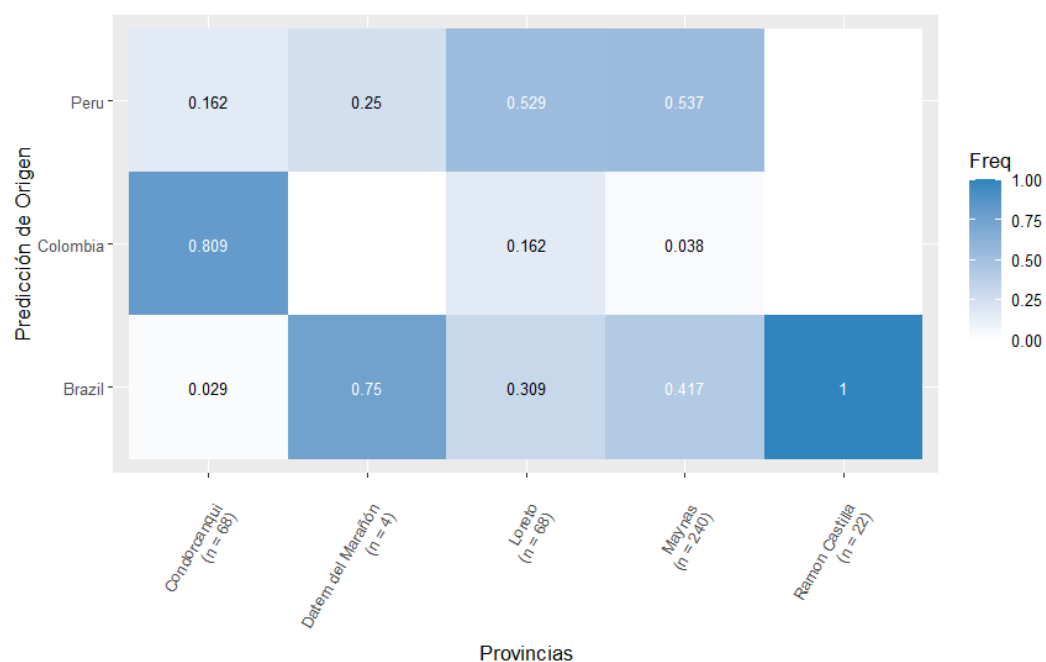


Figura 16. Predicción de origen geográfico de las muestras de *P. vivax* por provincia. Frecuencias de predicción utilizando los SNPs del barcode GEO33.

VII. DISCUSIÓN

La Amazonía peruana presenta una alta carga de pacientes con síndromes febriles de etiología patogénica diversa. En ese contexto, la malaria continua siendo la infección predominante, con mayor prevalencia de *Plasmodium vivax*, y se mantiene como un problema prioritario de salud pública (Sanchez et al., 2025). En el Perú, la mayor parte de los casos de malaria se concentran en Loreto y Amazonas, a pesar de la puesta en marcha del Plan hacia la Eliminación de la Malaria 2022–2030 del MINSA (MINSA, 2022). Con el aumento de infecciones en los últimos años (MINSA, 2026), persisten interrogantes sobre los mecanismos que permiten la dispersión de *P. vivax* hacia comunidades remotas y la existencia de flujos genéticos entre áreas alejadas. Así, la vigilancia genómica emerge como una herramienta clave para comprender la estructura y flujo genético de las poblaciones parasitarias y fortalecer las estrategias de eliminación (Kattenberg et al., 2024b).

Con ello, esta tesis describe la dinámica poblacional de *P. vivax* en Loreto y Amazonas utilizando datos públicos de secuencias NGS y evaluó la existencia de flujos genéticos entre zonas remotas. Se encontró que las provincias endémicas de ambas regiones presentaron estructuración genética, particularmente en Condorcanqui y Ramón Castilla, lo cual sugiere que las unidades de análisis puede integrar aislados de las mismas provincias. Además, las poblaciones parasitarias de *P. vivax* de la región de Loreto presentan una alta diversidad genética, en contraste con las poblaciones parasitarias de la región de Amazonas, donde se encontró un linaje clonal predominante. Finalmente, se evidenció que la conectividad parasitaria entre ambas regiones es limitada.

7.1. Calidad de las secuencias de *P. vivax*

Los análisis se realizaron con muestras que hayan pasado el umbral de calidad establecido. Así solo el 87.4% cumplieron con tener una profundidad de cobertura mayor a 10x y un porcentaje de genotipos faltantes mayor a 35%. Estos criterios de calidad estrictos son fundamentales para garantizar la confiabilidad en los análisis genómicos de *P. vivax*, especialmente cuando se trabaja con paneles de SNPs derivados de AmpliSeq o secuenciación de genomas completos (WGS) (Hupalo et al., 2016; Kattenberg et al., 2022; Pearson et al., 2016), asumiendo estos como mínimo estándar metodológico para la comparación entre estudios.

La cobertura mínima de 10X asegura que cada posición del genoma sea leída al menos diez veces, reduciendo la probabilidad de llamadas de SNP erróneas debidas a errores de secuenciación (Bright et al., 2012). De manera complementaria, la exclusión de muestras con más del 35% de datos faltantes permite evitar sesgos derivados de genotipos incompletos que pueden distorsionar los análisis de estructura poblacional o diversidad, porcentaje que está dentro del rango (15 – 50%) a lo realizado en otros estudios con muestras de *P. vivax* (Ibrahim et al., 2023; Kattenberg et al., 2024b; Pearson et al., 2016; Sutanto et al., 2023). Este umbral balancea la retención de información genética con la eliminación de muestras de baja calidad, una decisión especialmente relevante en *P. vivax*, cuya baja parasitemia y genomas multiclonales dificultan la obtención de datos de alta cobertura (Auburn et al., 2013).

Después de aplicar los filtros, el total de muestras son 68 y 334 para las regiones de Amazonas y Loreto, respectivamente. Según los cálculos el mínimo número de muestras era 96 para cada región, por lo que el “n” total de Amazonas queda por

debajo del umbral proyectado, lo cual puede afectar la precisión de las estimaciones de la diversidad genética y aumentar la varianza de las métricas, un efecto ampliamente documentado cuando el tamaño de muestra es reducido (Auburn et al., 2013).

Sin embargo, como la variabilidad genética de las muestras de Amazonas es más homogénea y tiene una estructura poblacional poco compleja, los estimadores de diversidad y hasta algunas medidas de diferenciación pueden estabilizarse con tamaños muestrales reducidos en comparación el “n” ideal calculado, lo que sugiere que el conjunto de muestras utilizado en los análisis puede capturar razonablemente la diversidad de esta región (Cotter et al., 2023). Aunque esto último debe validarse mediante análisis de sensibilidad, como la rarefacción o submuestreo, que permiten estandarizar el tamaño de muestra entre regiones y evaluar si la diversidad genética estimada en Amazonas se mantiene al comparar con un número equivalente de muestras, junto con el reporte de intervalos de confianza para reflejar la incertidumbre de las estimaciones (Cotter et al., 2023).

Por otro lado, en Loreto, el número de muestras (334) excede el mínimo estimado, lo que aumenta la confiabilidad de las estimaciones (menor error estándar, mayor captura de alelos raros y de estructura fina) y facilita análisis comparativos más robustos (Bashalkhanov et al., 2009).

7.2. Estructuración Poblacional

Con el fin de identificar los patrones de estructuración genética de *P. vivax* en distintos niveles espaciales y temporales, se comparó la capacidad de dos conjuntos de marcadores genómicos: el panel completo de SNPs bialélicos y el barcode *ITM*-

Peru derivado del *Pv AmpliSeq v2 Peru*.

Inicialmente, la comparación entre el conjunto completo de SNPs y el barcode *ITM-Peru* presentó diferencias significativas en las capacidades resolutivas para detectar estructuración genética en el contexto de poblaciones parasitarias de *P. vivax* en la Amazonia peruana. Así, en los tres primeros análisis de estructuración se observó que el conjunto completo de SNPs tiene una mayor capacidad para identificar clústeres genéticos coherentes geográficamente a nivel de provincia, mientras que el barcode *ITM-Peru* solo captura estructuraciones parciales.

Este resultado es consistente con las observaciones previas con paneles de distinto número de marcadores, donde los PCA y DAPC detectan clústeres coherentes con el origen geográfico cuando se utiliza paneles de SNPs más amplios en comparación con un número reducidos de SNPs (Fola et al., 2020; Kattenberg, et al., 2024b). Esta pérdida de resolución en paneles reducidos se explica porque los barcodes de tamaño fijo son diseñados con objetivos de clasificación a escala de país o regiones específicas y pueden carecer de suficiente información a escalas geográficas más grandes (Harrison et al., 2023), por lo que tienen tienden a subestimar la diferenciación poblacional (Kattenberg et al., 2024b; Pearson et al., 2016). Incluso se ha demostrado que el uso de un número reducido de marcadores, entre SNPs y microsatélites, no hay diferencias significativas en la identificación de estructuración genética concordantes con la ubicación (Cabrera-Sosa, et al., 2024b). También, estudios de genómica de *P. vivax* han demostrado que paneles genómicos extensos tienden a detectar subestructuras más sutiles incluso en contextos de alta conectividad (Kattenberg et al., 2024a).

Los análisis de estructuración genética avaló que los aislados siguen consistentemente los límites provinciales, con cuatro clústeres principales identificados: Condorcanqui, Ramón Castilla, Maynas y Loreto. Esta coherencia geográfica de los agrupamientos permite validar a la provincia como la unidad de análisis más apropiada para realizar inferencias poblacionales dentro de este estudio. Este nivel de resolución es epidemiológicamente relevante porque las provincias de la Amazonía peruana no son unidades administrativas arbitrarias, sino que sus comunidades están conectadas por sistemas fluviales que actúan como ejes de movilidad, comercio e interacción social, determinando el grado de conectividad entre poblaciones humanas y, por ende, entre poblaciones parasitarias (Villena et al., 2019). Esta dinámica de conectividad mediada por ríos ha sido documentada principalmente entre comunidades ubicadas dentro de las mismas cuencas ribereñas, así como entre cuencas distantes dentro de la región de Loreto (Delgado-Ratto et al., 2016).

La provincia de Condorcanqui mostró la mayor diferenciación genética respecto al resto de las provincias analizadas, formando un clúster bien definido y consistente en todos los métodos de análisis utilizados. Esta diferenciación es coherente con la epidemiología local y la geografía de la provincia, puesto que Condorcanqui alberga principalmente comunidades indígenas, con acceso casi exclusivamente fluvial, asentadas en las riberas del río Santiago y sus afluentes (Saavedra-Samillán et al., 2024). La alta conectividad mediada por el río Santiago junto con el marcado aislamiento externo de la provincia explican tanto la clusterización genética observada como la ausencia de subestructura genética entre sus distritos.

La provincia de Mariscal Ramón Castilla también se presentó como un clúster diferenciado del resto, con evidencia de subestructura interna entre los distritos de Ramón Castilla y Yavarí. Esta provincia ubicada en la zona de triple frontera con Brasil y Colombia, es articulada principalmente por los ríos Amazonas y Yavarí, y con limitada conectividad hacia los centros urbanos de la región (Kattenberg et al., 2024b). Estudios previos con el *Pv AmpliSeq v2 Peru* señalaron características genéticas particulares en esta zona, en el distrito de Yavarí se identificó un clúster clonal de *P. vivax* con haplotipos más similares a aislados brasileños, interpretado como una probable introducción de parásitos desde Brasil (Kattenberg et al., 2024b). La aparente diferenciación intra-provincial de Ramón Castilla puede interpretarse como dos dinámicas de transmisión distintas, una más clonal en Yavarí, coherente con su aislamiento relativo y otra en el distrito capital, articulado por el río Amazonas; sin embargo, debido al limitado muestreo dentro del distrito Ramón Castilla, no se puede realizar interpretaciones poblacionales concluyentes a nivel de distrito.

En contraste con Condorcanqui y Ramón Castilla, las provincias de Maynas y Loreto presentaron amplia dispersión en los árboles filogenéticos y considerable solapamiento en los análisis multivariados (principalmente en el PCA), reflejando un bajos niveles de estructura genética entre ambas provincias. Maynas concentra una red de comunidades ribereñas distribuidas a lo largo de los ríos Napo, Nanay, Itaya, Momón y Mazán, entre otros, que estructuran la movilidad y el comercio local; esta conectividad multidireccional facilita el intercambio parasitario entre comunidades de distintos distritos y propicia la baja estructuración genética por distritos (Van den Eede et al., 2010). Los aislados de la provincia de Loreto solo

contemplan a muestras del distrito de Trompeteros, específicamente de la comunidad de Nueva Jerusalen. Aunque hay una proximidad genética con Maynas, los resultados del DAPC muestran una diferenciación en la estructuración genética con los aislados de dicha provincia. Esta estructuración moderada es coherente con la conectividad fluvial multidireccional que caracteriza a la provincia de Maynas, puesto que Maynas concentra una red de comunidades ribereñas distribuidas a lo largo de los ríos Napo, Nanay, Itaya, Momón y Mazán, mientras que Loreto (en este caso, solo las muestras del 2019-2020 de la comunidad de Nueva Jerusalen) abarca a lo largo del río el Corrientes (Aramburú-Guarda et al., 1999). Los aislados del Datem del Marañón representan un ejemplo crítico sobre la subrepresentación de esta unidad de análisis y las deficiencias sobre los procesos de vigilancia en regiones remotas.

La dimensión temporal no emergió como una unidad de análisis robusta en los análisis poblacionales realizados. La aparente diferenciación entre los periodos 2006-2010 y 2022-2023 observada en el DAPC está probablemente confundida con el origen geográfico de los aislados, dado que el muestreo no es balanceado por tiempo y ubicación. Por ello, la agrupación por periodos fue retenida únicamente con fines descriptivos, sin derivar conclusiones poblacionales temporales independientes de la distribución geográfica.

El único escenario donde la dimensión temporal pudo ser evaluada de manera relativamente aislada fue el distrito de Mazán, el único distrito con aislados representados en los cuatro periodos de muestreo definidos. Los análisis de PCA y DAPC mostraron que los periodos 2006-2010 y 2022-2023 exhiben la mayor separación genética entre sí, mientras que los periodos 2011-2016 y 2017-2021

forman clústeres parcialmente solapados. Mazán es un distrito de alta transmisión con elevada movilidad poblacional, cuyas comunidades se distribuyen en las cuencas de los ríos Mazán y Napo, y fue uno de los distritos de Loreto con el mayor índice parasitario anual en 2017, con más del 80% de los casos ocurriendo en comunidades ribereñas (Rosas-aguirre et al., 2021b). Estas características propician condiciones para que se produzca diferenciación temporal, consistente con lo documentado en otras comunidades ribereñas del Amazonas peruano con transmisión persistente (Ramirez et al., 2025). La diferenciación entre el periodo más antiguo y el más reciente en Mazán podría reflejar la acumulación de deriva genética y posibles eventos de reintroducción a lo largo de casi dos décadas, aunque su interpretación debe hacerse con cautela dado el tamaño de muestra por periodo y la heterogeneidad espacial, es decir, número de casos por comunidad, cuyos patrones de transmisión difieren entre sí (Rosas-Aguirre et al., 2021b).

Por otro lado, los resultados del AMOVA aportan una visión jerárquica de la estructuración genética. El conjunto completo de SNPs bialélicos mostró que la mayor proporción de la variación (68,4 %) reside dentro de las provincias, mientras que los niveles entre provincias dentro de regiones aportan 16,2 % y entre regiones 15,4 %. Los estadísticos Φ_{CT} y Φ_{SC} resultaron significativos, indicando diferenciación genética tanto entre regiones como entre provincias, aunque el Φ_{ST} global no fue significativo. Esto sugiere que la estructura genética es jerárquica pero no profunda; es decir, las diferencias locales existen, pero la conectividad genómica dentro provincias sigue siendo alta. El análisis realizado con el barcode *ITM-Peru* mostró tendencias análogas, aunque con menor magnitud de varianza entre grupos, resultado coherente con su menor número de loci y menor poder informativo.

Este patrón es característico de *P. vivax* en escenarios de transmisión intermedia, donde la mayoría de la variación genética suele concentrarse dentro de las subpoblaciones debido a la elevada diversidad de linajes y la persistencia de focos clonales locales (Delgado-Ratto et al., 2016; Koepfli et al., 2011). Estudios en la Amazonía peruana y en Brasil también han reportado proporciones de varianza similares (60–70 % dentro de localidades) con diferenciación moderada entre regiones, atribuida a la mezcla de poblaciones parcialmente conectadas por movilidad fluvial y fronteriza (de Oliveira et al., 2020; Koepfli et al., 2011).

En cuanto al análisis de ancestría genética mediante ADMIXTURE, la ausencia de un valor óptimo claro de K sugiere una estructura genética continua, sin límites poblacionales definidos entre las provincias. Este patrón es característico de *P. vivax*, cuya historia demográfica se ve moldeada por mezcla de linajes, recombinación frecuente y flujo génico persistente en áreas de transmisión intermedia (Hupalo et al., 2016; Kattenberg et al., 2024b). La exploración de $K = 5$ y $K = 6$ permitió visualizar distintos niveles de subestructura: Condorcanqui mantuvo un clúster predominante persistente a lo largo de los periodos, consistente con un linaje localmente establecido; mientras que Loreto, Maynas y Datem del Marañón mostraron composiciones ancestrales mixtas, reflejando una posible conectividad genética sostenida entre provincias de la región de Loreto. En Ramón Castilla, la persistencia e incremento de un linaje ancestral tras un intervalo sin muestreo sugiere una expansión clonal local de un genotipo dominante, posiblemente mantenido por transmisión focal o brotes recurrentes (Kattenberg et al., 2024b).

Este patrón de persistencia ancestral local junto con mezclas en otras provincias

puede reflejar las dinámicas de movilidad humana y conectividad fluvial en áreas específicas de la Amazonía peruana, combinadas con focos endémicos estables donde se observa coexistencia entre linajes persistentes y poblaciones mezcladas (Delgado-Ratto et al., 2014). En regiones de baja transmisión, se ha descrito la formación de “*clonal pockets*” o subpoblaciones dominadas por pocos haplotipos, aun cuando existe flujo genético entre localidades (Ngwana-Joseph et al., 2024; Waltmann et al., 2018). Tales hallazgos apoyan la idea de que la reducción del intercambio efectivo entre parásitos, ya sea por geografía, dinámica de transmisión o intervención sanitaria, no elimina completamente la conectividad, sino que produce una estructura reticulada y dinámica, en la que linajes ancestrales pueden mantenerse y expandirse localmente.

Se evaluó el coeficiente de endogamia intrapoblacional (F_{IS}) que, aunque conceptualmente se emplea para detectar desviaciones del equilibrio de Hardy–Weinberg en oocistos, presenta una interpretación particular en *P. vivax* dado su ciclo de vida clonal y la alta proporción de infecciones monoclonales presentes en los ciclos sanguíneos (Soontarawirat et al., 2017).

Dado que la mayoría de las muestras analizadas correspondieron a infecciones monoclonales, el coeficiente de endogamia dentro de poblaciones (F_{IS}) posee una limitación biológica para la interpretación como medida de endogamia o desviaciones del equilibrio Hardy–Weinberg. Por ello, la estimación de F_{IS} se calculó únicamente con fines descriptivos, cuyos valores bajos y/o negativos refleja un exceso de heterocigotos (infecciones policlonales o recombinación) y valores altos, un déficit de heterocigotos (baja diversidad clonal). Esto debido a que el F_{IS}

en *Plasmodium* se interpreta más como una medida de clonalidad que de endogamia mendeliana y estructura intrapoblacional (Rousset et al., 2005). En este contexto, F_{IS} se puede usar como indicador de mezcla clonal o diversidad intrapoblacional más que como estructura poblacional clásica.

El barcode *ITM-Peru* proporciona una estimación más representativa de los patrones de recombinación y diversidad clonal consistente con los patrones de clonalidad detectados mediante F_{WS} . No obstante, ambos conjuntos poseen patrones de F_{IS} que sugieren una estructura clonal heterogénea en la Amazonía peruana, con focos de transmisión restringida en Condorcanqui y Ramón Castilla, y mayor mezcla genética en Loreto y Maynas. Dado las limitaciones de F_{IS} , los estudios recomiendan emplear el índice F_{WS} , que cuantifica la mezcla clonal dentro de cada infección y se correlaciona estrechamente con el F_{IS} descriptivo (Manske et al., 2012).

La comparación entre el conjunto completo de SNPs frente al barcode *ITM-Peru* evidencia que los SNPs bialélicos ofrecen una mayor capacidad resolutive para detectar estructura poblacional, incluso cuando los valores de F_{ST} o la topología de los árboles filogenéticos resultan similares. Esto se debe a que los SNPs bialélicos abarcan un número mayor de posiciones informativas distribuidas por todo el panel del *Pv AmpliSeq v2 Peru*, lo cual reduce el efecto de varianza muestral y mejora la capacidad para discernir patrones finos de diferenciación.

Los patrones de F_{ST} no cambiaron al compararlo ambos conjuntos de SNPs. Esto es debido a que, a pesar que las regiones bajo presión de selección suelen tener bajos niveles de diversidad (Labbé et al., 2023), la proporción de SNPs en presión de

selección en los ensayos AmpliSeq (pocos genes de resistencia) no fue alta en comparación a los SNPs en regiones neutrales como para influir los valores de diversidad (Kattenberg et al., 2024b).

7.3. Diversidad Genética

Los patrones de diversidad genética principalmente se estimaron con dos métricas complementarias como la heterocigocidad esperada (H_e , mediante el uso de SNPs neutros del barcode *ITM-Peru*) y diversidad nucleotídica (π , usando panel completo *Pv AmpliSeq v2 Peru*). Calculamos la H_e con el barcode *ITM-Peru* de 41 SNPs porque este conjunto fue curado para incluir marcadores bialélicos, independientes (bajo desequilibrio de ligamiento) y putativamente neutrales distribuidos en el genoma (Kattenberg et al., 2024b). Eso hace que H_e sea robusta a sesgos por ligamiento o selección, requiera menos cobertura para llamadas confiables y, sobre todo, sea comparable entre estudios y periodos.

También se estima π usando todas las regiones del panel *Pv-AmpliSeq v2 Peru* porque π es una medida por sitio que gana precisión con mayor número de posiciones y capta variantes raras y multialélicas fuera del barcode (Hupalo et al., 2016). Al integrar regiones neutrales y funcionales, π ofrece una visión de mayor resolución de la estructura genética y posibles gradientes espaciales/temporales o señales de selección que el barcode, por diseño, no persigue (Argyropoulos et al., 2023).

Con dos métricas complementarias como la H_e y π se observa un gradiente espacial de diversidad. A nivel regional, Loreto presenta mayor diversidad ($H_e = 0.339$; $\pi = 8 \times 10^{-4}$) frente a Amazonas ($H_e = 0.065$; $\pi = 4 \times 10^{-4}$). A nivel provincial, destacan

Maynas ($H_e = 0.351$) y Datem del Marañón ($H_e = 0.295$; $\pi = 14 \times 10^{-4}$), seguidas de la provincia de Loreto ($H_e = 0.277$; $\pi = 10 \times 10^{-4}$). En contraste, Ramón Castilla y Condorcanqui muestran H_e y π bajos (ambas con $\pi \approx 4 \times 10^{-4}$; $H_e = 0.090$ y 0.065 , respectivamente). Sin embargo, los valores estimados de Datem del Marañón no son representativos porque solo incluye 4 aislamientos. Por otro lado, aunque *Plasmodium* presenta una fase haploide durante la infección humana, las muestras fueron tratadas como diploides a efectos analíticos, dado que la presencia de múltiples alelos por locus puede interpretarse como heterocigosis aparente derivada de infecciones multiclonales (Manske et al., 2012). Así, la complejidad de infección basada en F_{ws} indica 16.17% de infecciones policlonales ($F_{ws} < 0.95$), concentradas en Loreto (0.35) y Datem (0.25), y proporciones bajas en Condorcanqui (0.03) y Ramón Castilla (0.14).

Al comparar los valores de diversidad genética con estudios previos de la Amazonía peruana se observa que, en general, los valores de H_e de la región de Loreto, en particular de la provincia de Maynas y Loreto, están dentro del rango moderada a alta, con excepciones de las muestras pertenecientes a la provincia de Ramón Castilla (Kattenberg et al., 2024b). Diversos estudios multicéntricos de *P. vivax* entre el 2006-2008 en los alrededores de Iquitos - Loreto reportan $H_e \sim 0.44$ – 0.69 mediante microsatélites (MS) (Van den Eede et al., 2010). En series posteriores y más amplias del alrededor de Iquitos y localidades vecinas se observan H_e de 0.66 – 0.76 , coherente con transmisión moderada y policlonalidad no trivial (Delgado-Ratto et al., 2016). Varios trabajos de cohorte focalizados (San Carlos, 2008–2010) muestran una diversidad $H_e \approx 0.49$ y predominio de infecciones monoclonales, ilustrando que sitios pequeños/aislados pueden tener diversidad más baja que el

promedio regional (Delgado-Ratto et al., 2014). Otros análisis con 16 MS en zonas distantes también sitúan $He \approx 0.54\text{--}0.60$ y resaltan conectividad entre focos alejados (Manrique et al., 2019).

En la aplicación del *Pv AmpliSeq v2 Peru* a 11 distritos de la región de Loreto, la diversidad por el barcode *ITM-Peru* fue moderada y con caídas marcadas en Yavarí (Ramón Castilla) ($He \approx 0.012$), mientras π (también calculada sobre todas las variantes del panel) distinguió mejor Iquitos (más alta) de Mazán/San Juan y cayó también en Yavarí; además, π fue mayor en 2007–2008 que en años recientes (Kattenberg et al., 2024b). Posteriormente, este mismo panel fue evaluado al comparar MS vs. barcode IMT en tres áreas de Loreto, $He_{MS} = 0.68\text{--}0.78$ vs. $He_{SNP} = 0.36\text{--}0.38$, encontrando tendencias concordantes pero menor amplitud con SNP por menor variabilidad por locus (Cabrera-Sosa et al., 2024b).

Al considerar la escala provincial, la situación de Condorcanqui y Ramón Castilla resulta crítica para la interpretación epidemiológica genética. Estas provincias aisladas exhiben valores bajos de heterocigosidad (He) y altas medidas de diferenciación precisamente por su aislamiento geográfico, transmisión local restringida y expansión clonal de linajes dominantes. Así, en congruencia con el contexto peruano, se reportaron linajes predominante en los distritos fronterizos Yavarí y Ramón Castilla, indicando un linaje genético persistente en esa zona (Kattenberg et al., 2024b). En contextos parasitarios, la baja diversidad y clonalidad local suelen asociarse con focos de transmisión estables de baja intensidad, donde pocos linajes dominan (Waltmann et al., 2018), aunque esa hipótesis se probó usando microsatélites en *P. vivax* en Oceanía, el principio es aplicable en entornos con restricción de transmisión.

A escala Latinoamérica, la π media fue $\sim 7.1 \times 10^{-4}$ ($\pm 4.6 \times 10^{-4}$), con Perú mostrando niveles comparables a Colombia y algo por debajo de Brasil; se ubica a Perú en el mismo orden de magnitud (10^{-3} – 10^{-4}) y confirma la hipótesis de que hay mayor frecuencia de infecciones policlonales en países con más casos (Colombia, Brasil, Perú) (Kattenberg et al., 2024a). Los valores de π en este trabajo son coherentes con valores estimados mediante WGS en Latinoamérica ($\sim 7 \times 10^{-4}$). Estos patrones muestran a π como métrica más estable/útil para contrastes espaciales/temporales dentro de un panel bien diseñado vs WGS.

Por otro lado, este estudio constituye entre los primeros análisis de genética poblacional de *P. vivax* en la región Amazonas. Los estimados muestran una población altamente homogénea, con bajos valores de diversidad y predominio de infecciones monoclonales. Este patrón es congruente con escenarios de transmisión focal y recombinación limitada, descritos en zonas periféricas o de frontera amazónica, donde los tamaños efectivos reducidos, los cuellos de botella y/o efectos fundador favorecen la estructura clonal (Brashear & Cui, 2022; Kattenberg et al., 2024b).

7.4. Conectividad Genética

Los resultados de conectividad genética obtenidos mediante los diferentes enfoques apuntan a un escenario de flujo génico bajo y direccional entre las provincias amazónicas del norte del Perú. Los análisis basados en identidad por descendencia (IBD) mostraron la coexistencia de clones persistentes y parásitos parcialmente relacionados entre provincias, lo cual evidencia una combinación de transmisión local estable y dispersión ocasional de linajes. Los clústeres de alto IBD (> 0.9) observados en Condorcanqui y Ramón Castilla sugieren persistencia clonal de

genotipos dominantes, similar a lo reportado en la Amazonía peruana anteriormente (Delgado-Ratto et al., 2016; Kattenberg et al., 2024b) donde linajes idénticos se detectaron repetidamente en áreas de baja transmisión. Al reducir el umbral ($IBD \approx 0.5$), la aparición de vínculos interprovinciales (principalmente entre los distritos de Maynas y Loreto) indica parentesco parcial y flujo genético histórico, compatible con un modelo de conectividad fluvial y movilidad humana, mecanismo clave de dispersión parasitaria en la Amazonía (Carrasco-Escobar et al., 2022, 2024; Manrique et al., 2019).

El patrón jerárquico evidenciado por las redes de StrainHub confirma la existencia de nodos fuente y sumidero de transmisión. Esta medición es una inferencia indirecta sobre cómo los linajes se relacionan entre distintas agrupaciones (en este caso, por provincias) a partir de los patrones de proximidad derivada de un árbol filogenético. Aunque reflejan relaciones compatibles con dispersión y/o intercambio entre poblaciones, no necesariamente implica transmisión directa entre individuos (De Bernardi Schneider et al., 2020).

Sin embargo, es posible identificar áreas fuentes y sumideros de las infecciones de *P. vivax*. Las provincias de Maynas y Loreto exhiben altos valores de outdegree y Source Hub Ratio, desempeñando roles de emisores de transmisión, mientras que Condorcanqui, Ramón Castilla y Datem del Marañón actúan como receptores terminales. Esta arquitectura coincide con los resultados conectividad espacial inferidos por IBD, donde se demostraron que los focos de alta transmisión próximos a ejes fluviales funcionan como fuentes de infección regional, mientras las zonas periféricas mantienen transmisión residual dependiente de la introducción de linajes externos (Carrasco-Escobar et al., 2024). Además, la asimetría direccional

observada (Maynas → provincias rurales) se alinea con la hipótesis de dispersión desde núcleos urbanos hacia zonas periurbanas y rurales de difícil acceso, un proceso también evidenciado en reportes sobre movilidad humana y riesgo de malaria en comunidades amazónicas (Gomez et al., 2025).

Notablemente, la provincia de Loreto mostró un papel como fuente y receptora de transmisión con respecto a Maynas. Este patrón se explicaría en el árbol filogenético de todos los SNPs bialélicos, los aislados de Loreto se distribuyen dentro de los mismos clados que los aislados de Maynas, y en el bajo valor que el F_{ST} observado entre ambas provincias (0.07). Adicionalmente, se observó que las poblaciones parasitaria por provincias que mantienen una baja diferenciación genética ($0.0 < F_{ST} < 0.08$) no mantienen un flujo genético alto. Esto se podría explicar que las poblaciones parasitarias no cumplen con el modelo clásico de la relación F_{ST} y N_m . El modelo clásico que vincula los valores de F_{ST} con el flujo génico mediante N_m asume poblaciones ideales y en equilibrio. Estos supuestos se cumplen rara vez y tampoco es el caso de las poblaciones parasitarias, por lo que los valores de F_{ST} describen bien la diferenciación más no se correlaciona directamente con las estimaciones del N_m de forma precisa (Whitlock & McCauley, 1999).

Sin embargo, del modelo clásico puede rescatarse un principio útil: valores de $N_m > 1$ indican un flujo génico suficiente para contrarrestar la diferenciación entre poblaciones, mientras que $N_m < 1$ sugiere un intercambio limitado de migrantes y, por tanto, una mayor subdivisión poblacional (Nielsen & Slatkin, 2013). Los valores de $N_m < 1$ estimados entre provincias confirman que el flujo génico es

insuficiente para homogenizar las poblaciones, congruente con los valores de diferenciación genética y las subdivisiones descritas.

Sin embargo, es necesario indicar que los niveles de flujo genético entre distritos de cada provincia está determinado principalmente por los corredores de malaria. Este patrón observado refleja la conectividad entre zonas distantes mediada por corredores fluviales y desplazamientos humanos (Manrique et al., 2019).

Así también, dado que los valores de N_m entre las provincias de Maynas y Condorcanqui es aproximadamente 1 (principalmente con dirección de Maynas → Condorcanqui), y considerando que las redes de IBD muestran individuos de ambas provincias con señales de conectividad genética, los resultados indicarían una posibilidad de introducción mediado por las movilizaciones humanas desde Loreto (particularmente desde Maynas) hacia Amazonas. Probablemente la introducción habría correspondido a un linaje que posteriormente experimentó una expansión clonal focalizada en Amazonas, explicando así los bajos niveles de diversidad genética.

En este contexto, Maynas actuaría como reservorio genético central, redistribuyendo linajes hacia provincias más aisladas. Estudios genómicos recientes en Latinoamérica refuerzan este modelo, donde se describe que las poblaciones de *P. vivax* en Perú, Brasil y Colombia están estructuradas pero conectadas, con señales de flujo genético moderado y selección local (Kattenberg et al., 2024a). De manera análoga, se evidenciaron en Colombia una red inter-país con cuellos de botella locales (Sutanto et al., 2023), patrón que coincide con la baja conectividad de Ramón Castilla y Condorcanqui.

El análisis del origen geográfico mediante el barcode GEO33 respalda parcialmente esta interpretación. Las asignaciones de origen de Condorcanqui a Colombia y Ramón Castilla a Brasil reflejan la conectividad transfronteriza del linaje de *P. vivax*, lo que concuerda con su localización fronteriza y su papel como zona de intercambio transfronterizo de linajes parasitarios (Kattenberg et al., 2024a, Kattenberg et al., 2024b).

Esta afinidad genética binacional coincide con los hallazgos previos (Manrique et al., 2019), donde se evidenciaron conectividad significativa entre zonas peruanas y brasileñas distantes mediante microsatélites. Más aun, usando el mismo barcode que estaba integrado dentro del panel *Pv Ampliseq v2 Peru*, se detectaron linajes compartidos en la frontera Perú–Brasil (Kattenberg et al., 2024b).

Sin embargo, se debe considerar que el barcode GEO33 fue calibrado principalmente con genomas de *P. vivax* de Asia y Oceanía, por lo que su poder discriminante podría verse reducido en regiones no representadas o poco representadas en el dataset original, tales como las Sudaméricas. Por lo tanto, su capacidad para predecir el origen de poblaciones circulantes en Perú podría estar limitada. A esto se suma que, en infecciones mixtas o con clones múltiples, la asignación de origen pierde fiabilidad, lo que restringe su utilidad para estudios finos de estructura poblacional o transmisión local (Trimarsanto et al., 2022).

Aun así, la convergencia entre los patrones inferidos por el GEO33 y la estructura poblacional observada mediante SNPs bialélicos refuerza la validez de una conectividad transfronteriza regional, donde la dispersión fluvial y la movilidad humana mantienen un equilibrio entre aislamiento local y flujo génico continental,

coherente con un modelo de una población heterogénea de *P. vivax* en la amazonía peruana (Delgado-Ratto et al., 2016; Kattenberg, et al., 2024b). Además, las muestras aisladas de Ramón Castilla mostraron baja conectividad con otras provincias peruanas y una similitud mayor con cepas brasileñas, reforzando la hipótesis de aislamiento parcial, diferenciación local y conectividad con cepas fronterizas (Kattenberg, et al., 2024b).

Finalmente, los patrones de conectividad genética descritos aquí tienen implicancias epidemiológicas directas para la transmisión de *P. vivax* en la Amazonía peruana. La persistencia de linajes clonales en provincias aisladas como Condorcanqui y Ramón Castilla sugiere una transmisión focal sostenida localmente, característica de zonas donde la circulación del parásito depende de unos pocos genotipos dominantes que se mantienen a través del tiempo (Delgado-Ratto et al., 2016; Kattenberg et al., 2024b). En contraste, el flujo génico bajo desde Maynas y Loreto evidencia una conectividad regional restringida, pero si capaz de reintroducir linajes en provincias periféricas, lo que representa un riesgo constante de reinfecciones importadas y un obstáculo para las estrategias de eliminación de malaria (Carrasco-Escobar et al., 2024; Guerra et al., 2019).

Este tipo de red jerarquizada, compuesta por nodos fuente y sumidero, concuerda con modelos de transmisión observados en otras regiones amazónicas, donde los focos de alta densidad poblacional y conectividad fluvial actúan como reservorios genéticos regionales, y las áreas más remotas albergan bolsones clonales de baja diversidad (Kattenberg et al., 2024b; Manrique et al., 2019). Así, la estructura observada refleja un sistema poblacional parcialmente conectado, donde la

movilidad humana, la hidrografía regional y la adaptación local de linajes determinan los patrones de dispersión del parásito (Sanchez et al., 2025)

En este contexto, el patrón de subdivisión genética identificado, con flujos génicos restringidos y diferenciación provincial marcada, es consistente con los escenarios de transmisión reducida, en los que la disminución del flujo de parásitos intensifica la estructuración poblacional y favorece la aparición de linajes localmente adaptados (Sutanto et al., 2023). Bajo estas condiciones, la detección de infecciones causadas por parásitos pertenecientes a una misma subpoblación sugiere que los individuos que los portan comparten factores epidemiológicos comunes, como ocupación, proximidad geográfica o patrones de movilidad. Estos hallazgos refuerzan la necesidad de integrar vigilancia genómica, movilidad humana y ecología del paisaje parasitario para comprender la dinámica de transmisión y orientar intervenciones focalizadas hacia los nodos de mayor conectividad.

7.5. Relevancia

La relevancia de este estudio constituye en que es entre uno de los primeros esfuerzos en caracterizar la diversidad y conectividad genética de *Plasmodium vivax* a múltiples escalas espaciales (regional, provincial y distrital) entre las regiones de Amazonas y Loreto. La generación de este tipo de evidencia genómica representa un aporte estratégico al conocimiento de la epidemiología molecular del *P. vivax*, ya que permite visualizar los patrones de dispersión, persistencia y diversificación del parásito en escenarios donde la vigilancia convencional tiene limitaciones operativas. Además, este enfoque que homogeniza data de parcial y genoma completo y analiza la conectividad genética, permite dilucidar conocimiento detallado de la distribución espacial y temporal de los linajes, así

como de su carga genética y diversidad, lo cual fundamental para establecer líneas de base genómicas que permitan monitorear el impacto de las intervenciones y medir el progreso hacia la eliminación (Arnott et al., 2012; Kattenberg et al., 2024b; Sutanto et al., 2023).

Por otro lado, desde una perspectiva epidemiológica, los resultados revelan la coexistencia de focos periféricos aislados (Condorcanqui, Ramón Castilla) y corredores de conectividad con roles diferenciados en la transmisión:

- En los focos periféricos, la baja diversidad y alta endogamia evidencian transmisión focal y expansión clonal, típica de escenarios de baja incidencia pero con persistencia local de linajes adaptados. Estas zonas deben ser priorizadas para intervenciones focalizadas, vigilancia activa y rastreo genético de introducciones, especialmente en áreas fronterizas donde la movilidad humana es alta pero los recursos de diagnóstico son limitados.
- En los corredores conectados, principalmente dentro de provincias, la alta policlonalidad y recombinación reflejan una transmisión sostenida y un flujo genético continuo entre áreas endémicas amazónicas. Esto exige mantener una vigilancia genómica continua, capaz de detectar linajes emergentes, reintroducciones transfronterizas y posibles adaptaciones locales a tratamientos o vectores (Carrasco-Escobar et al., 2024; Manrique et al., 2019).

Metodológicamente, este estudio resalta el potencial del *Pv AmpliSeq v2 Peru* como una herramienta intermedia y costo-efectiva entre los métodos tradicionales de microsatélites y el secuenciamiento genómico completo. Esta plataforma ofrece una

alternativa viable para la vigilancia molecular rutinaria de *P. vivax* en contextos de transmisión residual, ya que puede aplicarse incluso en muestras con baja parasitemia, típicas de infecciones vivax en etapa de eliminación. Su implementación podría fortalecer la capacidad diagnóstica y analítica del sistema de salud peruano, permitiendo integrar la vigilancia genómica al marco operativo del Plan Nacional de Eliminación de la Malaria 2022–2030 del MINSA.

7.5. Limitaciones

Este estudio presenta algunas limitaciones que deben considerarse al interpretar los resultados. En primer lugar, el desbalance muestral entre regiones y periodos limita la comparación directa de la diversidad genética y podría introducir sesgos en los índices de estructuración. En particular, la subrepresentación de los aislados de Condorcanqui y del Datem del Marañón restringe la capacidad de detectar con precisión la diversidad intrapoblacional y puede sobreestimar la diferenciación genética observada en estas provincias. En poblaciones pequeñas o con bajo número de genomas, las estimaciones de H_e , F_{st} y N_m tienden a ser más sensibles al error muestral y pueden reflejar parcialmente el efecto del tamaño de muestra más que una divergencia biológica real. Esto es especialmente relevante para Condorcanqui, donde la cantidad limitada de muestras puede subestimar la variabilidad genética local, y para el Datem del Marañón ($n = 4$), donde el bajo número de aislamientos reduce la robustez de las inferencias sobre conectividad.

Asimismo, la presencia potencial de infecciones multiclonales no identificadas y la dependencia de muestras clínicas sintomáticas, pueden influir en las estimaciones de heterocigosidad y de parentesco genético, subrepresentando la transmisión submicroscópica o asintomática. Estas últimas representan la mayor proporción de

infecciones en la Amazonía peruana y son epidemiológicamente relevante en la persistencia y propagación del parásito.

Una limitación crítica adicional es la ausencia de datos epidemiológicos y clínicos del hospedero humano en los repositorios genómicos utilizados. Al no disponer de información detallada sobre edad, movilidad, ocupación, historial de viaje o residencia, los análisis de conectividad genética no pueden vincularse directamente con los patrones reales de transmisión o con factores de riesgo humanos. Esto restringe la interpretación de la conectividad observada, que se basa exclusivamente en la información del parásito.

VIII. CONCLUSIONES

Las poblaciones de *Plasmodium vivax* en la Amazonía peruana presentan niveles bajos de flujo genético y una estructura geográfica marcada a nivel de provincias, con una clara separación entre las regiones de Loreto y Amazonas. La población de *P. vivax* de las provincias de Maynas y Loreto posee una mayor diversidad genética, consistente con una circulación más prolongada y variada de linajes, mientras que las poblaciones de Condorcanqui y Ramón Castilla se encuentran genéticamente aisladas, caracterizadas por baja heterocigosidad y expansión clonal, reflejo de transmisión local restringida.

El bajo parentesco genético y los valores reducidos de la cuantificación de la conectividad entre provincias evidencian un flujo génico muy limitado y asimétrico, posiblemente asociado a eventos de introducción mediado por movilizaciones humanas desde Loreto hacia Amazonas, seguido de la persistencia de linajes clonales en zonas aisladas de transmisión residual. El flujo genético detectado se concentra principalmente dentro de los distritos de cada provincia, aunque los patrones de transmisión potencial señalan a Maynas como el principal nodo emisor dentro de la red regional.

La distribución de la variabilidad genética sugiere que *P. vivax* en la Amazonía peruana conforma un sistema poblacional de baja conectividad, donde la mezcla genética y el aislamiento local coexisten bajo un equilibrio dinámico influido posiblemente por la movilidad humana restringida entre distritos endémicos cercanos, la geografía fluvial y la intensidad de heterogénea de transmisión.

IX. RECOMENDACIONES

Dado que los resultados de vigilancia genómica ofrecen una base sólida para comprender la dinámica espacial del *P. vivax*, se recomienda integrar los análisis genómicos con la vigilancia epidemiológica y entomológica nacional, incorporando información del hospedero, del vector y de la movilidad humana. Esta articulación permitiría interpretar con mayor precisión los patrones de transmisión y fortalecer la detección temprana de focos activos o reintroducciones.

Para comprender con mayor precisión la dinámica de transmisión de *P. vivax*, se recomienda realizar análisis genómicos en base a WGS en áreas geográficas que cuenten con series temporales extensas. Este enfoque permitiría evaluar, a microescala, los patrones de dispersión y la estructuración de las poblaciones a lo largo del tiempo, proporcionando una visión más completa de los procesos que sostienen la transmisión local y del posible desplazamiento del parásito entre entornos urbanos, semiurbanos y rurales.

Aunque el panel *Pv AmpliSeq v2 Peru* incluye marcadores asociados a genes de resistencia antimalárica, se recomienda que los futuros estudios profundicen en la caracterización y el estado actual de dichos genes y/o genes que no contempla el panel, con el propósito de evaluar posibles presiones selectivas y fortalecer la vigilancia terapéutica frente a *P. vivax*. La incorporación sistemática de estos análisis permitirá detectar tempranamente la emergencia o propagación de variantes resistentes y aportar evidencia complementaria para la optimización de esquemas de tratamiento y control antimalárico en la región amazónica.

Se recomienda priorizar la vigilancia genómica y el monitoreo continuo en las

provincias de Condorcanqui, Ramón Castilla y Datem del Marañón, debido a su bajo número de aislamientos y su papel como focos genéticamente aislados dentro de la red regional. El fortalecimiento de la capacidad de muestreo a lo largo de los años y análisis en estas zonas permitirá mejorar la detección de linajes locales, evaluar posibles reintroducciones y comprender la dinámica de transmisión en áreas donde el flujo génico es restringido.

En futuros estudios, se debería considerarse la inclusión de información sobre la ecología y comportamiento de los vectores, conectividad fluvial, variables climáticas y actividades antropogénicas (como el uso del suelo), con el fin de evaluar cómo los factores ecológicos influyen en la estructura genética del parásito y en la dispersión de linajes entre provincias.

Y por último, en base a los resultados de la vigilancia molecular, tales como este, donde se observa *hotstops* de malaria en la zona fronteriza, se sugiere fortalecer la cooperación binacional con Brasil y Colombia, a fin de establecer mecanismos coordinados de intercambio de datos y monitoreo conjunto de linajes parasitarios transfronterizos.

X. REFERENCIAS BIBLIOGRÁFICAS

- Adams, J. H., & Mueller, I. (2017). The biology of plasmodium vivax. *Cold Spring Harbor Perspectives in Medicine*, 7(9). <https://doi.org/10.1101/cshperspect.a025585>
- Akoniyon, O. P., Akiibinu, M., Adeleke, M. A., Maharaj, R., & Okpeku, M. (2024). A Comparative Study of Genetic Diversity and Multiplicity of Infection in Uncomplicated Plasmodium falciparum Infections in Selected Regions of Pre-Elimination and High Transmission Settings Using MSP1 and MSP2 Genes. *Pathogens*, 13(2), 172. <https://doi.org/10.3390/PATHOGENS13020172/S1>
- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655–1664. <https://doi.org/10.1101/GR.094052.109>
- Amir, A., Cheong, F. W., De Silva, J. R., & Lau, Y. L. (2018). Diagnostic tools in childhood malaria. *Parasites and Vectors*, 11(1), 1–12. <https://doi.org/10.1186/S13071-018-2617-Y/TABLES/1>
- Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*. Babraham Bioinformatics. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Aramburú-Guarda, J., Ramal Asayag, C., & Witzig, R. (1999). *Malaria Reemergence in the Peruvian Amazon Region*. 5(2), 209–215.
- Arnott, A., Barry, A. E., & Reeder, J. C. (2012). Understanding the population genetics of Plasmodium vivax is essential for malaria control and elimination. *Malaria Journal*, 11, 1–10. <https://doi.org/10.1186/1475-2875-11-14>
- Auburn, S., & Barry, A. E. (2017). Dissecting malaria biology and epidemiology using population genetics and genomics. *International Journal for Parasitology*, 47(2–3), 77–85. <https://doi.org/10.1016/j.ijpara.2016.08.006>
- Auburn, S., Campino, S., Miotto, O., Djimde, A. A., Zongo, I., Maslen, G., Mangano, V., Alcock, D., Macinnis, B., Rockett, K. A., Kwiatkowski, D. P.,

- Clark, T. G., Doumbo, O. K., & Oue, J. B. (2012). *Characterization of Within-Host Plasmodium falciparum Diversity Using Next-Generation Sequence Data*. 7(2). <https://doi.org/10.1371/journal.pone.0032891>
- Auburn, S., Cheng, Q., Marfurt, J., & Price, R. N. (2021). The changing epidemiology of Plasmodium vivax: Insights from conventional and novel surveillance tools. In *PLoS Medicine* (Vol. 18, Issue 4). Public Library of Science. <https://doi.org/10.1371/journal.pmed.1003560>
- Auburn, S., Marfurt, J., Maslen, G., Campino, S., Ruano Rubio, V., Manske, M., MacHunter, B., Kenangalem, E., Noviyanti, R., Trianty, L., Sebayang, B., Wirjanata, G., Sriprawat, K., Alcock, D., MacInnis, B., Miotto, O., Clark, T. G., Russell, B., Anstey, N. M., ... Price, R. N. (2013). Effective Preparation of Plasmodium vivax Field Isolates for High-Throughput Whole Genome Sequencing. *PLOS ONE*, 8(1), e53160. <https://doi.org/10.1371/JOURNAL.PONE.0053160>
- Barry, A. E., Schultz, L., Senn, N., Nale, J., Kiniboro, B., Siba, P. M., Mueller, I., & Reeder, J. C. (2013). High levels of genetic diversity of Plasmodium falciparum populations in Papua New Guinea despite variable infection prevalence. *American Journal of Tropical Medicine and Hygiene*, 88(4), 718–725. <https://doi.org/10.4269/ajtmh.12-0056>
- Barry, A. E., Waltmann, A., Koepfli, C., Barnadas, C., & Mueller, I. (2015). Uncovering the transmission dynamics of Plasmodium vivax using population genetics. *Pathogens and Global Health*, 109(3), 142–152. <https://doi.org/10.1179/2047773215Y.0000000012>
- Bashalkhanov, S., Pandey, M., & Rajora, O. P. (2009). A simple method for estimating genetic diversity in large populations from finite sample sizes. *BMC Genetics*, 10, 84. <https://doi.org/10.1186/1471-2156-10-84>
- Bhatt, S., Weiss, D. J., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., Battle, K. E., Moyes, C. L., Henry, A., Eckhoff, P. A., Wenger, E. A., Briët, O., Penny, M. A., Smith, T. A., Bennett, A., Yukich, J., Eisele, T. P., Griffin,

- J. T., Fergus, C. A., ... Gething, P. W. (2015). The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature*, 526(7572), 207–211. <https://doi.org/10.1038/nature15535>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Brashear, A. M., & Cui, L. (2022). Population genomics in neglected malaria parasites. *Frontiers in Microbiology*, 13(September), 1–17. <https://doi.org/10.3389/fmicb.2022.984394>
- Bright, A. T., Tewhey, R., Abeles, S., Chuquiyauri, R., Llanos-Cuentas, A., Ferreira, M. U., Schork, N. J., Vinetz, J. M., & Winzeler, E. A. (2012). Whole genome sequencing analysis of *Plasmodium vivax* using whole genome capture. *BMC Genomics*, 13(1), 262. <https://doi.org/10.1186/1471-2164-13-262>
- Broad Institute. (2019). *Picard Toolkit [Software]*. Broad Institute, GitHub repository. <https://broadinstitute.github.io/picard/>
- Brown, T. S., Arogbokun, O., Buckee, C. O., & Chang, H. H. (2021). Distinguishing gene flow between malaria parasite populations. *PLoS Genetics*, 17(12). <https://doi.org/10.1371/journal.pgen.1009335>
- Browning, B. L., & Browning, S. R. (2013). Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics*, 194(2), 459–471. <https://doi.org/10.1534/GENETICS.113.150029/-/DC1>
- Bustamante-Chauca, T. P. (2019). Caracterización epidemiológica de la transmisión de la malaria en comunidades indígenas del departamento de Amazonas. 2009 – 2019. *Revista Del Cuerpo Médico Del HNAAA*, 12(4), 325–331. <https://doi.org/10.35434/rcmhnaaa.2019.124.567>
- Cabrera-Sosa, L., Nolasco, O., Kattenberg, J. H., Fernandez-Miño, C., Valdivia, H. O., Barazorda, K., Arévalo de los Rios, S., Rodriguez-Ferrucci, H., Vinetz,

- J. M., Rosanas-Urgell, A., Van geertruyden, J. P., Gamboa, D., & Delgado-Ratto, C. (2024a). Genomic surveillance of malaria parasites in an indigenous community in the Peruvian Amazon. *Scientific Reports*, 14(1). <https://doi.org/10.1038/s41598-024-66925-x>
- Cabrera-Sosa, L., Safarpour, M., Kattenberg, J. H., Ramirez, R., Vinetz, J. M., Rosanas-Urgell, A., Gamboa, D., & Delgado-Ratto, C. (2024b). Comparing newly developed SNP barcode panels with microsatellites to explore population genetics of malaria parasites in the Peruvian Amazon. *Frontiers in Genetics*, 15. <https://doi.org/10.3389/fgene.2024.1488109>
- Carrasco-Escobar, G., Fornace, K., Wong, D., Padilla-Huamantínco, P. G., Saldaña-Lopez, J. A., Castillo-Meza, O. E., Caballero-Andrade, A. E., Manrique, E., Ruiz-Cabrejos, J., Barboza, J. L., Rodriguez, H., Henostroza, G., Gamboa, D., Castro, M. C., Vinetz, J. M., & Llanos-Cuentas, A. (2020). Open-Source 3D Printable GPS Tracker to Characterize the Role of Human Population Movement on Malaria Epidemiology in River Networks: A Proof-of-Concept Study in the Peruvian Amazon. *Frontiers in Public Health*, 8(September), 1–11. <https://doi.org/10.3389/fpubh.2020.526468>
- Carrasco-Escobar, G., Matta-Chuquisapon, J., Manrique, E., Ruiz-Cabrejos, J., Barboza, J. L., Wong, D., Henostroza, G., Llanos-Cuentas, A., & Benmarhnia, T. (2022). Quantifying the effect of human population mobility on malaria risk in the Peruvian Amazon. *Royal Society Open Science*, 9(7). <https://doi.org/10.1098/rsos.211611>
- Carrasco-Escobar, G., Villa, D., Barja, A., Lowe, R., Llanos-Cuentas, A., & Benmarhnia, T. (2024). The role of connectivity on malaria dynamics across areas with contrasting control coverage in the Peruvian Amazon. *PLoS Neglected Tropical Diseases*, 18(11), e0012560. <https://doi.org/10.1371/journal.pntd.0012560>
- Cavasini, C. E., Mattos, L. C. de, Couto, Á. A. D. A., Bonini-Domingos, C. R., Valencia, S. H., Neiras, W. C. de S., Alves, R. T., Rossit, A. R. B., Castilho, L., & Machado, R. L. D. (2007). Plasmodium vivax infection among Duffy

- antigen-negative individuals from the Brazilian Amazon region: an exception? *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 101(10), 1042–1044. <https://doi.org/10.1016/J.TRSTMH.2007.04.011>,
- Chenet, S. M., Tapia, L. L., Escalante, A. A., Durand, S., Lucas, C., & Bacon, D. J. (2012). Genetic diversity and population structure of genes encoding vaccine candidate antigens of *Plasmodium vivax*. *Malaria Journal*, 11, 6–8. <https://doi.org/10.1186/1475-2875-11-68>
- Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., & De Hoon, M. J. L. (2009). Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11), 1422–1423. <https://doi.org/10.1093/bioinformatics/btp163>
- Cohuet, A., Harris, C., Robert, V., & Fontenille, D. (2010). Evolutionary forces on *Anopheles*: what makes a malaria vector? In *Trends in Parasitology* (Vol. 26, Issue 3, pp. 130–136). <https://doi.org/10.1016/j.pt.2009.12.001>
- Cotter, D. J., Hofgard, E. F., Novembre, J., Szpiech, Z. A., & Rosenberg, N. A. (2023). A rarefaction approach for measuring population differences in rare and common variation. *Genetics*, 224(2), iyad070. <https://doi.org/10.1093/GENETICS/IYAD070>
- Cowell, A. N., Loy, D. E., Sundararaman, S. A., Valdivia, H., Fisch, K., Lescano, A. G., Baldeviano, G. C., Durand, S., Gerbasi, V., Sutherland, C. J., Nolder, D., Vinetz, J. M., Hahn, B. H., & Winzeler, E. A. (2017). Selective whole-genome amplification is a robust method that enables scalable whole-genome sequencing of *Plasmodium vivax* from unprocessed clinical samples. *MBio*, 8(1). <https://doi.org/10.1128/mBio.02257-16>
- Cowell, A. N., Valdivia, H. O., Bishop, D. K., & Winzeler, E. A. (2018). Exploration of *Plasmodium vivax* transmission dynamics and recurrent infections in the Peruvian Amazon using whole genome sequencing. *Genome*

- Medicine*, 10(1). <https://doi.org/10.1186/S13073-018-0563-0>,
- Csardi, G., & Nepusz, T. (2006). *The igraph software package for complex network research. InterJournal, Complex Systems*, 1695. <https://igraph.org/>
- Dalmat, R., Naughton, B., Kwan-Gett, T. S., Slyker, J., & Stuckey, E. M. (2019). Use cases for genetic epidemiology in malaria elimination. In *Malaria Journal* (Vol. 18, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s12936-019-2784-0>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/BIOINFORMATICS/BTR330>
- De Bernardi Schneider, A., Ford, C. T., Hostager, R., Williams, J., Cioce, M., Çatalyürek, Ü. V., Wertheim, J. O., & Janies, D. (2020). StrainHub: a phylogenetic tool to construct pathogen transmission networks. *Bioinformatics*, 36(3), 945–947. <https://doi.org/10.1093/BIOINFORMATICS/BTZ646>
- De Meulenaere, K., Cuypers, B., Gamboa, D., Laukens, K., & Rosanas-Urgell, A. (2023). A new Plasmodium vivax reference genome for South American isolates. *BMC Genomics*, 24(1). <https://doi.org/10.1186/s12864-023-09707-5>
- de Oliveira, T. C., Corder, R. M., Early, A., Rodrigues, P. T., Ladeia-Andrade, S., Alves, J. M. P., Neafsey, D. E., & Ferreira, M. U. (2020). Population genomics reveals the expansion of highly inbred plasmodium vivax lineages in the main malaria hotspot of Brazil. *PLoS Neglected Tropical Diseases*, 14(10), 1–16. <https://doi.org/10.1371/journal.pntd.0008808>
- Delgado-Ratto, C., Gamboa, D., Soto-Calle, V. E., Van den Eede, P., Torres, E., Sánchez-Martínez, L., Contreras-Mancilla, J., Rosanas-Urgell, A., Rodriguez Ferrucci, H., Llanos-Cuentas, A., Erhart, A., Van geertruyden, J. P., & D'Alessandro, U. (2016). Population Genetics of Plasmodium vivax in the Peruvian Amazon. *PLoS Neglected Tropical Diseases*, 10(1).

<https://doi.org/10.1371/journal.pntd.0004376>

- Delgado-Ratto, C., Soto-Calle, V. E., Van Den Eede, P., Gamboa, D., Rosas, A., Abatih, E. N., Ferrucci, H. R., Llanos-Cuentas, A., Van Geertruyden, J.-P., Erhart, A., & D'alessandro, U. (2014). *Population structure and spatio-temporal transmission dynamics of Plasmodium vivax after radical cure treatment in a rural village of the Peruvian Amazon*. <http://www.malariajournal.com/content/13/1/8>
- Dharia, N. V, Bright, A. T., Westenberger, S. J., Barnes, S. W., Batalov, S., Kuhen, K., Borboa, R., Federe, G. C., McClean, C. M., Vinetz, J. M., Neyra, V., Llanos-Cuentas, A., Barnwell, J. W., Walker, J. R., & Winzeler, E. A. (2010). Whole-genome sequencing and microarray analysis of ex vivo Plasmodium vivax reveal selective pressure on putative drug resistance genes. *Proceedings of the National Academy of Sciences of the United States of America*, 107(46), 20045–20050. <https://doi.org/10.1073/pnas.1003776107>
- Díaz-Salazar, G. E., Vidarte Cano, R. R., Guerra Cabrera, J. M., Segami Makiya, J. M., Rojas Ramos, C. A., Romero Espinoza, J. R., Molina Rosas, O. R., Medrano Colmenares, R. C., Díaz Salazar, G. E., Vidarte Cano, R. R., Guerra Cabrera, J. M., Segami Makiya, J. M., Rojas Ramos, C. A., Romero Espinoza, J. R., Molina Rosas, O. R., & Medrano Colmenares, R. C. (2024). Incidencia acumulada y distribución espaciotemporal de la malaria en el Perú en el periodo 2000 al 2022. *Anales de La Facultad de Medicina*, 85(4), 414–420. <https://doi.org/10.15381/ANALES.V85I4.29502>
- Feo, O., Tobar, K., Pineda, B., Rosas-Aguirre, A., Vargas, D., & Soto, G. (2009). *Compartiendo lecciones aprendidas. Proyecto control de malaria en zonasfronterizas de la región andina: un enfoque comunitario-PAMAFRO*. <https://doi.org/10.13140/RG.2.1.2847.8486>
- Fernandez-Miñope, C., Delgado-Ratto, C., Contreras-Mancilla, J., Ferrucci, H. R., Llanos-Cuentas, A., Gamboa, D., & Van geertruyden, J. P. (2021). Towards one standard treatment for uncomplicated Plasmodium falciparum and Plasmodium vivax malaria: Perspectives from and for the Peruvian Amazon.

In *International Journal of Infectious Diseases* (Vol. 105, pp. 293–297). Elsevier B.V. <https://doi.org/10.1016/j.ijid.2021.02.042>

Fola, A. A., Kattenberg, E., Razook, Z., Gumal, D. L., Lee, S., Mehra, S., Bahlo, M., Kazura, J., Robinson, L. J., Laman, M., Mueller, I., & Barry, A. E. (2020). SNP barcodes provide higher resolution than microsatellite markers to measure *Plasmodium vivax* population genetics. *Malaria Journal*, 1–15. <https://doi.org/10.1186/s12936-020-03440-0>

Garcia Castillo, S. S., Abanto Alvarez, C., Rosas-Aguirre, Á., Acosta, C., Corder, R. M., Gómez, J., Guzmán, M., Speybroeck, N., Llanos-Cuentas, A., Castro, M. C., Rosanas-Urgell, A., Ferreira, M. U., Vinetz, J. M., Gamboa, D., & Torres, K. (2024). Recurrence patterns and evolution of submicroscopic and asymptomatic *Plasmodium vivax* infections in malaria-endemic areas of the Peruvian Amazon. *PLoS Neglected Tropical Diseases*, 18(10). <https://doi.org/10.1371/journal.pntd.0012566>

Gomez, J., Grosso, A., Guzman-Guzman, M., Castillo, S. G., Castro, M. C., Torres, K., Vinetz, J. M., & Gamboa, D. (2025). Human mobility and malaria risk in peri-urban and rural communities in the Peruvian Amazon. *PLoS Neglected Tropical Diseases*, 19(1). <https://doi.org/10.1371/journal.pntd.0012058>

Goudet, J. (2005). hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, 5(1), 184–186. <https://doi.org/10.1111/J.1471-8286.2004.00828.X>

Griffin, J. T., Ferguson, N. M., & Ghani, A. C. (2014). Estimates of the changing age-burden of *Plasmodium falciparum* malaria disease in sub-Saharan Africa. *Nature Communications*, 5. <https://doi.org/10.1038/ncomms4136>

Guerra, C. A., Citron, D. T., García, G. A., & Smith, D. L. (2019). Characterising malaria connectivity using malaria indicator survey data. *Malaria Journal*, 1–12. <https://doi.org/10.1186/s12936-019-3078-2>

Gunderson, A. K., Kumar, R. E., Recalde-Coronel, C., Vasco, L. E., Valle-Campos, A., Mena, C. F., Zaitchik, B. F., Lescano, A. G., Pan, W. K., & Janko, M. M.

- (2020). Malaria transmission and spillover across the peru–ecuador border: A spatiotemporal analysis. *International Journal of Environmental Research and Public Health*, 17(20), 1–9. <https://doi.org/10.3390/ijerph17207434>
- Gunderson, A. K., Recalde-Coronel, C., Zaitchick, B. F., Yori, P. P., Rengifo Pinedo, S., Paredes Olortegui, M., Kosek, M., Vinetz, J. M., & Pan, W. K. (2023). A prospective cohort study linking migration, climate, and malaria risk in the Peruvian Amazon. *Epidemiology and Infection*, 151. <https://doi.org/10.1017/S0950268823001838>
- Hajian-Tilaki, K. (2014). Sample size estimation in diagnostic test studies of biomedical informatics. *Journal of Biomedical Informatics*, 48, 193–204. <https://doi.org/10.1016/j.jbi.2014.02.013>
- Harrison, A. G. L., Mehra, S., Razook, Z., Tessier, N., Lee, S., Hetzel, M. W., Tavul, L., Laman, M., Makita, L., Amato, R., Miotto, O., Burke, N., Jensen, A., Kwiatkowski, D., Betuela, I., Siba, P. M., Bahlo, M., Mueller, I., & Barry, A. E. (2023). *Defining malaria parasite population subdivisions , transmission dynamics and infection origins using SNP barcodes*. April, 1–24.
- Hartl, D. L. ., & Clark, A. G. . (2007). *Principles of Population Genetics*.
- Henden, L., Lee, S., Mueller, I., Barry, A., & Bahlo, M. (2018). Identity-by-descent analyses for measuring population dynamics and selection in recombining pathogens. *PLOS Genetics*, 14(5), e1007279. <https://doi.org/10.1371/JOURNAL.PGEN.1007279>
- Hendry, J. A., Kwiatkowski, D., & McVean, G. (2021). Elucidating relationships between P.falciparum prevalence and measures of genetic diversity with a combined genetic-epidemiological model of malaria. *PLOS Computational Biology*, 17(8), e1009287. <https://doi.org/10.1371/JOURNAL.PCBI.1009287>
- Hupalo, D. N., Luo, Z., Melnikov, A., Sutton, P. L., Rogov, P., Escalante, A., Vallejo, A. F., Herrera, S., Arévalo-Herrera, M., Fan, Q., Wang, Y., Cui, L., Lucas, C. M., Durand, S., Sanchez, J. F., Christian Baldeviano, G., Lescano, A. G., Laman, M., Barnadas, C., ... Carlton, J. M. (2016). Population

genomics studies identify signatures of global dispersal and drug resistance in *Plasmodium vivax*. *Nature Genetics*, 48(8), 953–958. <https://doi.org/10.1038/ng.3588>

Ibrahim, A., Manko, E., Dombrowski, J. G., Campos, M., Benavente, E. D., Nolder, D., Sutherland, C. J., Nosten, F., Fernandez, D., Vélez-Tobón, G., Castaño, A. T., Aguiar, A. C. C., Pereira, D. B., da Silva Santos, S., Suarez-Mutis, M., Di Santi, S. M., Regina de Souza Baptista, A., Dantas Machado, R. L., Marinho, C. R. F., ... Campino, S. (2023). Population-based genomic study of *Plasmodium vivax* malaria in seven Brazilian states and across South America. *The Lancet Regional Health - Americas*, 18, 100420. <https://doi.org/10.1016/J.LANA.2022.100420>

Instituto Nacional de Estadística e Informática. (2018). *III Censo de Comunidades Nativas 2017. TOMO I*. <https://censo2017.inei.gob.pe/resultados-definitivos-de-las-comunidades-nativas-y-campesinas-2017/>

Janko, M. M., Araujo, A. L., Ascencio, E. J., Guedes, G. R., Vasco, L. E., Santos, R. O., Damasceno, C. P., Medrano, P. G., Chacón-Uscamaita, P. R., Gunderson, A. K., O'Malley, S., Kansara, P. H., Narvaez, M. B., Coombes, C., Pizzitutti, F., Salmon-Mulanovich, G., Zaitchik, B. F., Mena, C. F., Lescano, A. G., ... Pan, W. K. (2024). Study protocol: improving response to malaria in the Amazon through identification of inter-community networks and human mobility in border regions of Ecuador, Peru and Brazil. *BMJ Open*, 14(4). <https://doi.org/10.1136/bmjopen-2023-078911>

Janko, M. M., Recalde-Coronel, G. C., Damasceno, C. P., Salmón-Mulanovich, G., Barbieri, A. F., Lescano, A. G., Zaitchik, B. F., & Pan, W. K. (2023). The impact of sustained malaria control in the Loreto region of Peru: a retrospective, observational, spatially-varying interrupted time series analysis of the PAMAFRO program. *The Lancet Regional Health - Americas*, 20. <https://doi.org/10.1016/j.lana.2023.100477>

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21), 3070–3071.

<https://doi.org/10.1093/BIOINFORMATICS/BTR521>

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2014(1), 1–14. <https://doi.org/10.7717/PEERJ.281/TABLE-6>

Kanaka, K. K., Sukhija, N., Goli, R. C., Singh, S., Ganguly, I., Dixit, S. P., Dash, A., & Malik, A. A. (2023). On the concepts and measures of diversity in the genomics era. *Current Plant Biology*, 33, 100278. <https://doi.org/10.1016/J.CPB.2023.100278>

Kar, N. P., Kumar, A., Singh, O. P., Carlton, J. M., & Nanda, N. (2014). A review of malaria transmission dynamics in forest ecosystems. *Parasites & Vectors* 2014 7:1, 7(1), 1–12. <https://doi.org/10.1186/1756-3305-7-265>

Kassambara, A. (2018). fastqcr: Quality Control of Sequencing Data. R package version 0.1.2. CRAN: Contributed Packages. <https://doi.org/10.32614/CRAN.PACKAGE.FASTQCR>

Kassambara, A. (2020). ggpubr: ‘ggplot2’ Based Publication Ready Plots. R package version 0.4.0. CRAN: Contributed Packages. <https://doi.org/10.32614/CRAN.PACKAGE.GGPUBR>

Kassambara, A. (2023a). Pipe-Friendly Framework for Basic Statistical Tests [R package rstatix version 0.7.2]. CRAN: Contributed Packages. <https://doi.org/10.32614/CRAN.PACKAGE.RSTATIX>

Kassambara, A. (2023b). Visualization of a Correlation Matrix using “ggplot2” [R package ggcorrplot version 0.1.4.1]. <https://doi.org/10.32614/CRAN.PACKAGE.GGCORRPLOT>

Kattenberg, J. H., Cabrera-Sosa, L., Figueroa-Ildefonso, E., Mutsaers, M., Monsieurs, P., Guetens, P., Infante, B., Delgado-Ratto, C., Gamboa, D., & Rosanas-Urgell, A. (2024b). Plasmodium vivax genomic surveillance in the Peruvian Amazon with Pv AmpliSeq assay. *PLoS Neglected Tropical*

Diseases, 18(7), e0011879. <https://doi.org/10.1371/journal.pntd.0011879>

- Kattenberg, J. H., Dijk, N. J. Van, Fernández-miño, C. A., & Guetens, P. (2023). *Molecular Surveillance of Malaria Using the PF AmpliSeq Custom Assay for Plasmodium falciparum Parasites from Dried Blood Spot DNA Isolates from Peru*. 13. <https://doi.org/10.1128/spectrum.00960-22>
- Kattenberg, J. H., Fernandez-Miño, C., van Dijk, N. J., Llacsahuanga Allcca, L., Guetens, P., Valdivia, H. O., Van geertruyden, J.-P., Rovira-Vallbona, E., Monsieurs, P., Delgado-Ratto, C., Gamboa, D., & Rosanas-Urgell, A. (2023). Malaria Molecular Surveillance in the Peruvian Amazon with a Novel Highly Multiplexed Plasmodium falciparum AmpliSeq Assay. *Microbiology Spectrum*, 11(2). <https://doi.org/10.1128/spectrum.00960-22>
- Kattenberg, J. H., Monsieurs, P., De Meyer, J., De Meulenaere, K., Sauve, E., de Oliveira, T. C., Ferreira, M. U., Gamboa, D., & Rosanas-Urgell, A. (2024a). Population genomic evidence of structured and connected Plasmodium vivax populations under host selection in Latin America. *Ecology and Evolution*, 14(3). <https://doi.org/10.1002/ece3.11103>
- Kattenberg, J. H., Nguyen, H. Van, Nguyen, H. L., Sauve, E., Nguyen, N. T. H., Chopo-Pizarro, A., Trimarsanto, H., Monsieurs, P., Guetens, P., Nguyen, X. X., Esbroeck, M. Van, Auburn, S., Nguyen, B. T. H., & Rosanas-Urgell, A. (2022). Novel highly-multiplexed AmpliSeq targeted assay for Plasmodium vivax genetic surveillance use cases at multiple geographical scales. *Frontiers in Cellular and Infection Microbiology*, 12. <https://doi.org/10.3389/fcimb.2022.953187>
- Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., & Prodöhl, P. A. (2013). diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, 4(8), 782–788. <https://doi.org/10.1111/2041-210X.12067>
- Knaus, B. J., & Grünwald, N. J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, 17(1), 44–53.

<https://doi.org/10.1111/1755-0998.12549>

- Kodama, Y., Shumway, M., & Leinonen, R. (2012). The sequence read archive: Explosive growth of sequencing data. *Nucleic Acids Research*, 40(D1), 2011–2013. <https://doi.org/10.1093/nar/gkr854>
- Koepfli, C., Ross, A., Kiniboro, B., Smith, T. A., Zimmerman, P. A., Siba, P., Mueller, I., & Felger, I. (2011). Multiplicity and diversity of plasmodium vivax infections in a highly endemic region in papua New Guinea. *PLoS Neglected Tropical Diseases*, 5(12), 1–7. <https://doi.org/10.1371/journal.pntd.0001424>
- Krueger, F. (2015). *Trim Galore [Software]*. Babraham Bioinformatics. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
- Labbé, F., He, Q., Zhan, Q., Tiedje, K. E., Argyropoulos, D. C., Tan, M. H., Ghansah, A., Day, K. P., & Pascual, M. (2023). Neutral vs. non-neutral genetic footprints of Plasmodium falciparum multiclonal infections. *PLoS Computational Biology*, 19(1), 1–28. <https://doi.org/10.1371/journal.pcbi.1010816>
- LaVerriere, E., Schwabl, P., Carrasquilla, M., Taylor, A. R., Johnson, Z. M., Shieh, M., Panchal, R., Straub, T. J., Kuzma, R., Watson, S., Buckee, C. O., Andrade, C. M., Portugal, S., Crompton, P. D., Traore, B., Rayner, J. C., Corredor, V., James, K., Cox, H., ... Neafsey, D. E. (2022). Design and implementation of multiplexed amplicon sequencing panels to serve genomic epidemiology of infectious disease: A malaria case study. *Molecular Ecology Resources*, 22(6), 2285–2303. <https://doi.org/10.1111/1755-0998.13622>
- Lee, S., & Bahlo, M. (2016). moimix: an R package for assessing clonality in malaria infections. *Bioinformatics*, 32(13), 2056–2058. <https://bahlolab.github.io/moimix/>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760.

<https://doi.org/10.1093/bioinformatics/btp324>

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>

Lima-Junior, J. da C., & Pratt-Riccio, L. R. (2016). Major histocompatibility complex and malaria: Focus on *Plasmodium vivax* Infection. In *Frontiers in Immunology* (Vol. 7, Issue JAN). Frontiers Media S.A. <https://doi.org/10.3389/fimmu.2016.00013>

Lucchi, N. W., Oberstaller, J., Kissinger, J. C., & Udhayakumar, V. (2013). Malaria Diagnostics and Surveillance in the Post-Genomic Era. *Public Health Genomics*, 16(1–2), 37–43. <https://doi.org/10.1159/000345607>. Malaria

Manrique, P., Miranda-Alban, J., Alarcon-Baldeon, J., Ramirez, R., Carrasco-Escobar, G., Herrera, H., Guzman-Guzman, M., Rosas-Aguirre, A., Llanos-Cuentas, A., Vinetz, J. M., Escalante, A. A., & Gamboa, D. (2019). Microsatellite analysis reveals connectivity among geographically distant transmission zones of *Plasmodium vivax* in the Peruvian Amazon: A critical barrier to regional malaria elimination. *PLoS Neglected Tropical Diseases*, 13(11). <https://doi.org/10.1371/journal.pntd.0007876>

Manrique Valverde, P. C. (2018). *Epidemiología molecular de Plasmodium vivax en cuatro comunidades ribereñas del distrito de Mazán en la Amazonía Peruana* [Universidad Peruana Cayetano Heredia]. <http://repositorio.upch.edu.pe/handle/upch/3671>

Manske, M., Miotto, O., Campino, S., Auburn, S., Almagro-Garcia, J., Maslen, G., O'Brien, J., Djimde, A., Doumbo, O., Zongo, I., Ouedraogo, J. B., Michon, P., Mueller, I., Siba, P., Nzila, A., Borrmann, S., Kiara, S. M., Marsh, K., Jiang, H., ... Kwiatkowski, D. P. (2012). Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature*, 487(7407), 375–

379. <https://doi.org/10.1038/nature11174>

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110.20>

Mendes, C., Dias, F., Figueiredo, J., Mora, V. G., Cano, J., de Sousa, B., do Rosário, V. E., Benito, A., Berzosa, P., & Arez, A. P. (2011). Duffy negative antigen is no longer a barrier to *Plasmodium vivax* - molecular evidences from the African West Coast (Angola and Equatorial Guinea). *PLoS Neglected Tropical Diseases*, 5(6). <https://doi.org/10.1371/journal.pntd.0001192>

Ministerio de Salud del Perú (MINSA). (2017). *Plan Malaria Cero Periodo 2017–2021: Resolución Ministerial No 244–2017*. https://docs.bvsalud.org/biblioref/2019/04/965065/rm_244-2017-minsa.pdf

Ministerio de Salud del Perú (MINSA). (2022). *Documento técnico: Plan hacia la eliminación de la malaria en el Perú 2022-2030*. 1–60. [https://bvs.minsa.gob.pe/local/fi-admin/RM-034-2022 MINSA.pdf](https://bvs.minsa.gob.pe/local/fi-admin/RM-034-2022%20MINSA.pdf)

Ministerio de Salud del Perú (MINSA). (2024). *Sala de Situación de malaria SE52*. <https://www.dge.gob.pe/portal/docs/vigilancia/sala/2024/SE52/malaria.pdf>

Ministerio de Salud del Perú (MINSA). (2025). *NORMA TÉCNICA DE SALUD PARA LA ATENCIÓN INTEGRAL DE MALARIA NO COMPLICADA Y MALARIA GRAVE EN EL PERÚ*.

Ministerio de Salud del Perú (MINSA). (2026). *Sala Situacional de Enfermedades Metaxénicas - Malaria*. https://app7.dge.gob.pe/maps2/shiny_metaxenicas_web/

Montenegro, C. C., Bustamante-Chauca, T. P., Pajuelo Reyes, C., Bernal, M., Gonzales, L., Tapia-Limonchi, R., Tejedo, J. R., & Chenet, S. M. (2021). *Plasmodium falciparum* outbreak in native communities of Condorcanqui,

- Amazonas, Perú. *Malaria Journal*, 20(1). <https://doi.org/10.1186/s12936-021-03608-2>
- Müller, K., & Wickham, H. (2023). Simple Data Frames [R package tibble version 3.3.1]. CRAN: *Contributed Packages*. <https://doi.org/10.32614/CRAN.package.tibble>
- National Center for Biotechnology Information (NCBI). (2024). *SRA Toolkit Documentation*. <https://www.ncbi.nlm.nih.gov/sra/docs/sradownload/>
- Neafsey, D. E., & Volkman, S. K. (2017). Malaria Genomics in the Era of Eradication. *Cold Spring Harbor Perspectives in Medicine*, 7(8), a025544. <https://doi.org/10.1101/CSHPERSPECT.A025544>
- Nei, M. (1973). Analysis of Gene Diversity in Subdivided Populations. *Proceedings of the National Academy of Sciences*, 70(12), 3321–3323. <https://doi.org/10.1073/PNAS.70.12.3321>
- Ngwana-Joseph, G. C., Phelan, J. E., Manko, E., Dombrowski, J. G., da Silva Santos, S., Suarez-Mutis, M., Vélez-Tobón, G., Tobón Castaño, A., Machado, R. L. D., Marinho, C. R. F., Nolder, D., Nosten, F., Sutherland, C. J., Campino, S., & Clark, T. G. (2024). Genomic analysis of global *Plasmodium vivax* populations reveals insights into the evolution of drug resistance. *Nature Communications*, 15(1). <https://doi.org/10.1038/s41467-024-54964-x>
- Nielsen, R., & Slatkin, M. (2013). *An Introduction to Population Genetics*.
- Ntumngia, F. B., Thomson-Luque, R., Pires, C. V., & Adams, J. H. (2016). The role of the human duffy antigen receptor for chemokines in malaria susceptibility: Current opinions and future treatment prospects. *Journal of Receptor, Ligand and Channel Research*, 9. <https://doi.org/10.2147/JRLCR.S99725>,
- Oyegoke, O. O., Maharaj, L., Akoniyon, O. P., Kwoji, I., Roux, A. T., Adewumi, T. S., Maharaj, R., Oyebola, B. T., Adeleke, M. A., & Okpeku, M. (2022). Malaria diagnostic methods with the elimination goal in view. *Parasitology*

Research, 121(7), 1867. <https://doi.org/10.1007/S00436-022-07512-9>

- Palma-Cuero, M., Machado, M. B., Graça, J. T. B., Dos Anjos, N. B., Pereira, R. S., & Suárez-Mutis, M. C. (2022). Malaria at international borders: challenges for elimination on the remote Brazil-Peru border. *Revista Do Instituto de Medicina Tropical de Sao Paulo*, 64. <https://doi.org/10.1590/S1678-9946202264029>
- Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3), 526–528. <https://doi.org/10.1093/BIOINFORMATICS/BTY633>
- Pearson, R. D., Amato, R., Auburn, S., Miotto, O., Almagro-Garcia, J., Amaratunga, C., Suon, S., Mao, S., Noviyanti, R., Trimarsanto, H., Marfurt, J., Anstey, N. M., William, T., Boni, M. F., Dolecek, C., Tran, H. T., White, N. J., Michon, P., Siba, P., ... Kwiatkowski, D. P. (2016). Genomic analysis of local variation and recent evolution in *Plasmodium vivax*. *Nature Genetics*, 48(8), 959. <https://doi.org/10.1038/NG.3599>
- Pebesma, E. (2018). Simple features for R: Standardized support for spatial vector data. *R Journal*, 10(1), 439–446. <https://doi.org/10.32614/RJ-2018-009>
- Pinedo-Cancino, V., Arista, K. M., Baldeviano, G. C., Saavedra-Langer, R., Arana, A., Vásquez-Chasnamote, M. E., Valle-Campos, A., Castro, J. C., Ventocilla, J. A., Smith, E. S., Lescano, A. G., & Ruíz-Mesia, L. (2024). Unravelling heterogeneous malaria transmission dynamics in the Peruvian Amazon: insights from a cross-sectional survey. *Malaria Journal*, 23(1). <https://doi.org/10.1186/s12936-024-05032-8>
- Price, R. N., Tjitra, E., Guerra, C. A., Yeung, S., White, N. J., & Anstey, N. M. (2007). *Vivax malaria: neglected and not benign*. <https://pmc.ncbi.nlm.nih.gov/articles/PMC2653940/>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based

linkage analyses. *American Journal of Human Genetics*, 81(3), 559–575.
<https://doi.org/10.1086/519795>

Ramirez, R. (2020). *ESTRATEGIAS DE CONTROL BASADO EN LA DETECCIÓN DE INFECCIONES POR PCR EN TIEMPO REAL Y SU EFECTO EN LA DIVERSIDAD GENÉTICA Y LA REPRODUCCIÓN DE Plasmodium vivax EN LAS CUENCAS MAZAN Y NAPO*. Universidad Peruana Cayetano Heredia.

Ramirez, R. (2024). *MICRO-EPIDEMIOLOGÍA MOLECULAR DE LAS INFECCIONES QUE CONTRIBUYEN A MANTENER LA TRANSMISIÓN DE Plasmodium vivax EN COMUNIDADES EN VÍAS DE ELIMINACIÓN DE LA MALARIA EN LA REGIÓN LORETO*. Universidad Peruana Cayetano Heredia.

Ramirez, R., Torres, K., Rodríguez, P., Acosta, C., Guzmán-Guzmán, M., Llanos-Cuentas, A., Vinetz, J. M., Escalante, A. A., & Gamboa, D. (2025). High Prevalence, Genetic Diversity and Temporal Differentiation of Plasmodium vivax in a Remote Hard-to-Reach Community from the Peruvian Amazon Region. *The American Journal of Tropical Medicine and Hygiene*.
<https://doi.org/10.4269/AJTMH.24-0662>

Rodríguez, J. C. P., Olivera, M. J., Cantillo, L. A., & Chaparro-Narváez, P. (2024). Changes in the endemic-epidemic pattern of malaria in Colombia, 1978-2021. *Revista Da Sociedade Brasileira de Medicina Tropical*, 57, e00405.
<https://doi.org/10.1590/0037-8682-0364-2023>

Rodriguez, M. H. (2021). Residual Malaria: Limitations of Current Vector Control Strategies to Eliminate Transmission in Residual Foci. *Journal of Infectious Diseases*, 223, S55–S60. <https://doi.org/10.1093/infdis/jiaa582>

Rojas Muro, L. M. (2024). *Factores asociados a la prevalencia de malaria en tres comunidades nativas del distrito de Rio Santiago, Condorcanqui* [Universidad Peruana Cayetano Heredia].
<https://repositorio.upch.edu.pe/handle/20.500.12866/15234>

Roques, M., Bindschedler, A., Beyeler, R., & Heussler, V. T. (2023). Same, same

but different: Exploring Plasmodium cell division during liver stage development. In *PLoS Pathogens* (Vol. 19, Issue 3). Public Library of Science. <https://doi.org/10.1371/JOURNAL.PPAT.1011210>

Rosado, J., White, M. T., Longley, R. J., Lacerda, M., Monteiro, W., Brewster, J., Sattabongkot, J., Guzman-Guzman, M., Llanos-Cuentas, A., Vinetz, J. M., Gamboa, D., & Mueller, I. (2021). Heterogeneity in response to serological exposure markers of recent plasmodium vivax infections in contrasting epidemiological contexts. *PLoS Neglected Tropical Diseases*, 15(2). <https://doi.org/10.1371/journal.pntd.0009165>

Rosas-Aguirre, A., Gamboa, D., Manrique, P., Conn, J. E., Moreno, M., Lescano, A. G., Sanchez, J. F., Rodriguez, H., Silva, H., Llanos-Cuentas, A., & Vinetz, J. M. (2016). Epidemiology of Plasmodium vivax malaria in Peru. *American Journal of Tropical Medicine and Hygiene*, 95, 133–144. <https://doi.org/10.4269/ajtmh.16-0268>

Rosas-Aguirre, A., Guzman-Guzman, M., Chuquiyauri, R., Moreno, M., Manrique, P., Ramirez, R., Carrasco-Escobar, G., Rodriguez, H., Speybroeck, N., Conn, J. E., Gamboa, D., Vinetz, J. M., & Llanos-Cuentas, A. (2021). Temporal and Microspatial Heterogeneity in Transmission Dynamics of Coendemic Plasmodium vivax and Plasmodium falciparum in Two Rural Cohort Populations in the Peruvian Amazon. *Journal of Infectious Diseases*, 223(8), 1466–1477. <https://doi.org/10.1093/infdis/jiaa526>

Rosas-Aguirre, A., Guzman-Guzman, M., Gamboa, D., Chuquiyauri, R., Ramirez, R., Manrique, P., Carrasco-Escobar, G., Puemape, C., Llanos-Cuentas, A., & Vinetz, J. M. (2017). Micro-heterogeneity of malaria transmission in the Peruvian Amazon: A baseline assessment underlying a population-based cohort study. *Malaria Journal*, 16(1), 1–19. <https://doi.org/10.1186/S12936-017-1957-Y/FIGURES/4>

Rosas-aguirre, A., Moreno, M., Moreno-gutierrez, D., Llanos-cuentas, A., Saavedra, M., Contreras-mancilla, J., Barboza, J., Alava, F., Aguirre, K., Carrasco, G., Prussing, C., Vinetz, J., Conn, J. E., & Speybroeck, N. (2021).

Integrating Parasitological and Entomological Observations to Understand Malaria Transmission in Riverine Villages in the Peruvian Amazon. 223(Suppl 2), 99–110. <https://doi.org/10.1093/infdis/jiaa496>

Rosas-Aguirre, A., Speybroeck, N., Llanos-Cuentas, A., Rosanas-Urgell, A., Carrasco-Escobar, G., Rodriguez, H., Gamboa, D., Contreras-Mancilla, J., Alava, F., Soares, I. S., Remarque, E., D'Alessandro, U., & Erhart, A. (2015). Hotspots of Malaria Transmission in the Peruvian Amazon: Rapid Assessment through a Parasitological and Serological Survey. *PLOS ONE*, 10(9), e0137458. <https://doi.org/10.1371/JOURNAL.PONE.0137458>

Rousset, F., Ayala, F. J., & Renaud, F. (2005). “‘ Clonal ’” population structure of the malaria agent *Plasmodium falciparum* in high-infection regions. <https://doi.org/10.1073/pnas.0508871102>

Saavedra-Samillán, M., Burgos, F., García Huamán, F., Valdivia, H. O., Gamboa, D., & Chenet, S. M. (2024). Spatiotemporal dynamics of malaria and climate influence on its incidence in Condorcanqui Province, 2005–2022. *Malaria Journal* , 23(1). <https://doi.org/10.1186/s12936-024-05193-6>

Sanchez, J. F., Carnero, A. M., Rivera, E., Rosales, L. A., Christian Baldeviano, G., Asencios, J. L., Edgel, K. A., Vinetz, J. M., & Lescano, A. G. (2017). Unstable Malaria transmission in the southern peruvian amazon and its association with gold mining, madre de dios, 2001-2012. *American Journal of Tropical Medicine and Hygiene*, 96(2), 304–311. <https://doi.org/10.4269/AJTMH.16-0030/-/DC2/SD2.PDF>

Sanchez, J. F., Valdivia, H. O., & Pannebaker, D. L. (2025). Malaria Surveillance: A Cross-sectional Study in the Peruvian Amazon Basin, 2012–2023. *Journal of Infectious Diseases*, 231(Suppl 1), S58–S65. <https://doi.org/10.1093/infdis/jiae570>

Sato, S. (2021). Plasmodium—a brief introduction to the parasites causing human malaria and their basic biology. In *Journal of Physiological Anthropology* (Vol. 40, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s40101-020->

- Sayers, E. W., Beck, J., Bolton, E. E., Brister, J. R., Chan, J., Connor, R., Feldgarden, M., Fine, A. M., Funk, K., Hoffman, J., Kannan, S., Kelly, C., Klimke, W., Kim, S., Lathrop, S., Marchler-Bauer, A., Murphy, T. D., O’Sullivan, C., Schmieder, E., ... Pruitt, K. D. (2025). Database resources of the National Center for Biotechnology Information in 2025. *Nucleic Acids Research*, 53(D1), D20–D29. <https://doi.org/10.1093/nar/gkae979>
- Segura-Alabart, N., Serratosa, F., Gómez, S., & Fernández, A. (2022). Nonunique UPGMA clusterings of microsatellite markers. *Briefings in Bioinformatics*, 23(5), bbac312. <https://doi.org/10.1093/BIB/BBAC312>
- Slatkin, M. (1987). Gene flow and the geographic structure of natural populations. *Science*, 236(4803), 787–792. <https://doi.org/10.1126/science.3576198>
- Soontarawirat, I., Andolina, C., Paul, R., Day, N. P. J., Nosten, F., Woodrow, C. J., & Imwong, M. (2017). Plasmodium vivax genetic diversity and heterozygosity in blood samples and resulting oocysts at the Thai – Myanmar border. *Malaria Journal*, 1–11. <https://doi.org/10.1186/s12936-017-2002-x>
- Soto-Calle, V., Rosas-Aguirre, A., Llanos-Cuentas, A., Abatih, E., Dedeken, R., Rodriguez, H., Rosanas-Urgell, A., Gamboa, D., Dálessandro, U., Erhart, A., & Speybroeck, N. (2017). Spatio-Temporal analysis of malaria incidence in the Peruvian Amazon Region between 2002 and 2013. *Scientific Reports*, 7. <https://doi.org/10.1038/srep40350>
- Sundqvist, L., Keenan, K., Zackrisson, M., Prodöhl, P., & Kleinhans, D. (2016). Directional genetic differentiation and relative migration. *Ecology and Evolution*, 6(11), 3461–3475. <https://doi.org/10.1002/ECE3.2096>
- Sutanto, E., Pava, Z., Echeverry, D. F., Lopera-Mesa, T. M., Montenegro, L. M., Yasnot-Acosta, M. F., Benavente, E. D., Pearson, R. D., Herrera, S., Arévalo-Herrera, M., Trimarsanto, H., Rumaseb, A., Noviyanti, R., Kwiatkowski, D. P., Price, R. N., & Auburn, S. (2023). Genomics of Plasmodium vivax in Colombia reveals evidence of local bottle-necking and inter-country

- connectivity in the Americas. *Scientific Reports*, 13(1), 1–17. <https://doi.org/10.1038/S41598-023-46076-1>;SUBJMETA
- Sutton, P. L., Neyra, V., Hernandez, J. N., & Branch, O. L. H. (2009). Plasmodium falciparum and Plasmodium vivax infections in the Peruvian Amazon: Propagation of complex, multiple allele-type infections without super-infection. *American Journal of Tropical Medicine and Hygiene*, 81(6), 950–960. <https://doi.org/10.4269/ajtmh.2009.09-0132>
- Suwanarusk, R., Cooke, B. M., Dondorp, A. M., Silamut, K., Sattabongkot, J., White, N. J., & Udomsangpetch, R. (2004). The Deformability of Red Blood Cells Parasitized by Plasmodium falciparum and P. vivax. *Journal of Infectious Diseases*, 189(2), 190–194. <https://doi.org/10.1086/380468>,
- Tambo, M., Mwinga, M., & Mumbengegwi, D. R. (2018). Loop-mediated isothermal amplification (LAMP) and Polymerase Chain Reaction (PCR) as quality assurance tools for Rapid Diagnostic Test (RDT) malaria diagnosis in Northern Namibia. *PLoS ONE*, 13(12). <https://doi.org/10.1371/JOURNAL.PONE.0206848>,
- Tang, Y., Horikoshi, M., & Li, W. (2016). Ggfortify: Unified interface to visualize statistical results of popular r packages. *R Journal*, 8(2), 478–489. <https://doi.org/10.32614/RJ-2016-060>
- Tangpukdee, N., Duangdee, C., Wilairatana, P., & Krudsood, S. (2009). Malaria Diagnosis: A Brief Review. *The Korean Journal of Parasitology*, 47(2), 93. <https://doi.org/10.3347/KJP.2009.47.2.93>
- Tessema, S., Wesolowski, A., Chen, A., Murphy, M., Wilhelm, J., Mupiri, A. R., Ruktanonchai, N. W., Alegana, V. A., Tatem, A. J., Tambo, M., Didier, B., Cohen, J. M., Bennett, A., Sturrock, H. J., Gosling, R., Hsiang, M. S., Smith, D. L., Mumbengegwi, D. R., Smith, J. L., & Greenhouse, B. (2019). Using parasite genetic and human mobility data to infer local and cross- border malaria connectivity in Southern Africa. *ELife*, 8, 1–20.

<https://doi.org/10.7554/eLife.43510>

- Thrusfield, M., Christley, R., Brown, H., Diggle, P. J., French, N., Howe, K., Kelly, L., O'Connor, A., Sargeant, J., & Wood, H. (2017). Veterinary Epidemiology: Fourth Edition. *Veterinary Epidemiology: Fourth Edition*, 1–861. <https://doi.org/10.1002/9781118280249>
- Torres, K., Alava, F., Soto-Calle, V., Llanos-Cuentas, A., Rodriguez, H., Llacsahuanga, L., Gamboa, D., & Vinetz, J. (2020). Malaria Situation in the Peruvian Amazon during the COVID-19 Pandemic. *The American Journal of Tropical Medicine and Hygiene*, 103(5), 1773. <https://doi.org/10.4269/AJTMH.20-0889>
- Trimarsanto, H., Amato, R., Pearson, R. D., Sutanto, E., Noviyanti, R., Trianty, L., Marfurt, J., Pava, Z., Echeverry, D. F., Lopera-Mesa, T. M., Montenegro, L. M., Tobón-Castaño, A., Grigg, M. J., Barber, B., William, T., Anstey, N. M., Getachew, S., Petros, B., Aseffa, A., ... Auburn, S. (2022). A molecular barcode and web-based data analysis tool to identify imported *Plasmodium vivax* malaria. *Communications Biology*, 5(1), 1–10. <https://doi.org/10.1038/s42003-022-04352-2>
- Van den Eede, P., Van der Auwera, G., Delgado, C., Huyse, T., Soto-Calle, V. E., Gamboa, D., Grande, T., Rodriguez, H., Llanos, A., Anné, J., & Erhart, A. (2010). *Multilocus genotyping reveals high heterogeneity and strong local population structure of the Plasmodium vivax population in the Peruvian Amazon*. <http://www.malariajournal.com/content/9/1/151>
- Villena, F. E., Sanchez, J. F., Nolasco, O., Braga, G., Ricopa, L., Barazorda, K., Salas, C. J., Lucas, C., Lizewski, S. E., Joya, C. A., Gamboa, D., Delgado-Ratto, C., & Valdivia, H. O. (2022). Drug resistance and population structure of *Plasmodium falciparum* and *Plasmodium vivax* in the Peruvian Amazon. *Scientific Reports*, 12(1). <https://doi.org/10.1038/s41598-022-21028-3>
- Waltmann, A., Koepfli, C., Tessier, N., Karl, S., Fola, A., Darcy, A. W., Wini, L., Harrison, G. L. A., Barnadas, C., Jennison, C., Karunajeewa, H., Boyd, S.,

- Whittaker, M., Kazura, J., Bahlo, M., Mueller, I., & Barry, A. E. (2018). Increasingly inbred and fragmented populations of *Plasmodium vivax* associated with the eastward decline in malaria transmission across the Southwest Pacific. *PLoS Neglected Tropical Diseases*, 12(1), 1–21. <https://doi.org/10.1371/journal.pntd.0006146>
- Weir, B. S., & Cockerham, C. C. (1984). ESTIMATING F-STATISTICS FOR THE ANALYSIS OF POPULATION STRUCTURE. *Evolution*, 38(6), 1358–1370. <https://doi.org/10.1111/J.1558-5646.1984.TB05657.X>
- Whitlock, M. C., & McCauley, D. E. (1999). *Indirect measures of gene flow and migration : $F_{ST} = 1 / (4 N m + 1)$* . 82(October 1998), 117–125.
- Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. *The Grammar of Graphics*. <https://doi.org/10.1007/0-387-28695-0>
- Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2014). dplyr: A Grammar of Data Manipulation. CRAN: Contributed Packages. <https://doi.org/10.32614/CRAN.package.dplyr>
- Wickham, H., Vaughan, D., & Girlich, M. (2023). Tidy Messy Data [R package tidyr version 1.3.2]. CRAN: Contributed Packages. <https://doi.org/10.32614/CRAN.package.tidyr>
- Wingett, S. W., & Andrews, S. (2018). Fastq screen: A tool for multi-genome mapping and quality control . *F1000Research*, 7(May), 1–14. <https://doi.org/10.12688/f1000research.15931.1>
- World Health Organization. (2011). *Universal access to malaria diagnostic testing, an Operational Manual*. <https://iris.who.int/handle/10665/44657>
- World Health Organization. (2025). World Malaria Report 2025. In *World Health: Vol. WHO/HTM/GM* (Issue December). <https://doi.org/ISBN 978 92 4 1564403>
- Xu, S., Dai, Z., Guo, P., Fu, X., Liu, S., Zhou, L., Tang, W., Feng, T., Chen, M.,

- Zhan, L., Wu, T., Hu, E., Jiang, Y., Bo, X., & Yu, G. (2021). ggtreeExtra: Compact Visualization of Richly Annotated Phylogenetic Data. *Molecular Biology and Evolution*, 38(9), 4039–4042. <https://doi.org/10.1093/MOLBEV/MSAB166>
- Yang, N., Zhang, H., Han, X., Liu, Z., & Lu, Y. (2024). Advancements and applications of loop-mediated isothermal amplification technology: a comprehensive overview. *Frontiers in Microbiology*, 15, 1406632. <https://doi.org/10.3389/FMICB.2024.1406632/BIBTEX>
- Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. Y. (2017). ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8(1), 28–36. <https://doi.org/10.1111/2041-210X.12628>
- Zheng, X., Gogarten, S. M., Lawrence, M., Stilp, A., Conomos, M. P., Weir, B. S., Laurie, C., & Levine, D. (2017). SeqArray-a storage-efficient high-performance data format for WGS variant calls. *Bioinformatics (Oxford, England)*, 33(15), 2251–2257.

<https://doi.org/10.1093/BIOINFORMATICS/BTX145>

XI. ANEXOS

Anexo 1. Tabla de minería de datos. Relación de la data genómica disponible en bases de datos públicas, incluyendo información sobre el tipo de secuenciación, identificadores asociados (Biosample, SRA, Bioproject y RUN), así como datos de procedencia geográfica (comunidad, distrito, provincia y región), año de colección y referencia bibliográfica correspondiente.

Biosample	Nombre de la muestra	Tipo de Data	Plataforma de secuenciación	SRA	Bioproject	RUN	Comunidad	Distrito	Provincia	Región	Año	Referencia
SAMN00109877	IQ07	WGS	ILLUMINA	SRS113792	PRJNA51913	SRR073125		Iquitos	Maynas	Loreto	2007	Dharia et al., 2010
SAMN08452356	PQPC-029-0	WGS	ILLUMINA	SRS2910041	PRJNA432819	SRR6668513	Padrecocha	Punchana	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452357	PQPC-029-84	WGS	ILLUMINA	SRS2909997	PRJNA432819	SRR6668560	Padrecocha	Punchana	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452358	PQPC-035-0	WGS	ILLUMINA	SRS2910026	PRJNA432819	SRR6668528	Padrecocha	Punchana	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452359	PQPC-035-35	WGS	ILLUMINA	SRS2910023	PRJNA432819	SRR6668531	Padrecocha	Punchana	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452295	PQPC-045-0	WGS	ILLUMINA	SRS2909991	PRJNA432819	SRR6668566	Padrecocha	Punchana	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452297	PQPC-047-43	WGS	ILLUMINA	SRS2909989	PRJNA432819	SRR6668568	Padrecocha	Punchana	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452300	PQPC-071-105	WGS	ILLUMINA	SRS2909999	PRJNA432819	SRR6668558	Padrecocha	Punchana	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452302	PQPC-088-91	WGS	ILLUMINA	SRS2910011	PRJNA432819	SRR6668544	Padrecocha	Punchana	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452304	PQPC-113-76	WGS	ILLUMINA	SRS2910013	PRJNA432819	SRR6668542	Padrecocha	Punchana	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452306	PQPC-117-52	WGS	ILLUMINA	SRS2910007	PRJNA432819	SRR6668548	Padrecocha	Punchana	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452307	PQPC-123-0	WGS	ILLUMINA	SRS2910016	PRJNA432819	SRR6668547	Padrecocha	Punchana	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452308	PQPC-125-0	WGS	ILLUMINA	SRS2910009	PRJNA432819	SRR6668546	Padrecocha	Punchana	Maynas	Loreto	2008	Cowell et al., 2018

SAMN08452309	PQPC-125-80	WGS	ILLUMINA	SRS2910002	PRJNA432819	SRR6668553	Padrecocha	Punchana	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452311	PQPC-128-87	WGS	ILLUMINA	SRS2910049	PRJNA432819	SRR6668504	Padrecocha	Punchana	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452313	PQPC-135-120	WGS	ILLUMINA	SRS2910051	PRJNA432819	SRR6668502	Padrecocha	Punchana	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452315	PQPC-139-121	WGS	ILLUMINA	SRS2910045	PRJNA432819	SRR6668508	Padrecocha	Punchana	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452316	PQSC-010-0	WGS	ILLUMINA	SRS2910057	PRJNA432819	SRR6668509	Santa Clara	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452318	PQSC-013-52	WGS	ILLUMINA	SRS2910046	PRJNA432819	SRR6668507	Santa Clara	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452319	PQSC-032-0	WGS	ILLUMINA	SRS2910053	PRJNA432819	SRR6668500	Santa Clara	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452320	PQSC-034-0	WGS	ILLUMINA	SRS2910052	PRJNA432819	SRR6668501	Santa Clara	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452322	PQSC-042-134	WGS	ILLUMINA	SRS2910021	PRJNA432819	SRR6668534	Santa Clara	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452324	PQSC-058-30	WGS	ILLUMINA	SRS2910019	PRJNA432819	SRR6668536	Santa Clara	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452325	PQSC-088-178	WGS	ILLUMINA	SRS2910027	PRJNA432819	SRR6668539	Santa Clara	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452326	PQSC-105-0	WGS	ILLUMINA	SRS2910017	PRJNA432819	SRR6668538	Santa Clara	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452327	PQSC-105-32	WGS	ILLUMINA	SRS2910014	PRJNA432819	SRR6668541	Santa Clara	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452291	H-032-28	WGS	ILLUMINA	SRS2909995	PRJNA432819	SRR6668562		Iquitos	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452292	H-071-0	WGS	ILLUMINA	SRS2909996	PRJNA432819	SRR6668561		Iquitos	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452328	PQSJ-004-28	WGS	ILLUMINA	SRS2910024	PRJNA432819	SRR6668540	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452329	PQSJ-058-0	WGS	ILLUMINA	SRS2910022	PRJNA432819	SRR6668533	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452330	PQSJ-058-40	WGS	ILLUMINA	SRS2910031	PRJNA432819	SRR6668532	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452331	PQSJ-076-0	WGS	ILLUMINA	SRS2910036	PRJNA432819	SRR6668518	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452332	PQSJ-103-0	WGS	ILLUMINA	SRS2910035	PRJNA432819	SRR6668519	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018

SAMN08452333	PQSJ-122-0	WGS	ILLUMINA	SRS2910034	PRJNA432819	SRR6668520	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452334	PQSJ-122-144	WGS	ILLUMINA	SRS2910054	PRJNA432819	SRR6668521	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452335	PQSJ-124-0	WGS	ILLUMINA	SRS2910032	PRJNA432819	SRR6668522	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452336	PQSJ-125-0	WGS	ILLUMINA	SRS2910042	PRJNA432819	SRR6668523	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452337	PQSJ-126-0	WGS	ILLUMINA	SRS2910030	PRJNA432819	SRR6668524	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452338	PQSJ-131-0	WGS	ILLUMINA	SRS2910029	PRJNA432819	SRR6668525	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452339	PQSJ-132-0	WGS	ILLUMINA	SRS2910044	PRJNA432819	SRR6668510	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452340	PQSJ-134-0	WGS	ILLUMINA	SRS2910043	PRJNA432819	SRR6668511	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452341	PQSJ-171-0	WGS	ILLUMINA	SRS2910008	PRJNA432819	SRR6668557	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452342	PQSJ-171-129	WGS	ILLUMINA	SRS2910000	PRJNA432819	SRR6668556	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452343	PQSJ-190-0	WGS	ILLUMINA	SRS2910015	PRJNA432819	SRR6668555	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452344	PQSJ-190-98	WGS	ILLUMINA	SRS2910001	PRJNA432819	SRR6668554	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452345	PQSJ-199-0	WGS	ILLUMINA	SRS2910040	PRJNA432819	SRR6668527	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452346	PQSJ-199-56	WGS	ILLUMINA	SRS2910028	PRJNA432819	SRR6668526	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452347	PQSJ-202-0	WGS	ILLUMINA	SRS2910004	PRJNA432819	SRR6668551	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452348	PQSJ-202-100	WGS	ILLUMINA	SRS2910005	PRJNA432819	SRR6668550	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452349	PQSJ-213-0	WGS	ILLUMINA	SRS2910033	PRJNA432819	SRR6668530	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452350	PQSJ-227-0	WGS	ILLUMINA	SRS2910025	PRJNA432819	SRR6668529	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452351	PQSJ-270-51	WGS	ILLUMINA	SRS2910038	PRJNA432819	SRR6668516	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452352	PQSJ-284-0	WGS	ILLUMINA	SRS2910037	PRJNA432819	SRR6668517	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018

SAMN08452353	PQSJ-284-71	WGS	ILLUMINA	SRS2910055	PRJNA432819	SRR6668514	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452354	PQSJ-294-0	WGS	ILLUMINA	SRS2910039	PRJNA432819	SRR6668515	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN08452355	PQSJ-294-73	WGS	ILLUMINA	SRS2910056	PRJNA432819	SRR6668512	San Juan	San Juan Bautista	Maynas	Loreto	2008	Cowell et al., 2018
SAMN02677068	Peru01	WGS	ILLUMINA	SRS693277	PRJNA240370	SRR1562535	Santo Tomas	San Juan Bautista	Maynas	Loreto	2008	Hupalo et al., 2016
SAMN02677072	Peru02	WGS	ILLUMINA	SRS694228	PRJNA240374	SRR1568179	Santo Tomas	San Juan Bautista	Maynas	Loreto	2008	Hupalo et al., 2016
SAMN03284397	Peru03	WGS	ILLUMINA	SRS819436	PRJNA240530	SRR1759047	Santo Tomas	San Juan Bautista	Maynas	Loreto	2008	Hupalo et al., 2016
SAMN03284398	Peru04	WGS	ILLUMINA	SRS819437	PRJNA240530	SRR1759122	Santo Tomas	San Juan Bautista	Maynas	Loreto	2008	Hupalo et al., 2016
SAMN03284579	Peru05	WGS	ILLUMINA	SRS819479	PRJNA240530	SRR1759139	Santo Tomas	San Juan Bautista	Maynas	Loreto	2008	Hupalo et al., 2016
SAMN02677100	Peru06	WGS	ILLUMINA	SRS696218	PRJNA240403	SRR1568211	Santo Tomas	San Juan Bautista	Maynas	Loreto	2008	Hupalo et al., 2016
SAMN02677101	Peru07	WGS	ILLUMINA	SRS693979	PRJNA240404	SRR1568172	Santo Tomas	San Juan Bautista	Maynas	Loreto	2008	Hupalo et al., 2016
SAMN02677102	Peru08	WGS	ILLUMINA	SRS693263	PRJNA240405	SRR1562512	Santo Tomas	San Juan Bautista	Maynas	Loreto	2008	Hupalo et al., 2016
SAMN02677103	Peru09	WGS	ILLUMINA	SRS693264	PRJNA240406	SRR1562513	Santo Tomas	San Juan Bautista	Maynas	Loreto	2008	Hupalo et al., 2016
SAMN02677104	Peru10	WGS	ILLUMINA	SRS693270	PRJNA240407	SRR1562521	Santo Tomas	San Juan Bautista	Maynas	Loreto	2008	Hupalo et al., 2016
SAMN02677065	Peru0839	WGS	ILLUMINA	SRS693923	PRJNA240367	SRR1568187		Mazan	Maynas	Loreto	2008	Hupalo et al., 2016
SAMN03284594	Peru612	WGS	ILLUMINA	SRS819493	PRJNA240530	SRR1759307		Mazan	Maynas	Loreto	2008	Hupalo et al., 2016
SAMN03284614	Peru697	WGS	ILLUMINA	SRS819579	PRJNA240530	SRR1759336		Mazan	Maynas	Loreto	2008	Hupalo et al., 2016
SAMN02677075	Peru721	WGS	ILLUMINA	SRS693911	PRJNA240377	SRR1568234		Mazan	Maynas	Loreto	2008	Hupalo et al., 2016
SAMN02677070	Peru791	WGS	ILLUMINA	SRS693268	PRJNA240372	SRR1568203		Mazan	Maynas	Loreto	2008	Hupalo et al., 2016
SAMN02677067	Peru830	WGS	ILLUMINA	SRS693953	PRJNA240369	SRR1568183		Mazan	Maynas	Loreto	2008	Hupalo et al., 2016
SAMN02677096	Peru852	WGS	ILLUMINA	SRS693941	PRJNA240399	SRR1568182		Mazan	Maynas	Loreto	2008	Hupalo et al., 2016

SAMN02677097	Peru858	WGS	ILLUMINA	SRS693949	PRJNA240400	SRR1568195		Mazan	Maynas	Loreto	2008	Hupalo et al., 2016
SAMN02677098	Peru869	WGS	ILLUMINA	SRS693276	PRJNA240401	SRR1562534		Mazan	Maynas	Loreto	2008	Hupalo et al., 2016
SAMN02677099	Peru872	WGS	ILLUMINA	SRS693977	PRJNA240402	SRR1568122		Mazan	Maynas	Loreto	2008	Hupalo et al., 2016
SAMN02677125	Peru1008	WGS	ILLUMINA	SRS693972	PRJNA240519	SRR1568206	Puerto Alegre	Morona	Datem del Marañón	Loreto	2010	Hupalo et al., 2016
SAMN02677128	Peru2025	WGS	ILLUMINA	SRS696216	PRJNA240522	SRR1568162	Puerto Alegre	Morona	Datem del Marañón	Loreto	2010	Hupalo et al., 2016
SAMN02677126	Peru1021	WGS	ILLUMINA	SRS693307	PRJNA240520	SRR1562567	Puerto Alegre	Morona	Datem del Marañón	Loreto	2011	Hupalo et al., 2016
SAMN02677127	Peru1022	WGS	ILLUMINA	SRS693476	PRJNA240521	SRR1562851	Puerto Alegre	Morona	Datem del Marañón	Loreto	2011	Hupalo et al., 2016
SAMN02677115	Peru3043_D0	WGS	ILLUMINA	SRS696213	PRJNA240509	SRR1568163	Padrecocha	Punchana	Maynas	Loreto	2011	Hupalo et al., 2016
SAMN02677122	Peru3133	WGS	ILLUMINA	SRS693925	PRJNA240516	SRR1564630	Padrecocha	Punchana	Maynas	Loreto	2011	Hupalo et al., 2016
SAMN02677123	Peru3136	WGS	ILLUMINA	SRS696217	PRJNA240517	SRR1568157	Padrecocha	Punchana	Maynas	Loreto	2011	Hupalo et al., 2016
SAMN02677114	Peru3232	WGS	ILLUMINA	SRS693926	PRJNA240508	SRR1568232	Padrecocha	Punchana	Maynas	Loreto	2012	Hupalo et al., 2016
SAMN02677118	Peru3270_D0	WGS	ILLUMINA	SRS693364	PRJNA240512	SRR1562624	Padrecocha	Punchana	Maynas	Loreto	2012	Hupalo et al., 2016
SAMN02677124	Peru3280	WGS	ILLUMINA	SRS693355	PRJNA240518	SRR1562615	Padrecocha	Punchana	Maynas	Loreto	2012	Hupalo et al., 2016
SAMN02677121	Peru3323	WGS	ILLUMINA	SRS693921	PRJNA240515	SRR1568184	Padrecocha	Punchana	Maynas	Loreto	2012	Hupalo et al., 2016
SAMN02677135	Peru4017	WGS	ILLUMINA	SRS693281	PRJNA240529	SRR1562538		Iquitos	Maynas	Loreto	2012	Hupalo et al., 2016
SAMN02677136	Peru4023	WGS	ILLUMINA	SRS693948	PRJNA240530	SRR1568117		Iquitos	Maynas	Loreto	2012	Hupalo et al., 2016
SAMN02677116	Peru3043_D87	WGS	ILLUMINA	SRS693581	PRJNA240510	SRR1562972	Padrecocha	Punchana	Maynas	Loreto	2011	Hupalo et al., 2016
SAMN02677117	Peru3043_D106	WGS	ILLUMINA	SRS693354	PRJNA240511	SRR1562614	Padrecocha	Punchana	Maynas	Loreto	2011	Hupalo et al., 2016
SAMN02677119	Peru3270_D43	WGS	ILLUMINA	SRS693345	PRJNA240513	SRR1562606	Padrecocha	Punchana	Maynas	Loreto	2012	Hupalo et al., 2016
SAMN02677120	Peru3270_D59	WGS	ILLUMINA	SRS693952	PRJNA240514	SRR1568787	Padrecocha	Punchana	Maynas	Loreto	2013	Hupalo et al., 2016

SAMN05853095	DBS-4	WGS	ILLUMINA	SRS1885920	PRJNA344889	SRR5136530	Iquitos	Iquitos	Maynas	Loreto	2015	N/A (PRJNA344889)
SAMN05853094	DBS-3	WGS	ILLUMINA	SRS1885936	PRJNA344889	SRR5136546	Padrecocha	Punchana	Maynas	Loreto	2015	N/A (PRJNA344889)
SAMN05853093	DBS-2	WGS	ILLUMINA	SRS1885924	PRJNA344889	SRR5136534	Padrecocha	Punchana	Maynas	Loreto	2015	N/A (PRJNA344889)
SAMN05853092	DBS-1	WGS	ILLUMINA	SRS1885934	PRJNA344889	SRR5136544	Padrecocha	Punchana	Maynas	Loreto	2015	N/A (PRJNA344889)
SAMN05853091	Sample_18	WGS	ILLUMINA	SRS1885927	PRJNA344889	SRR5136537	Padrecocha	Punchana	Maynas	Loreto	2007	N/A (PRJNA344889)
SAMN05853090	Sample_17	WGS	ILLUMINA	SRS1885935	PRJNA344889	SRR5136545	Padrecocha	Punchana	Maynas	Loreto	2007	N/A (PRJNA344889)
SAMN05853089	Sample_16	WGS	ILLUMINA	SRS1885939	PRJNA344889	SRR5136550	Padrecocha	Punchana	Maynas	Loreto	2007	N/A (PRJNA344889)
SAMN05853088	Sample_15	WGS	ILLUMINA	SRS1885937	PRJNA344889	SRR5136547	Padrecocha	Punchana	Maynas	Loreto	2007	N/A (PRJNA344889)
SAMN05853087	Sample_14	WGS	ILLUMINA	SRS1885921	PRJNA344889	SRR5136531	Padrecocha	Punchana	Maynas	Loreto	2007	N/A (PRJNA344889)
SAMN05853086	Sample_13	WGS	ILLUMINA	SRS1885923	PRJNA344889	SRR5136533	Padrecocha	Punchana	Maynas	Loreto	2007	N/A (PRJNA344889)
SAMN05853085	Sample_12	WGS	ILLUMINA	SRS1885933	PRJNA344889	SRR5136543	Padrecocha	Punchana	Maynas	Loreto	2007	N/A (PRJNA344889)
SAMN05853084	Sample_11	WGS	ILLUMINA	SRS1885930	PRJNA344889	SRR5136540	Padrecocha	Punchana	Maynas	Loreto	2007	N/A (PRJNA344889)
SAMN05853083	Sample_10	WGS	ILLUMINA	SRS1885922	PRJNA344889	SRR5136532	Padrecocha	Punchana	Maynas	Loreto	2007	N/A (PRJNA344889)
SAMN05853082	Sample_9	WGS	ILLUMINA	SRS1885932	PRJNA344889	SRR5136542	Padrecocha	Punchana	Maynas	Loreto	2007	N/A (PRJNA344889)
SAMN05853081	Sample_8	WGS	ILLUMINA	SRS1885925	PRJNA344889	SRR5136535	Padrecocha	Punchana	Maynas	Loreto	2007	N/A (PRJNA344889)
SAMN05853080	Sample_7	WGS	ILLUMINA	SRS1885940	PRJNA344889	SRR5136549	Santa Clara	San Juan Bautista	Maynas	Loreto	2006	N/A (PRJNA344889)
SAMN05853079	Sample_6	WGS	ILLUMINA	SRS1885926	PRJNA344889	SRR5136536	Santa Clara	San Juan Bautista	Maynas	Loreto	2007	N/A (PRJNA344889)
SAMN05853078	Sample_5	WGS	ILLUMINA	SRS1885941	PRJNA344889	SRR5136551	Padrecocha	Punchana	Maynas	Loreto	2008	N/A (PRJNA344889)
SAMN05853077	Sample_4	WGS	ILLUMINA	SRS1885931	PRJNA344889	SRR5136541	Padrecocha	Punchana	Maynas	Loreto	2007	N/A (PRJNA344889)
SAMN05853076	Sample_3	WGS	ILLUMINA	SRS1885938	PRJNA344889	SRR5136548	Padrecocha	Punchana	Maynas	Loreto	2007	N/A (PRJNA344889)

SAMN05853075	Sample_2	WGS	ILLUMINA	SRS1885928	PRJNA344889	SRR5136538	Padrecocha	Punchana	Maynas	Loreto	2007	N/A (PRJNA344889)
SAMN05853074	Sample_1	WGS	ILLUMINA	SRS1885929	PRJNA344889	SRR5136539	Padrecocha	Punchana	Maynas	Loreto	2007	N/A (PRJNA344889)
SAMN33235697	Pv07-19	WGS	ILLUMINA	SRS16725315	PRJNA853729	SRR23389636	Progreso	San Juan Bautista	Maynas	Loreto	2019	De Meulenaere et al., 2023; De Meulenaere et al., 2022
SAMN33235698	Pv12-19	WGS	ILLUMINA	SRS16725316	PRJNA853729	SRR23389635	Progreso	Belen	Maynas	Loreto	2019	De Meulenaere et al., 2023; De Meulenaere et al., 2022
SAMN33235699	Pv20-19	WGS	ILLUMINA	SRS16725317	PRJNA853729	SRR23389634	Santa Clara	San Juan Bautista	Maynas	Loreto	2019	De Meulenaere et al., 2023; De Meulenaere et al., 2022
SAMN33285180	Pv21-19	WGS	ILLUMINA	SRS16760795	PRJNA853729	SRR23455098	Zungarococha	San Juan Bautista	Maynas	Loreto	2019	De Meulenaere et al., 2023; De Meulenaere et al., 2022
SAMN29401255	Pv004	WGS	ILLUMINA	SRS13614552	PRJNA853729	SRR19884082	Moronacocha	Iquitos	Maynas	Loreto	2016	De Meulenaere et al., 2023; De Meulenaere et al., 2022
SAMN33235695	Pv005	WGS	ILLUMINA	SRS16725313	PRJNA853729	SRR23389638	Moronacocha	Iquitos	Maynas	Loreto	2016	De Meulenaere et al., 2023; De Meulenaere et al., 2022
SAMN29401257	Pv017	WGS	ILLUMINA	SRS13614563	PRJNA853729	SRR23390573	Progreso	Belen	Maynas	Loreto	2016	De Meulenaere et al., 2023; De Meulenaere et al., 2022
SAMN33235696	Pv019	WGS	ILLUMINA	SRS16725314	PRJNA853729	SRR23389637	Progreso	Belen	Maynas	Loreto	2016	De Meulenaere et al., 2023; De Meulenaere et al., 2022
SAMN29401258	Pv020	WGS	ILLUMINA	SRS13614567	PRJNA853729	SRR19884081	Progreso	Belen	Maynas	Loreto	2016	De Meulenaere et al., 2023; De Meulenaere et al., 2022

SAMN29401259	Pv021	WGS	ILLUMINA	SRS13614554	PRJNA853729	SRR19884080	Quistococha	San Juan Bautista	Maynas	Loreto	2016	De Meulenaere et al., 2023; De Meulenaere et al., 2022
SAMN29401260	Pv022	WGS	ILLUMINA	SRS13614553	PRJNA853729	SRR19884079	Quistococha	San Juan Bautista	Maynas	Loreto	2016	De Meulenaere et al., 2023; De Meulenaere et al., 2022
SAMN29401262	Pv024	WGS	ILLUMINA	SRS13614556	PRJNA853729	SRR19884078	Varillal	San Juan Bautista	Maynas	Loreto	2016	De Meulenaere et al., 2023; De Meulenaere et al., 2022
SAMN29401263	Pv025	WGS	ILLUMINA	SRS13614557	PRJNA853729	SRR19884077	Varillal	San Juan Bautista	Maynas	Loreto	2016	De Meulenaere et al., 2023; De Meulenaere et al., 2022
SAMN29401265	Pv032	WGS	ILLUMINA	SRS13614560	PRJNA853729	SRR19884076	Quistococha	San Juan Bautista	Maynas	Loreto	2016	De Meulenaere et al., 2023; De Meulenaere et al., 2022
SAMN29401266	Pv033	WGS	ILLUMINA	SRS13614561	PRJNA853729	SRR23390572	Quistococha	San Juan Bautista	Maynas	Loreto	2016	De Meulenaere et al., 2023; De Meulenaere et al., 2022
SAMN29401267	Pv041	WGS	ILLUMINA	SRS13614562	PRJNA853729	SRR19884075	Progreso	Belen	Maynas	Loreto	2016	De Meulenaere et al., 2023; De Meulenaere et al., 2022
SAMN29590264	Kattenberg_Peru_1	WGS	ILLUMINA	SRS13977783	PRJNA856808	SRR20306056		Iquitos	Maynas	Loreto	2008	Kattenberg et al., 2022
SAMN29590270	Kattenberg_Peru_2	WGS	ILLUMINA	SRS13977790	PRJNA856808	SRR20306049		Alto Nanay	Maynas	Loreto	2008	Kattenberg et al., 2022
SAMN29590271	Kattenberg_Peru_3	WGS	ILLUMINA	SRS13977791	PRJNA856808	SRR20306048	Barrio Florido	Punchana	Maynas	Loreto	2008	Kattenberg et al., 2022
SAMN29590272	Kattenberg_Peru_4	WGS	ILLUMINA	SRS13977792	PRJNA856808	SRR20306047		Mazan	Maynas	Loreto	2008	Kattenberg et al., 2022
SAMN29590273	Kattenberg_Peru_5	WGS	ILLUMINA	SRS13977793	PRJNA856808	SRR20306046	San Juan	San Juan Bautista	Maynas	Loreto	2008	Kattenberg et al., 2022
SAMN29590274	Kattenberg_Peru_6	WGS	ILLUMINA	SRS13977794	PRJNA856808	SRR20306017		Iquitos	Maynas	Loreto	2008	Kattenberg et al., 2022

SAMN29590275	Kattenberg_Peru_7	WGS	ILLUMINA	SRS13977824	PRJNA856808	SRR20306016	Curaray	Napo	Maynas	Loreto	2008	Kattenberg et al., 2022
SAMN29590276	Kattenberg_Peru_8	WGS	ILLUMINA	SRS13977823	PRJNA856808	SRR20306015	San Juan	San Juan Bautista	Maynas	Loreto	2008	Kattenberg et al., 2022
SAMN29590277	Kattenberg_Peru_9	WGS	ILLUMINA	SRS13977825	PRJNA856808	SRR20306014	San Juan	San Juan Bautista	Maynas	Loreto	2008	Kattenberg et al., 2022
SAMN29590265	Kattenberg_Peru_10	WGS	ILLUMINA	SRS13977784	PRJNA856808	SRR20306055	San Juan	San Juan Bautista	Maynas	Loreto	2008	Kattenberg et al., 2022
SAMN29590266	Kattenberg_Peru_11	WGS	ILLUMINA	SRS13977785	PRJNA856808	SRR20306054	Curaray	Napo	Maynas	Loreto	2008	Kattenberg et al., 2022
SAMN29590267	Kattenberg_Peru_12	WGS	ILLUMINA	SRS13977786	PRJNA856808	SRR20306053		Iquitos	Maynas	Loreto	2008	Kattenberg et al., 2022
SAMN29590268	Kattenberg_Peru_13	WGS	ILLUMINA	SRS13977787	PRJNA856808	SRR20306052		Iquitos	Maynas	Loreto	2008	Kattenberg et al., 2022
SAMN29590269	Kattenberg_Peru_14	WGS	ILLUMINA	SRS13977789	PRJNA856808	SRR20306050	Caballococha	Ramon Castilla	Ramon Castilla	Loreto	2008	Kattenberg et al., 2022
SAMEA104366051	VVX01421	WGS	ILLUMINA	ERS1991029	PRJEB2140	ERR2309761		Iquitos	Maynas	Loreto	2016	Adam et al., 2022
SAMEA104366053	VVX01439	WGS	ILLUMINA	ERS1991031	PRJEB2140	ERR2309762		Iquitos	Maynas	Loreto	2016	Adam et al., 2022
SAMEA104366050	VVX01445	WGS	ILLUMINA	ERS1991028	PRJEB2140	ERR2309760		Iquitos	Maynas	Loreto	2016	Adam et al., 2022
SAMEA104366032	VVX01456	WGS	ILLUMINA	ERS1991010	PRJEB2140	ERR2309903		Iquitos	Maynas	Loreto	2016	Adam et al., 2022
SAMEA3305207	PP0097-C	WGS	ILLUMINA	ERS683881	PRJEB2140	ERR979859		Iquitos	Maynas	Loreto	2014	Adam et al., 2022
SAMEA104366352	VVX01252	WGS	ILLUMINA	ERS1991330	PRJEB2140	ERR2309877		Iquitos	Maynas	Loreto	2015	Adam et al., 2022
SAMEA104366347	VVX01266	WGS	ILLUMINA	ERS1991325	PRJEB2140	ERR2309872		Iquitos	Maynas	Loreto	2015	Adam et al., 2022
SAMEA104366246	VVX01893	WGS	ILLUMINA	ERS1991224	PRJEB2140	ERR2309769		Iquitos	Maynas	Loreto	2015	Adam et al., 2022
SN	7040503	WGS	ILLUMINA	SN	SN	7040503	Libertad	Mazan	Maynas	Loreto	2022	Ramirez. R. (2024)
S/A	7040510	WGS	ILLUMINA	SN	SN	7040510	Libertad	Mazan	Maynas	Loreto	2022	Ramirez. R. (2024)
S/A	7040520	WGS	ILLUMINA	SN	SN	7040520	Libertad	Mazan	Maynas	Loreto	2022	Ramirez. R. (2024)
S/A	7040522	WGS	ILLUMINA	SN	SN	7040522	Libertad	Mazan	Maynas	Loreto	2022	Ramirez. R. (2024)

S/A	7040523	WGS	ILLUMINA	SN	SN	7040523	Libertad	Mazan	Maynas	Loreto	2022	Ramirez. R. (2024)
S/A	7040524	WGS	ILLUMINA	SN	SN	7040524	Libertad	Mazan	Maynas	Loreto	2022	Ramirez. R. (2024)
S/A	7040533	WGS	ILLUMINA	SN	SN	7040533	Libertad	Mazan	Maynas	Loreto	2022	Ramirez. R. (2024)
S/A	7040548	WGS	ILLUMINA	SN	SN	7040548	Libertad	Mazan	Maynas	Loreto	2022	Ramirez. R. (2024)
S/A	7040551	WGS	ILLUMINA	SN	SN	7040551	Libertad	Mazan	Maynas	Loreto	2022	Ramirez. R. (2024)
S/A	7040563	WGS	ILLUMINA	SN	SN	7040563	Libertad	Mazan	Maynas	Loreto	2022	Ramirez. R. (2024)
S/A	7050585	WGS	ILLUMINA	SN	SN	7050585	Urco Mirañó	Mazan	Maynas	Loreto	2022	Ramirez. R. (2024)
S/A	7050586	WGS	ILLUMINA	SN	SN	7050586	Urco Mirañó	Mazan	Maynas	Loreto	2022	Ramirez. R. (2024)
S/A	904021-16	WGS	ILLUMINA	SN	SN	904021-16	Libertad	Mazan	Maynas	Loreto	2022	Ramirez. R. (2024)
S/A	904022-16	WGS	ILLUMINA	SN	SN	904022-16	Libertad	Mazan	Maynas	Loreto	2022	Ramirez. R. (2024)
S/A	905047-18	WGS	ILLUMINA	SN	SN	905047-18	Urco Mirañó	Mazan	Maynas	Loreto	2022	Ramirez. R. (2024)
S/A	102032044	WGS	ILLUMINA	SN	SN	102032044	Libertad	Mazan	Maynas	Loreto	2015	Ramirez. R. (2024)
S/A	102067075	WGS	ILLUMINA	SN	SN	102067075	Libertad	Mazan	Maynas	Loreto	2015	Ramirez. R. (2024)
S/A	110022035	WGS	ILLUMINA	SN	SN	110022035	Urco Mirañó	Mazan	Maynas	Loreto	2015	Ramirez. R. (2024)
S/A	110022065	WGS	ILLUMINA	SN	SN	110022065	Urco Mirañó	Mazan	Maynas	Loreto	2015	Ramirez. R. (2024)
S/A	110047144	WGS	ILLUMINA	SN	SN	110047144	Urco Mirañó	Mazan	Maynas	Loreto	2015	Ramirez. R. (2024)
S/A	110053044	WGS	ILLUMINA	SN	SN	110053044	Urco Mirañó	Mazan	Maynas	Loreto	2015	Ramirez. R. (2024)
S/A	231014058	WGS	ILLUMINA	SN	SN	231014058	Libertad	Mazan	Maynas	Loreto	2017	Ramirez. R. (2024)
S/A	CON21-065	AmpliSeq	ILLUMINA	SN	SN	ot1_S43	Candungos	Río Santiago	Condorcanqui	Amazonas	2021	Este estudio
S/A	CON21-066	AmpliSeq	ILLUMINA	SN	SN	ot2_S54	Candungos	Río Santiago	Condorcanqui	Amazonas	2021	Este estudio

S/A	CON21-067	AmpliSeq	ILLUMINA	SN	SN	ot3_S65	Candungos	Río Santiago	Condorcanqui	Amazonas	2021	Este estudio
S/A	CON22-248	AmpliSeq	ILLUMINA	SN	SN	ot4_S77	Candungos	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-312	AmpliSeq	ILLUMINA	SN	SN	ot5_S89	Cucuasa	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-320	AmpliSeq	ILLUMINA	SN	SN	ot6_S101	Cucuasa	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-335	AmpliSeq	ILLUMINA	SN	SN	ot7_S113	Soledad	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-336	AmpliSeq	ILLUMINA	SN	SN	ot8_S125	Soledad	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-338	AmpliSeq	ILLUMINA	SN	SN	ot9_S44	Ayambis	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-343	AmpliSeq	ILLUMINA	SN	SN	ot10_S55	Palometa	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-344	AmpliSeq	ILLUMINA	SN	SN	ot11_S66	Cucuasa	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-345	AmpliSeq	ILLUMINA	SN	SN	ot12_S78	Ayambis	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-346	AmpliSeq	ILLUMINA	SN	SN	ot13_S90	Cucuasa	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-347	AmpliSeq	ILLUMINA	SN	SN	ot14_S102	Palometa	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-364	AmpliSeq	ILLUMINA	SN	SN	ot15_S114	Ayambis	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-366	AmpliSeq	ILLUMINA	SN	SN	ot16_S126	Nauta	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-367	AmpliSeq	ILLUMINA	SN	SN	ot17_S45	Nauta	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-376	AmpliSeq	ILLUMINA	SN	SN	ot18_S56	Ayambis	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-377	AmpliSeq	ILLUMINA	SN	SN	ot19_S67	Ajachim	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-379	AmpliSeq	ILLUMINA	SN	SN	ot20_S79	Nauta	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-446	AmpliSeq	ILLUMINA	SN	SN	ot21_S91	Caterpiza	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-451	AmpliSeq	ILLUMINA	SN	SN	ot22_S103	Caterpiza	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio

S/A	CON22-459	AmpliSeq	ILLUMINA	SN	SN	ot23_S115	Caterpiza	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-462	AmpliSeq	ILLUMINA	SN	SN	ot24_S127	Caterpiza	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-463	AmpliSeq	ILLUMINA	SN	SN	ot25_S46	Cucuasa	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-465	AmpliSeq	ILLUMINA	SN	SN	ot26_S57	Cucuasa	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-466	AmpliSeq	ILLUMINA	SN	SN	ot27_S68	Cucuasa	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CT035-22	AmpliSeq	ILLUMINA	SN	SN	ot28_S80	Caterpiza	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	GY23-066	AmpliSeq	ILLUMINA	SN	SN	ot29_S92	Guayabal	Río Santiago	Condorcanqui	Amazonas	2023	Este estudio
S/A	GY23-079	AmpliSeq	ILLUMINA	SN	SN	ot30_S104	Guayabal	Río Santiago	Condorcanqui	Amazonas	2023	Este estudio
S/A	GY23-113	AmpliSeq	ILLUMINA	SN	SN	ot31_S116	Guayabal	Río Santiago	Condorcanqui	Amazonas	2023	Este estudio
S/A	CON23-062	AmpliSeq	ILLUMINA	SN	SN	ot32_S128	Kayamas	Río Santiago	Condorcanqui	Amazonas	2023	Este estudio
S/A	CON23-065	AmpliSeq	ILLUMINA	SN	SN	ot33_S47	Belén	Río Santiago	Condorcanqui	Amazonas	2023	Este estudio
S/A	CON23-071	AmpliSeq	ILLUMINA	SN	SN	ot34_S58	Kayamas	Río Santiago	Condorcanqui	Amazonas	2023	Este estudio
S/A	CON23-097	AmpliSeq	ILLUMINA	SN	SN	ot35_S69	Alto Kashap	Nieva	Condorcanqui	Amazonas	2023	Este estudio
S/A	CON23-099	AmpliSeq	ILLUMINA	SN	SN	ot36_S81	Alto Kashap	Nieva	Condorcanqui	Amazonas	2023	Este estudio
S/A	CON23-100	AmpliSeq	ILLUMINA	SN	SN	ot37_S93	Alto Kashap	Nieva	Condorcanqui	Amazonas	2023	Este estudio
S/A	CON23-102	AmpliSeq	ILLUMINA	SN	SN	ot38_S105	Alto Kashap	Nieva	Condorcanqui	Amazonas	2023	Este estudio
S/A	CON23-107	AmpliSeq	ILLUMINA	SN	SN	ot39_S117	Alto Kashap	Nieva	Condorcanqui	Amazonas	2023	Este estudio
S/A	CON23-112	AmpliSeq	ILLUMINA	SN	SN	ot40_S129	Alto Kashap	Nieva	Condorcanqui	Amazonas	2023	Este estudio
S/A	CON23-114	AmpliSeq	ILLUMINA	SN	SN	ot41_S48	Alto Kashap	Nieva	Condorcanqui	Amazonas	2023	Este estudio
S/A	CON23-126	AmpliSeq	ILLUMINA	SN	SN	ot42_S59	Alto Kashap	Nieva	Condorcanqui	Amazonas	2023	Este estudio
S/A	CON23-128	AmpliSeq	ILLUMINA	SN	SN	ot43_S70	Alto Kashap	Nieva	Condorcanqui	Amazonas	2023	Este estudio
S/A	CON23-130	AmpliSeq	ILLUMINA	SN	SN	ot44_S82	Alto Kashap	Nieva	Condorcanqui	Amazonas	2023	Este estudio

S/A	CON23-133	AmpliSeq	ILLUMINA	SN	SN	ot45_S94	Alto Kashap	Nieva	Condorcanqui	Amazonas	2023	Este estudio
S/A	CON23-148	AmpliSeq	ILLUMINA	SN	SN	ot46_S106	Guayabal	Río Santiago	Condorcanqui	Amazonas	2023	Este estudio
S/A	CON23-171	AmpliSeq	ILLUMINA	SN	SN	ot47_S118	Alto Kashap	Nieva	Condorcanqui	Amazonas	2023	Este estudio
S/A	CON23-172	AmpliSeq	ILLUMINA	SN	SN	ot48_S130	Alto Kashap	Nieva	Condorcanqui	Amazonas	2023	Este estudio
S/A	CON23-173	AmpliSeq	ILLUMINA	SN	SN	ot49_S49	Alto Kashap	Nieva	Condorcanqui	Amazonas	2023	Este estudio
S/A	CON23-177	AmpliSeq	ILLUMINA	SN	SN	ot50_S60	Alto Kashap	Nieva	Condorcanqui	Amazonas	2023	Este estudio
S/A	CON22-05	AmpliSeq	ILLUMINA	SN	SN	ch1_S71	Alianza Progreso	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-06	AmpliSeq	ILLUMINA	SN	SN	ch2_S83	Alianza Progreso	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	NE-860	AmpliSeq	ILLUMINA	SN	SN	ch4_S95	Nueva Esperanza	Río Santiago	Condorcanqui	Amazonas	2020	Este estudio
S/A	Ps.CH-019	AmpliSeq	ILLUMINA	SN	SN	ch5_S107	Alianza Progreso	Río Santiago	Condorcanqui	Amazonas	2019	Este estudio
S/A	Ps.CH-003	AmpliSeq	ILLUMINA	SN	SN	ch6_S119	Alianza Progreso	Río Santiago	Condorcanqui	Amazonas	2019	Este estudio
S/A	CON21-036	AmpliSeq	ILLUMINA	SN	SN	ch9_S131	Alianza Progreso	Río Santiago	Condorcanqui	Amazonas	2021	Este estudio
S/A	NE-020-04	AmpliSeq	ILLUMINA	SN	SN	ch10_S50	Nueva Esperanza	Río Santiago	Condorcanqui	Amazonas	2021	Este estudio
S/A	CON22-83	AmpliSeq	ILLUMINA	SN	SN	ch12_S61	Nueva Esperanza	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-228	AmpliSeq	ILLUMINA	SN	SN	ch13_S72	Alianza Progreso	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-285	AmpliSeq	ILLUMINA	SN	SN	ch15_S84	Alianza Progreso	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-324	AmpliSeq	ILLUMINA	SN	SN	ch17_S96	Alianza Progreso	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-383	AmpliSeq	ILLUMINA	SN	SN	ch19_S108	Alianza Progreso	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-430	AmpliSeq	ILLUMINA	SN	SN	ch20_S120	Nueva Esperanza	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON22-433	AmpliSeq	ILLUMINA	SN	SN	ch21_S132	Nueva Esperanza	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	AP-030-22	AmpliSeq	ILLUMINA	SN	SN	ch25_S51	Alianza Progreso	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio

S/A	AP-040-22	AmpliSeq	ILLUMINA	SN	SN	ch26_S62	Alianza Progreso	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	AP-053-22	AmpliSeq	ILLUMINA	SN	SN	ch27_S73	Alianza Progreso	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	NE-012-22	AmpliSeq	ILLUMINA	SN	SN	ch31_S85	Nueva Esperanza	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	NE-082-22	AmpliSeq	ILLUMINA	SN	SN	ch33_S97	Nueva Esperanza	Río Santiago	Condorcanqui	Amazonas	2022	Este estudio
S/A	CON23-043	AmpliSeq	ILLUMINA	SN	SN	ch37_S109	Nueva Esperanza	Río Santiago	Condorcanqui	Amazonas	2023	Este estudio
S/A	CON23-051	AmpliSeq	ILLUMINA	SN	SN	ch39_S121	Alianza Progreso	Río Santiago	Condorcanqui	Amazonas	2023	Este estudio
S/A	CON23-060	AmpliSeq	ILLUMINA	SN	SN	ch41_S133	Nueva Esperanza	Río Santiago	Condorcanqui	Amazonas	2023	Este estudio
S/A	CON23-061	AmpliSeq	ILLUMINA	SN	SN	ch42_S52	Alianza Progreso	Río Santiago	Condorcanqui	Amazonas	2023	Este estudio
SAMN38982299	PE-PV-3	AmpliSeq	ILLUMINA	SRS19948955	PRJNA1055117	SRR27306060	San Jose de Loreto Yacu	Ramon Castilla	Ramon Castilla	Loreto	2018	Kattenberg et al., 2023
SAMN38982300	PE-PV-4	AmpliSeq	ILLUMINA	SRS19948956	PRJNA1055117	SRR27306059	San Jose de Loreto Yacu	Ramon Castilla	Ramon Castilla	Loreto	2018	Kattenberg et al., 2023
SAMN38982301	PE-PV-8	AmpliSeq	ILLUMINA	SRS19949105	PRJNA1055117	SRR27305910	Islandia	Yavarí	Ramon Castilla	Loreto	2018	Kattenberg et al., 2023
SAMN38982302	PE-PV-10	AmpliSeq	ILLUMINA	SRS19949146	PRJNA1055117	SRR27305869	Islandia	Yavarí	Ramon Castilla	Loreto	2018	Kattenberg et al., 2023
SAMN38982303	PE-PV-11	AmpliSeq	ILLUMINA	SRS19949157	PRJNA1055117	SRR27305858	Islandia	Yavarí	Ramon Castilla	Loreto	2018	Kattenberg et al., 2023
SAMN38982304	PE-PV-17	AmpliSeq	ILLUMINA	SRS19949168	PRJNA1055117	SRR27305847	Islandia	Yavarí	Ramon Castilla	Loreto	2018	Kattenberg et al., 2023
SAMN38982305	PE-PV-21	AmpliSeq	ILLUMINA	SRS19949179	PRJNA1055117	SRR27305836	Islandia	Yavarí	Ramon Castilla	Loreto	2018	Kattenberg et al., 2023
SAMN38982306	PE-PV-22	AmpliSeq	ILLUMINA	SRS19948992	PRJNA1055117	SRR27306023	Islandia	Yavarí	Ramon Castilla	Loreto	2018	Kattenberg et al., 2023
SAMN38982307	PE-PV-23	AmpliSeq	ILLUMINA	SRS19949003	PRJNA1055117	SRR27306012	Islandia	Yavarí	Ramon Castilla	Loreto	2018	Kattenberg et al., 2023
SAMN38982308	PE-PV-24	AmpliSeq	ILLUMINA	SRS19949014	PRJNA1055117	SRR27306001	Islandia	Yavarí	Ramon Castilla	Loreto	2018	Kattenberg et al., 2023
SAMN38982309	PE-PV-25	AmpliSeq	ILLUMINA	SRS19948957	PRJNA1055117	SRR27306058	Islandia	Yavarí	Ramon Castilla	Loreto	2018	Kattenberg et al., 2023
SAMN38982310	PE-PV-26	AmpliSeq	ILLUMINA	SRS19948968	PRJNA1055117	SRR27306047	Islandia	Yavarí	Ramon Castilla	Loreto	2018	Kattenberg et al., 2023

SAMN38982311	PE-PV-28	AmpliSeq	ILLUMINA	SRS19948979	PRJNA1055117	SRR27306036	Islandia	Yavarí	Ramon Castilla	Loreto	2018	Kattenberg et al., 2023
SAMN38982312	PE-PV-32	AmpliSeq	ILLUMINA	SRS19949028	PRJNA1055117	SRR27305987	Islandia	Yavarí	Ramon Castilla	Loreto	2018	Kattenberg et al., 2023
SAMN38982313	PE-PV-33	AmpliSeq	ILLUMINA	SRS19949039	PRJNA1055117	SRR27305976	Islandia	Yavarí	Ramon Castilla	Loreto	2018	Kattenberg et al., 2023
SAMN38982314	PE-PV-34	AmpliSeq	ILLUMINA	SRS19949050	PRJNA1055117	SRR27305965	Islandia	Yavarí	Ramon Castilla	Loreto	2018	Kattenberg et al., 2023
SAMN38982315	PE-PV-35	AmpliSeq	ILLUMINA	SRS19949061	PRJNA1055117	SRR27305954	Islandia	Yavarí	Ramon Castilla	Loreto	2018	Kattenberg et al., 2023
SAMN38982316	PE-PV-37	AmpliSeq	ILLUMINA	SRS19949072	PRJNA1055117	SRR27305943	Islandia	Yavarí	Ramon Castilla	Loreto	2018	Kattenberg et al., 2023
SAMN38982317	PE-PV-38	AmpliSeq	ILLUMINA	SRS19949083	PRJNA1055117	SRR27305932	Islandia	Yavarí	Ramon Castilla	Loreto	2018	Kattenberg et al., 2023
SAMN38982318	PE-PV-39	AmpliSeq	ILLUMINA	SRS19949094	PRJNA1055117	SRR27305921	Islandia	Yavarí	Ramon Castilla	Loreto	2018	Kattenberg et al., 2023
SAMN38982319	PE-PV-40	AmpliSeq	ILLUMINA	SRS19949106	PRJNA1055117	SRR27305909	Islandia	Yavarí	Ramon Castilla	Loreto	2018	Kattenberg et al., 2023
SAMN38982320	PE-PV-42	AmpliSeq	ILLUMINA	SRS19949117	PRJNA1055117	SRR27305898	Islandia	Yavarí	Ramon Castilla	Loreto	2018	Kattenberg et al., 2023
SAMN38982321	PE-PV-45	AmpliSeq	ILLUMINA	SRS19949128	PRJNA1055117	SRR27305887	San Juan	San Juan Bautista	Maynas	Loreto	2007	Kattenberg et al., 2023
SAMN38982322	PE-PV-46	AmpliSeq	ILLUMINA	SRS19949139	PRJNA1055117	SRR27305876	San Juan	San Juan Bautista	Maynas	Loreto	2007	Kattenberg et al., 2023
SAMN38982323	PE-PV-47	AmpliSeq	ILLUMINA	SRS19949140	PRJNA1055117	SRR27305875	San Juan	San Juan Bautista	Maynas	Loreto	2007	Kattenberg et al., 2023
SAMN38982324	PE-PV-48	AmpliSeq	ILLUMINA	SRS19949141	PRJNA1055117	SRR27305874	San Juan	San Juan Bautista	Maynas	Loreto	2007	Kattenberg et al., 2023
SAMN38982325	PE-PV-49	AmpliSeq	ILLUMINA	SRS19949142	PRJNA1055117	SRR27305873	San Juan	San Juan Bautista	Maynas	Loreto	2007	Kattenberg et al., 2023
SAMN38982326	PE-PV-50	AmpliSeq	ILLUMINA	SRS19949143	PRJNA1055117	SRR27305872	San Juan	San Juan Bautista	Maynas	Loreto	2007	Kattenberg et al., 2023
SAMN38982327	PE-PV-51	AmpliSeq	ILLUMINA	SRS19949144	PRJNA1055117	SRR27305871	San Juan	San Juan Bautista	Maynas	Loreto	2007	Kattenberg et al., 2023
SAMN38982328	PE-PV-53	AmpliSeq	ILLUMINA	SRS19949145	PRJNA1055117	SRR27305870	San Juan	San Juan Bautista	Maynas	Loreto	2007	Kattenberg et al., 2023
SAMN38982329	PE-PV-55	AmpliSeq	ILLUMINA	SRS19949147	PRJNA1055117	SRR27305868	Iquitos	Iquitos	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982330	PE-PV-56	AmpliSeq	ILLUMINA	SRS19949148	PRJNA1055117	SRR27305867	San Juan	San Juan Bautista	Maynas	Loreto	2008	Kattenberg et al., 2023

SAMN38982331	PE-PV-57	AmpliSeq	ILLUMINA	SRS19949149	PRJNA1055117	SRR27305866	Bellaluz	San Juan Bautista	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982332	PE-PV-59	AmpliSeq	ILLUMINA	SRS19949150	PRJNA1055117	SRR27305865	El Milagro	San Juan Bautista	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982333	PE-PV-63	AmpliSeq	ILLUMINA	SRS19949151	PRJNA1055117	SRR27305864	Belen	Belen	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982334	PE-PV-65	AmpliSeq	ILLUMINA	SRS19949152	PRJNA1055117	SRR27305863	Alto Nanay	Alto Nanay	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982335	PE-PV-66	AmpliSeq	ILLUMINA	SRS19949154	PRJNA1055117	SRR27305862	San Juan	San Juan Bautista	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982336	PE-PV-67	AmpliSeq	ILLUMINA	SRS19949153	PRJNA1055117	SRR27305861	San Juan	San Juan Bautista	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982337	PE-PV-68	AmpliSeq	ILLUMINA	SRS19949155	PRJNA1055117	SRR27305860	Iquitos	Iquitos	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982338	PE-PV-69	AmpliSeq	ILLUMINA	SRS19949156	PRJNA1055117	SRR27305859	Iquitos	Iquitos	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982339	PE-PV-70	AmpliSeq	ILLUMINA	SRS19949158	PRJNA1055117	SRR27305857	Iquitos	Iquitos	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982340	PE-PV-71	AmpliSeq	ILLUMINA	SRS19949159	PRJNA1055117	SRR27305856	Punchana	Punchana	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982341	PE-PV-72	AmpliSeq	ILLUMINA	SRS19949160	PRJNA1055117	SRR27305855	San Juan	San Juan Bautista	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982342	PE-PV-73	AmpliSeq	ILLUMINA	SRS19949161	PRJNA1055117	SRR27305854	Barrio Florida	Punchana	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982343	PE-PV-74	AmpliSeq	ILLUMINA	SRS19949162	PRJNA1055117	SRR27305853	Mazan	Mazan	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982344	PE-PV-75	AmpliSeq	ILLUMINA	SRS19949163	PRJNA1055117	SRR27305852	San Juan	San Juan Bautista	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982345	PE-PV-76	AmpliSeq	ILLUMINA	SRS19949165	PRJNA1055117	SRR27305851	San Juan	San Juan Bautista	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982346	PE-PV-78	AmpliSeq	ILLUMINA	SRS19949164	PRJNA1055117	SRR27305850	Masusa	Punchana	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982347	PE-PV-79	AmpliSeq	ILLUMINA	SRS19949166	PRJNA1055117	SRR27305849	San Juan	San Juan Bautista	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982348	PE-PV-80	AmpliSeq	ILLUMINA	SRS19949167	PRJNA1055117	SRR27305848	Iquitos	Iquitos	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982349	PE-PV-81	AmpliSeq	ILLUMINA	SRS19949169	PRJNA1055117	SRR27305846	Curaray	Napo	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982350	PE-PV-82	AmpliSeq	ILLUMINA	SRS19949170	PRJNA1055117	SRR27305845	Iquitos	Iquitos	Maynas	Loreto	2008	Kattenberg et al., 2023

SAMN38982351	PE-PV-83	AmpliSeq	ILLUMINA	SRS19949171	PRJNA1055117	SRR27305844	San Juan	San Juan Bautista	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982352	PE-PV-84	AmpliSeq	ILLUMINA	SRS19949172	PRJNA1055117	SRR27305843	San Juan	San Juan Bautista	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982353	PE-PV-85	AmpliSeq	ILLUMINA	SRS19949173	PRJNA1055117	SRR27305842	San Juan	San Juan Bautista	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982354	PE-PV-86	AmpliSeq	ILLUMINA	SRS19949039	PRJNA1055117	SRR27305841	San Juan	San Juan Bautista	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982355	PE-PV-87	AmpliSeq	ILLUMINA	SRS19949175	PRJNA1055117	SRR27305840	San Juan	San Juan Bautista	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982356	PE-PV-88	AmpliSeq	ILLUMINA	SRS19949176	PRJNA1055117	SRR27305839	Iquitos	Iquitos	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982357	PE-PV-89	AmpliSeq	ILLUMINA	SRS19949177	PRJNA1055117	SRR27305838	San Juan	San Juan Bautista	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982358	PE-PV-90	AmpliSeq	ILLUMINA	SRS19949178	PRJNA1055117	SRR27305837	Iquitos	Iquitos	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982359	PE-PV-91	AmpliSeq	ILLUMINA	SRS19949181	PRJNA1055117	SRR27305835	San Juan	San Juan Bautista	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982360	PE-PV-92	AmpliSeq	ILLUMINA	SRS19949180	PRJNA1055117	SRR27305834	Iquitos	Iquitos	Maynas	Loreto	2008	Kattenberg et al., 2023
SAMN38982361	PE-PV-93	AmpliSeq	ILLUMINA	SRS19949182	PRJNA1055117	SRR27305833	Caballococha	Ramon Castilla	Ramon Castilla	Loreto	2008	Kattenberg et al., 2023
SAMN38982362	PE-PV-95	AmpliSeq	ILLUMINA	SRS19949183	PRJNA1055117	SRR27305832	Simon Bolivar	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982363	PE-PV-96	AmpliSeq	ILLUMINA	SRS19949184	PRJNA1055117	SRR27305831	America	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982364	PE-PV-97	AmpliSeq	ILLUMINA	SRS19948987	PRJNA1055117	SRR27306028	Shiriara	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982365	PE-PV-98	AmpliSeq	ILLUMINA	SRS19948988	PRJNA1055117	SRR27306027	Camelias	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982366	PE-PV-99	AmpliSeq	ILLUMINA	SRS19948989	PRJNA1055117	SRR27306026	Santa Clara	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982367	PE-PV-100	AmpliSeq	ILLUMINA	SRS19948990	PRJNA1055117	SRR27306025	Santa Clara	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982368	PE-PV-101	AmpliSeq	ILLUMINA	SRS19948991	PRJNA1055117	SRR27306024	San Juan	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982369	PE-PV-102	AmpliSeq	ILLUMINA	SRS19948993	PRJNA1055117	SRR27306022	Bella Luz	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982370	PE-PV-104	AmpliSeq	ILLUMINA	SRS19948994	PRJNA1055117	SRR27306021	Santa Clara	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023

SAMN38982371	PE-PV-105	AmpliSeq	ILLUMINA	SRS19948995	PRJNA1055117	SRR27306020	Santa Clara	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982372	PE-PV-106	AmpliSeq	ILLUMINA	SRS19948996	PRJNA1055117	SRR27306019	Santa Clara	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982373	PE-PV-107	AmpliSeq	ILLUMINA	SRS19948997	PRJNA1055117	SRR27306018	Azucenas	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982374	PE-PV-108	AmpliSeq	ILLUMINA	SRS19948998	PRJNA1055117	SRR27306017	Santa Clara	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982375	PE-PV-109	AmpliSeq	ILLUMINA	SRS19948999	PRJNA1055117	SRR27306016	Rosale	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982376	PE-PV-110	AmpliSeq	ILLUMINA	SRS19949000	PRJNA1055117	SRR27306015	Camelias	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982377	PE-PV-111	AmpliSeq	ILLUMINA	SRS19949001	PRJNA1055117	SRR27306014	Rumococha	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982378	PE-PV-112	AmpliSeq	ILLUMINA	SRS19949002	PRJNA1055117	SRR27306013	Camelias	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982379	PE-PV-113	AmpliSeq	ILLUMINA	SRS19949004	PRJNA1055117	SRR27306011	Manco Inca	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982380	PE-PV-114	AmpliSeq	ILLUMINA	SRS19949005	PRJNA1055117	SRR27306010	Quistococha	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982381	PE-PV-115	AmpliSeq	ILLUMINA	SRS19949006	PRJNA1055117	SRR27306009	Azucenas	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982382	PE-PV-116	AmpliSeq	ILLUMINA	SRS19949007	PRJNA1055117	SRR27306008	Azucenas	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982383	PE-PV-117	AmpliSeq	ILLUMINA	SRS19949008	PRJNA1055117	SRR27306007	25 de Enero	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982384	PE-PV-118	AmpliSeq	ILLUMINA	SRS19949009	PRJNA1055117	SRR27306006	Quistococha	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982385	PE-PV-119	AmpliSeq	ILLUMINA	SRS19949010	PRJNA1055117	SRR27306005	Camelias	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982386	PE-PV-120	AmpliSeq	ILLUMINA	SRS19949011	PRJNA1055117	SRR27306004	Monasi	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982387	PE-PV-121	AmpliSeq	ILLUMINA	SRS19949012	PRJNA1055117	SRR27306003	Camelias	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982388	PE-PV-122	AmpliSeq	ILLUMINA	SRS19949013	PRJNA1055117	SRR27306002	Simon Bolivar	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982389	PE-PV-123	AmpliSeq	ILLUMINA	SRS19949015	PRJNA1055117	SRR27306000	Laguna Azul	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982390	PE-PV-125	AmpliSeq	ILLUMINA	SRS19949016	PRJNA1055117	SRR27305999	Quistococha	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023

SAMN38982391	PE-PV-126	AmpliSeq	ILLUMINA	SRS19949017	PRJNA1055117	SRR27305998	Quistococha	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982392	PE-PV-127	AmpliSeq	ILLUMINA	SRS19949018	PRJNA1055117	SRR27305997	San Juan	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982393	PE-PV-128	AmpliSeq	ILLUMINA	SRS19949019	PRJNA1055117	SRR27305996	Santa Clara	San Juan Bautista	Maynas	Loreto	2014	Kattenberg et al., 2023
SAMN38982394	PE-PV-134	AmpliSeq	ILLUMINA	SRS19949020	PRJNA1055117	SRR27305995	Huaman Urco	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982395	PE-PV-141	AmpliSeq	ILLUMINA	SRS19949021	PRJNA1055117	SRR27305994	Urco Miraña	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982396	PE-PV-142	AmpliSeq	ILLUMINA	SRS19949022	PRJNA1055117	SRR27305993	Urco Miraña	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982397	PE-PV-143	AmpliSeq	ILLUMINA	SRS19949023	PRJNA1055117	SRR27305992	Urco Miraña	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982398	PE-PV-152	AmpliSeq	ILLUMINA	SRS19949024	PRJNA1055117	SRR27305991	Visto Bueno	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982399	PE-PV-153	AmpliSeq	ILLUMINA	SRS19948958	PRJNA1055117	SRR27306057	Visto Bueno	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982400	PE-PV-163	AmpliSeq	ILLUMINA	SRS19948959	PRJNA1055117	SRR27306056	Santa Cruz	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982401	PE-PV-164	AmpliSeq	ILLUMINA	SRS19948960	PRJNA1055117	SRR27306055	Santa Cruz	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982402	PE-PV-174	AmpliSeq	ILLUMINA	SRS19948961	PRJNA1055117	SRR27306054	Libertad	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982403	PE-PV-176	AmpliSeq	ILLUMINA	SRS19948962	PRJNA1055117	SRR27306053	Libertad	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982404	PE-PV-180	AmpliSeq	ILLUMINA	SRS19948963	PRJNA1055117	SRR27306052	Libertad	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982405	PE-PV-183	AmpliSeq	ILLUMINA	SRS19948964	PRJNA1055117	SRR27306051	Libertad	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982406	PE-PV-184	AmpliSeq	ILLUMINA	SRS19948965	PRJNA1055117	SRR27306050	Puerto Alegre	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982407	PE-PV-185	AmpliSeq	ILLUMINA	SRS19948966	PRJNA1055117	SRR27306049	Puerto Alegre	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982408	PE-PV-187	AmpliSeq	ILLUMINA	SRS19948967	PRJNA1055117	SRR27306048	Puerto Alegre	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982409	PE-PV-188	AmpliSeq	ILLUMINA	SRS19948969	PRJNA1055117	SRR27306046	Puerto Alegre	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982410	PE-PV-192	AmpliSeq	ILLUMINA	SRS19948970	PRJNA1055117	SRR27306045	Sucusari	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023

SAMN38982411	PE-PV-194	AmpliSeq	ILLUMINA	SRS19948971	PRJNA1055117	SRR27306044	Sucusari	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982412	PE-PV-196	AmpliSeq	ILLUMINA	SRS19948972	PRJNA1055117	SRR27306043	Huaman Urco	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982413	PE-PV-215	AmpliSeq	ILLUMINA	SRS19948973	PRJNA1055117	SRR27306042	Urco Miraña	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982414	PE-PV-220	AmpliSeq	ILLUMINA	SRS19948974	PRJNA1055117	SRR27306041	Urco Miraña	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982415	PE-PV-248	AmpliSeq	ILLUMINA	SRS19948975	PRJNA1055117	SRR27306040	Urco Miraña	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982416	PE-PV-315	AmpliSeq	ILLUMINA	SRS19948976	PRJNA1055117	SRR27306039	Puerto Obrero	Mazan	Maynas	Loreto	2017	Kattenberg et al., 2023
SAMN38982417	PE-PV-322	AmpliSeq	ILLUMINA	SRS19948977	PRJNA1055117	SRR27306038	Libertad	Mazan	Maynas	Loreto	2017	Kattenberg et al., 2023
SAMN38982418	PE-PV-325	AmpliSeq	ILLUMINA	SRS19948978	PRJNA1055117	SRR27306037	Libertad	Mazan	Maynas	Loreto	2017	Kattenberg et al., 2023
SAMN38982419	PE-PV-327	AmpliSeq	ILLUMINA	SRS19948980	PRJNA1055117	SRR27306035	Libertad	Mazan	Maynas	Loreto	2017	Kattenberg et al., 2023
SAMN38982420	PE-PV-344	AmpliSeq	ILLUMINA	SRS19948981	PRJNA1055117	SRR27306034	Primero de Enero	Mazan	Maynas	Loreto	2017	Kattenberg et al., 2023
SAMN38982421	PE-PV-348	AmpliSeq	ILLUMINA	SRS19948982	PRJNA1055117	SRR27306033	Santa Cruz	Mazan	Maynas	Loreto	2017	Kattenberg et al., 2023
SAMN38982422	PE-PV-349	AmpliSeq	ILLUMINA	SRS19948983	PRJNA1055117	SRR27306032	Santa Cruz	Mazan	Maynas	Loreto	2017	Kattenberg et al., 2023
SAMN38982423	PE-PV-352	AmpliSeq	ILLUMINA	SRS19948984	PRJNA1055117	SRR27306031	Santa Cruz	Mazan	Maynas	Loreto	2017	Kattenberg et al., 2023
SAMN38982424	PE-PV-353	AmpliSeq	ILLUMINA	SRS19948985	PRJNA1055117	SRR27306030	Bello Horizonte	Mazan	Maynas	Loreto	2017	Kattenberg et al., 2023
SAMN38982425	PE-PV-354	AmpliSeq	ILLUMINA	SRS19948986	PRJNA1055117	SRR27306029	Santa Cruz	Mazan	Maynas	Loreto	2017	Kattenberg et al., 2023
SAMN38982426	PE-PV-355	AmpliSeq	ILLUMINA	SRS19949025	PRJNA1055117	SRR27305990	Santa Cruz	Mazan	Maynas	Loreto	2017	Kattenberg et al., 2023
SAMN38982427	PE-PV-360	AmpliSeq	ILLUMINA	SRS19949026	PRJNA1055117	SRR27305989	Bello Horizonte	Mazan	Maynas	Loreto	2017	Kattenberg et al., 2023
SAMN38982428	PE-PV-363	AmpliSeq	ILLUMINA	SRS19949027	PRJNA1055117	SRR27305988	Santa Cruz	Mazan	Maynas	Loreto	2017	Kattenberg et al., 2023
SAMN38982429	PE-PV-364	AmpliSeq	ILLUMINA	SRS19949029	PRJNA1055117	SRR27305986	Libertad	Mazan	Maynas	Loreto	2017	Kattenberg et al., 2023
SAMN38982430	PE-PV-365	AmpliSeq	ILLUMINA	SRS19949030	PRJNA1055117	SRR27305985	Libertad	Mazan	Maynas	Loreto	2017	Kattenberg et al., 2023

SAMN38982431	PE-PV-367	AmpliSeq	ILLUMINA	SRS19949031	PRJNA1055117	SRR27305984	Libertad	Mazan	Maynas	Loreto	2017	Kattenberg et al., 2023
SAMN38982432	PE-PV-368	AmpliSeq	ILLUMINA	SRS19949032	PRJNA1055117	SRR27305983	Libertad	Mazan	Maynas	Loreto	2017	Kattenberg et al., 2023
SAMN38982433	PE-PV-378	AmpliSeq	ILLUMINA	SRS19949033	PRJNA1055117	SRR27305982	Puerto Obrero	Mazan	Maynas	Loreto	2017	Kattenberg et al., 2023
SAMN38982434	PE-PV-383	AmpliSeq	ILLUMINA	SRS19949034	PRJNA1055117	SRR27305981	Huaman Urco	Mazan	Maynas	Loreto	2017	Kattenberg et al., 2023
SAMN38982435	PE-PV-385	AmpliSeq	ILLUMINA	SRS19949035	PRJNA1055117	SRR27305980	Primero de Enero	Mazan	Maynas	Loreto	2017	Kattenberg et al., 2023
SAMN38982436	PE-PV-387	AmpliSeq	ILLUMINA	SRS19949036	PRJNA1055117	SRR27305979	Primero de Enero	Mazan	Maynas	Loreto	2017	Kattenberg et al., 2023
SAMN38982437	PE-PV-388	AmpliSeq	ILLUMINA	SRS19949037	PRJNA1055117	SRR27305978	San Antonio de Zambrano	Mazan	Maynas	Loreto	2017	Kattenberg et al., 2023
SAMN38982438	PE-PV-390	AmpliSeq	ILLUMINA	SRS19949038	PRJNA1055117	SRR27305977	Urco Miraña	Mazan	Maynas	Loreto	2017	Kattenberg et al., 2023
SAMN38982439	PE-PV-395	AmpliSeq	ILLUMINA	SRS19949040	PRJNA1055117	SRR27305975	Visto Bueno	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982440	PE-PV-396	AmpliSeq	ILLUMINA	SRS19949041	PRJNA1055117	SRR27305974	Visto Bueno	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982441	PE-PV-398	AmpliSeq	ILLUMINA	SRS19949042	PRJNA1055117	SRR27305973	Salvador	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982442	PE-PV-402	AmpliSeq	ILLUMINA	SRS19949043	PRJNA1055117	SRR27305972	Sucusari	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982443	PE-PV-404	AmpliSeq	ILLUMINA	SRS19949044	PRJNA1055117	SRR27305971	Sucusari	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982444	PE-PV-405	AmpliSeq	ILLUMINA	SRS19949045	PRJNA1055117	SRR27305970	Sucusari	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982445	PE-PV-409	AmpliSeq	ILLUMINA	SRS19949046	PRJNA1055117	SRR27305969	Puerto Alegre	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982446	PE-PV-411	AmpliSeq	ILLUMINA	SRS19949047	PRJNA1055117	SRR27305968	Libertad	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982447	PE-PV-412	AmpliSeq	ILLUMINA	SRS19949048	PRJNA1055117	SRR27305967	Libertad	Mazan	Maynas	Loreto	2016	Kattenberg et al., 2023
SAMN38982448	PE-PV-413	AmpliSeq	ILLUMINA	SRS19949049	PRJNA1055117	SRR27305966	Nueva Nazareth	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982449	PE-PV-414	AmpliSeq	ILLUMINA	SRS19949051	PRJNA1055117	SRR27305964	Nueva Nazareth	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982450	PE-PV-415	AmpliSeq	ILLUMINA	SRS19949052	PRJNA1055117	SRR27305963	Nueva Nazareth	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023

SAMN38982451	PE-PV-416	AmpliSeq	ILLUMINA	SRS19949053	PRJNA1055117	SRR27305962	Nueva Nazareth	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982452	PE-PV-417	AmpliSeq	ILLUMINA	SRS19949055	PRJNA1055117	SRR27305961	Nueva Nazareth	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982453	PE-PV-418	AmpliSeq	ILLUMINA	SRS19949054	PRJNA1055117	SRR27305960	Nueva Jerusalem	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982454	PE-PV-419	AmpliSeq	ILLUMINA	SRS19949056	PRJNA1055117	SRR27305959	Nueva Jerusalem	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982455	PE-PV-420	AmpliSeq	ILLUMINA	SRS19949057	PRJNA1055117	SRR27305958	Nueva Jerusalem	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982456	PE-PV-421	AmpliSeq	ILLUMINA	SRS19949058	PRJNA1055117	SRR27305957	Nueva Jerusalem	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982457	PE-PV-422	AmpliSeq	ILLUMINA	SRS19949059	PRJNA1055117	SRR27305956	Nueva Jerusalem	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982458	PE-PV-423	AmpliSeq	ILLUMINA	SRS19949060	PRJNA1055117	SRR27305955	Nueva Jerusalem	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982459	PE-PV-424	AmpliSeq	ILLUMINA	SRS19949062	PRJNA1055117	SRR27305953	Nueva Jerusalem	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982460	PE-PV-425	AmpliSeq	ILLUMINA	SRS19949063	PRJNA1055117	SRR27305952	Nueva Jerusalem	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982461	PE-PV-426	AmpliSeq	ILLUMINA	SRS19949064	PRJNA1055117	SRR27305951	Nueva Nazareth	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982462	PE-PV-427	AmpliSeq	ILLUMINA	SRS19949065	PRJNA1055117	SRR27305950	Nueva Nazareth	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982463	PE-PV-428	AmpliSeq	ILLUMINA	SRS19949066	PRJNA1055117	SRR27305949	Nueva Nazareth	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982464	PE-PV-429	AmpliSeq	ILLUMINA	SRS19949067	PRJNA1055117	SRR27305948	Nueva Jerusalem	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982465	PE-PV-430	AmpliSeq	ILLUMINA	SRS19949068	PRJNA1055117	SRR27305947	Nueva Jerusalem	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982466	PE-PV-431	AmpliSeq	ILLUMINA	SRS19949069	PRJNA1055117	SRR27305946	Nueva Jerusalem	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982467	PE-PV-432	AmpliSeq	ILLUMINA	SRS19949070	PRJNA1055117	SRR27305945	Nueva Jerusalem	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982468	PE-PV-434	AmpliSeq	ILLUMINA	SRS19949071	PRJNA1055117	SRR27305944	Nueva Jerusalem	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982469	PE-PV-435	AmpliSeq	ILLUMINA	SRS19949073	PRJNA1055117	SRR27305942	Nueva Jerusalem	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982470	PE-PV-436	AmpliSeq	ILLUMINA	SRS19949074	PRJNA1055117	SRR27305941	Nueva Jerusalem	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023

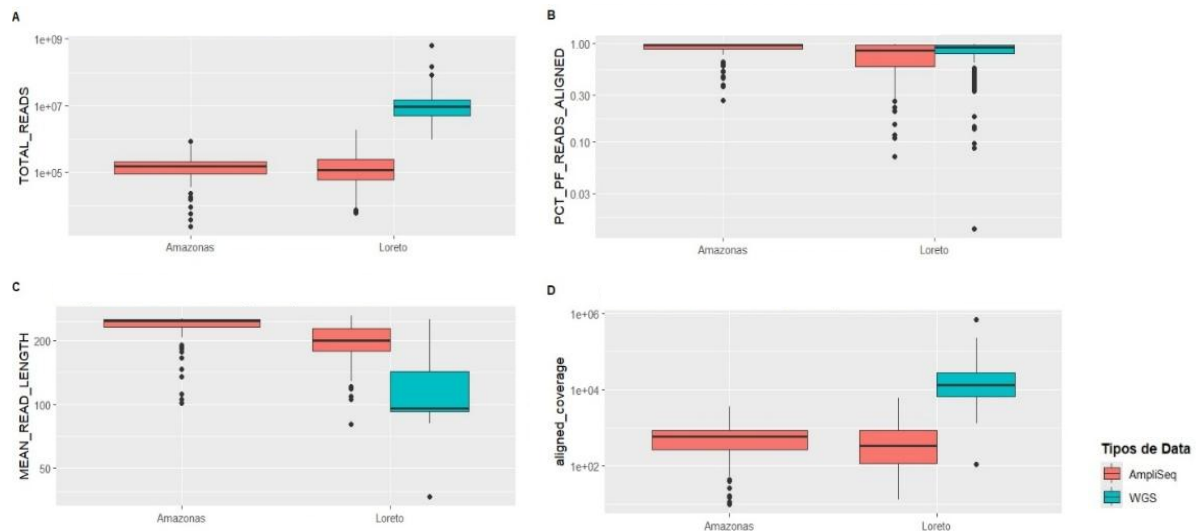
SAMN38982471	PE-PV-438	AmpliSeq	ILLUMINA	SRS19949075	PRJNA1055117	SRR27305940	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982472	PE-PV-439	AmpliSeq	ILLUMINA	SRS19949076	PRJNA1055117	SRR27305939	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982473	PE-PV-440	AmpliSeq	ILLUMINA	SRS19949077	PRJNA1055117	SRR27305938	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982474	PE-PV-441	AmpliSeq	ILLUMINA	SRS19949078	PRJNA1055117	SRR27305937	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982475	PE-PV-442	AmpliSeq	ILLUMINA	SRS19949079	PRJNA1055117	SRR27305936	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982476	PE-PV-443	AmpliSeq	ILLUMINA	SRS19949080	PRJNA1055117	SRR27305935	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982477	PE-PV-444	AmpliSeq	ILLUMINA	SRS19949081	PRJNA1055117	SRR27305934	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982478	PE-PV-445	AmpliSeq	ILLUMINA	SRS19949082	PRJNA1055117	SRR27305933	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982479	PE-PV-446	AmpliSeq	ILLUMINA	SRS19949084	PRJNA1055117	SRR27305931	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982480	PE-PV-447	AmpliSeq	ILLUMINA	SRS19949085	PRJNA1055117	SRR27305930	Nueva Nazareth	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982481	PE-PV-448	AmpliSeq	ILLUMINA	SRS19949086	PRJNA1055117	SRR27305929	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982482	PE-PV-449	AmpliSeq	ILLUMINA	SRS19949087	PRJNA1055117	SRR27305928	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982483	PE-PV-450	AmpliSeq	ILLUMINA	SRS19949088	PRJNA1055117	SRR27305927	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982484	PE-PV-451	AmpliSeq	ILLUMINA	SRS19949089	PRJNA1055117	SRR27305926	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982485	PE-PV-452	AmpliSeq	ILLUMINA	SRS19949090	PRJNA1055117	SRR27305925	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982486	PE-PV-453	AmpliSeq	ILLUMINA	SRS19949091	PRJNA1055117	SRR27305924	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982487	PE-PV-454	AmpliSeq	ILLUMINA	SRS19949092	PRJNA1055117	SRR27305923	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982488	PE-PV-455	AmpliSeq	ILLUMINA	SRS19949093	PRJNA1055117	SRR27305922	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982489	PE-PV-456	AmpliSeq	ILLUMINA	SRS19949095	PRJNA1055117	SRR27305920	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982490	PE-PV-457	AmpliSeq	ILLUMINA	SRS19949096	PRJNA1055117	SRR27305919	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023

SAMN38982491	PE-PV-458	AmpliSeq	ILLUMINA	SRS19949097	PRJNA1055117	SRR27305918	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982492	PE-PV-459	AmpliSeq	ILLUMINA	SRS19949098	PRJNA1055117	SRR27305917	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982493	PE-PV-460	AmpliSeq	ILLUMINA	SRS19949099	PRJNA1055117	SRR27305916	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982494	PE-PV-461	AmpliSeq	ILLUMINA	SRS19949100	PRJNA1055117	SRR27305915	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982495	PE-PV-462	AmpliSeq	ILLUMINA	SRS19949101	PRJNA1055117	SRR27305914	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2019	Kattenberg et al., 2023
SAMN38982496	PE-PV-463	AmpliSeq	ILLUMINA	SRS19949103	PRJNA1055117	SRR27305913	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2020	Kattenberg et al., 2023
SAMN38982497	PE-PV-464	AmpliSeq	ILLUMINA	SRS19949102	PRJNA1055117	SRR27305912	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2020	Kattenberg et al., 2023
SAMN38982498	PE-PV-465	AmpliSeq	ILLUMINA	SRS19949104	PRJNA1055117	SRR27305911	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2020	Kattenberg et al., 2023
SAMN38982499	PE-PV-466	AmpliSeq	ILLUMINA	SRS19949107	PRJNA1055117	SRR27305908	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2020	Kattenberg et al., 2023
SAMN38982500	PE-PV-467	AmpliSeq	ILLUMINA	SRS19949108	PRJNA1055117	SRR27305907	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2020	Kattenberg et al., 2023
SAMN38982501	PE-PV-468	AmpliSeq	ILLUMINA	SRS19949109	PRJNA1055117	SRR27305906	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2020	Kattenberg et al., 2023
SAMN38982502	PE-PV-469	AmpliSeq	ILLUMINA	SRS19949110	PRJNA1055117	SRR27305905	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2020	Kattenberg et al., 2023
SAMN38982503	PE-PV-470	AmpliSeq	ILLUMINA	SRS19949111	PRJNA1055117	SRR27305904	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2020	Kattenberg et al., 2023
SAMN38982504	PE-PV-471	AmpliSeq	ILLUMINA	SRS19949113	PRJNA1055117	SRR27305903	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2020	Kattenberg et al., 2023
SAMN38982505	PE-PV-472	AmpliSeq	ILLUMINA	SRS19949112	PRJNA1055117	SRR27305902	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2020	Kattenberg et al., 2023
SAMN38982506	PE-PV-473	AmpliSeq	ILLUMINA	SRS19949114	PRJNA1055117	SRR27305901	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2020	Kattenberg et al., 2023
SAMN38982507	PE-PV-474	AmpliSeq	ILLUMINA	SRS19949115	PRJNA1055117	SRR27305900	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2020	Kattenberg et al., 2023
SAMN38982508	PE-PV-475	AmpliSeq	ILLUMINA	SRS19949116	PRJNA1055117	SRR27305899	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2020	Kattenberg et al., 2023
SAMN38982509	PE-PV-476	AmpliSeq	ILLUMINA	SRS19949118	PRJNA1055117	SRR27305897	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2020	Kattenberg et al., 2023
SAMN38982510	PE-PV-477	AmpliSeq	ILLUMINA	SRS19949120	PRJNA1055117	SRR27305896	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2020	Kattenberg et al., 2023

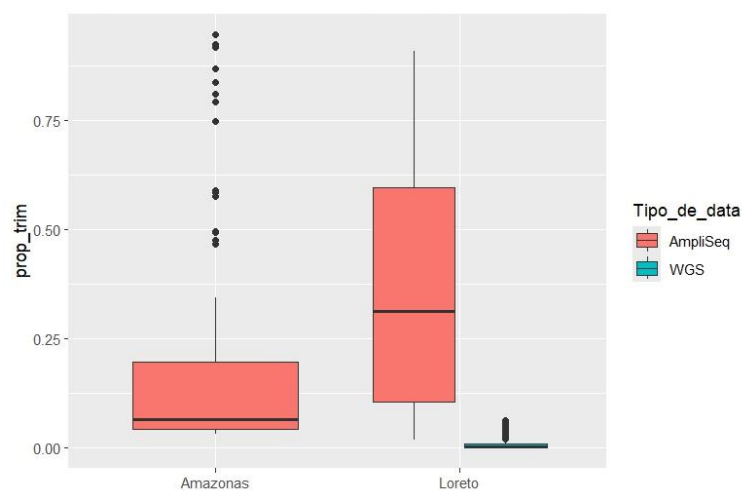
SAMN38982511	PE-PV-478	AmpliSeq	ILLUMINA	SRS19949119	PRJNA1055117	SRR27305895	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2020	Kattenberg et al., 2023
SAMN38982512	PE-PV-479	AmpliSeq	ILLUMINA	SRS19949121	PRJNA1055117	SRR27305894	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2020	Kattenberg et al., 2023
SAMN38982513	PE-PV-480	AmpliSeq	ILLUMINA	SRS19949122	PRJNA1055117	SRR27305893	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2020	Kattenberg et al., 2023
SAMN38982514	PE-PV-481	AmpliSeq	ILLUMINA	SRS19949123	PRJNA1055117	SRR27305892	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2020	Kattenberg et al., 2023
SAMN38982515	PE-PV-482	AmpliSeq	ILLUMINA	SRS19949124	PRJNA1055117	SRR27305891	Nueva Jerusalen	Trompeteros	Loreto	Loreto	2020	Kattenberg et al., 2023
SAMN38982516	PE-PV-483	AmpliSeq	ILLUMINA	SRS19949125	PRJNA1055117	SRR27305890	Gamitanacocha	Mazan	Maynas	Loreto	2018	Kattenberg et al., 2023
SAMN38982517	PE-PV-484	AmpliSeq	ILLUMINA	SRS19949127	PRJNA1055117	SRR27305889	Gamitanacocha	Mazan	Maynas	Loreto	2018	Kattenberg et al., 2023
SAMN38982518	PE-PV-486	AmpliSeq	ILLUMINA	SRS19949126	PRJNA1055117	SRR27305888	Gamitanacocha	Mazan	Maynas	Loreto	2018	Kattenberg et al., 2023
SAMN38982519	PE-PV-487	AmpliSeq	ILLUMINA	SRS19949129	PRJNA1055117	SRR27305886	Libertad	Mazan	Maynas	Loreto	2018	Kattenberg et al., 2023
SAMN38982520	PE-PV-488	AmpliSeq	ILLUMINA	SRS19949130	PRJNA1055117	SRR27305885	Libertad	Mazan	Maynas	Loreto	2018	Kattenberg et al., 2023
SAMN38982521	PE-PV-489	AmpliSeq	ILLUMINA	SRS19949131	PRJNA1055117	SRR27305884	Libertad	Mazan	Maynas	Loreto	2018	Kattenberg et al., 2023
SAMN38982522	PE-PV-490	AmpliSeq	ILLUMINA	SRS19949132	PRJNA1055117	SRR27305883	Libertad	Mazan	Maynas	Loreto	2018	Kattenberg et al., 2023
SAMN38982523	PE-PV-492	AmpliSeq	ILLUMINA	SRS19949133	PRJNA1055117	SRR27305882	Lago yuracyacu	Mazan	Maynas	Loreto	2018	Kattenberg et al., 2023
SAMN38982524	PE-PV-493	AmpliSeq	ILLUMINA	SRS19949134	PRJNA1055117	SRR27305881	Puerto Alegre	Mazan	Maynas	Loreto	2018	Kattenberg et al., 2023
SAMN38982525	PE-PV-495	AmpliSeq	ILLUMINA	SRS19949135	PRJNA1055117	SRR27305880	Libertad	Mazan	Maynas	Loreto	2018	Kattenberg et al., 2023
SAMN38982526	PE-PV-496	AmpliSeq	ILLUMINA	SRS19949136	PRJNA1055117	SRR27305879	Libertad	Mazan	Maynas	Loreto	2018	Kattenberg et al., 2023
SAMN38982527	PE-PV-497	AmpliSeq	ILLUMINA	SRS19949137	PRJNA1055117	SRR27305878	Libertad	Mazan	Maynas	Loreto	2018	Kattenberg et al., 2023
SAMN38982528	PE-PV-498	AmpliSeq	ILLUMINA	SRS19949138	PRJNA1055117	SRR27305877	Puerto Alegre	Mazan	Maynas	Loreto	2018	Kattenberg et al., 2023

**SN: sin código asignado*

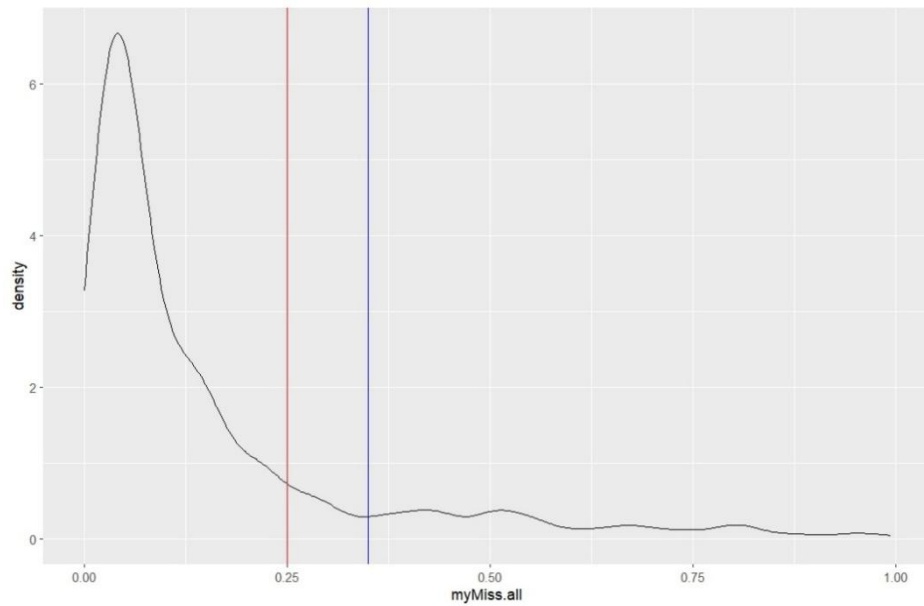
Anexo 2. Parámetros de secuenciación. Diagramas de caja que muestran las características de los reads obtenidos mediante las plataformas Amplicon y WGS, diferenciados por regiones (Amazonas y Loreto). A) Número total de reads; B) Porcentaje de reads alineados; C) Longitud media de los reads; y D) Cobertura de alineamiento.



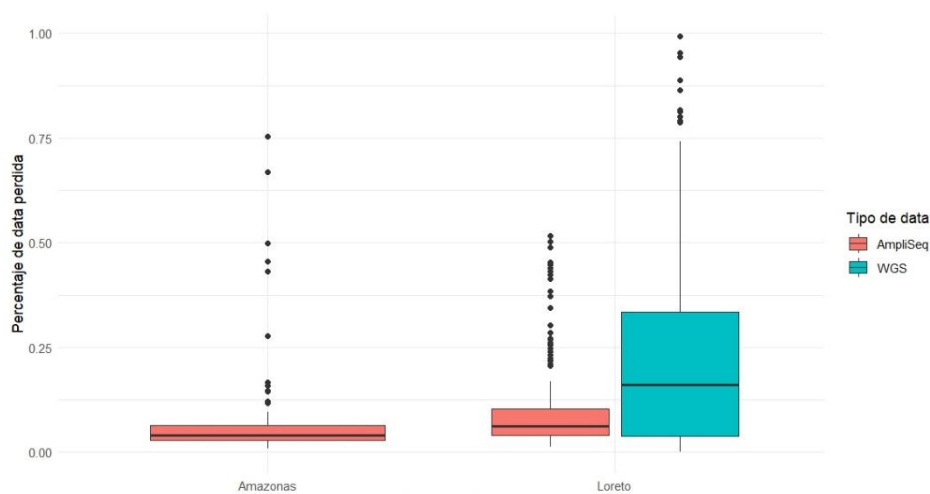
Anexo 3. Proporción de reads trimeados. Diagrama de caja que muestra la proporción de reads eliminados durante el proceso de trimming en las secuencias obtenidas mediante Amplicon y WGS, diferenciadas por región



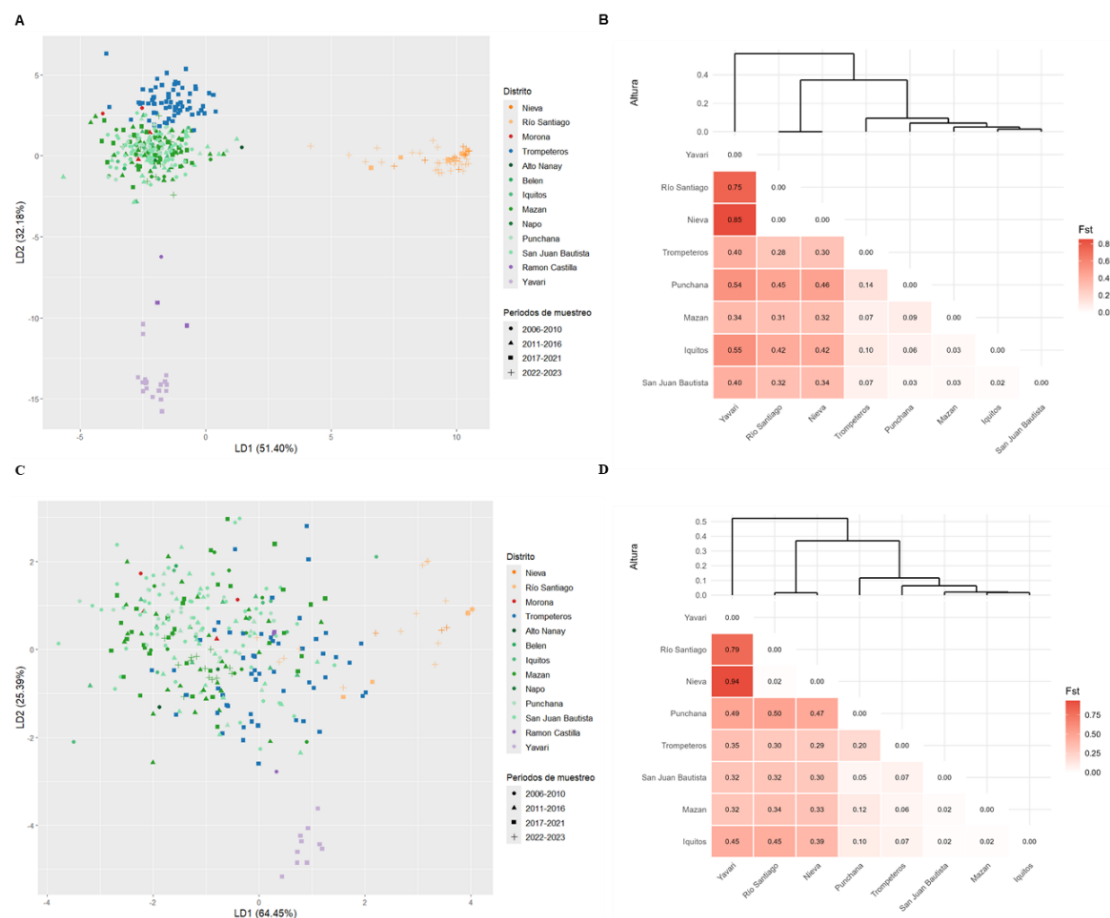
Anexo 4. Distribución de datos faltantes (missing genotypes). Curva de densidad que representa la proporción de genotipos faltantes en las muestras analizadas. La línea roja indica el promedio de datos faltantes y la línea azul el umbral de filtrado aplicado (35%). Aquellas muestras con un porcentaje de genotipos faltantes mayor al 35% fueron excluidas de los análisis.



Anexo 5. Porcentaje de datos faltantes. Diagrama de caja que muestra la proporción de datos faltantes en las secuencias obtenidas mediante AmpliSeq y WGS, diferenciadas por las regiones de Amazonas y Loreto.

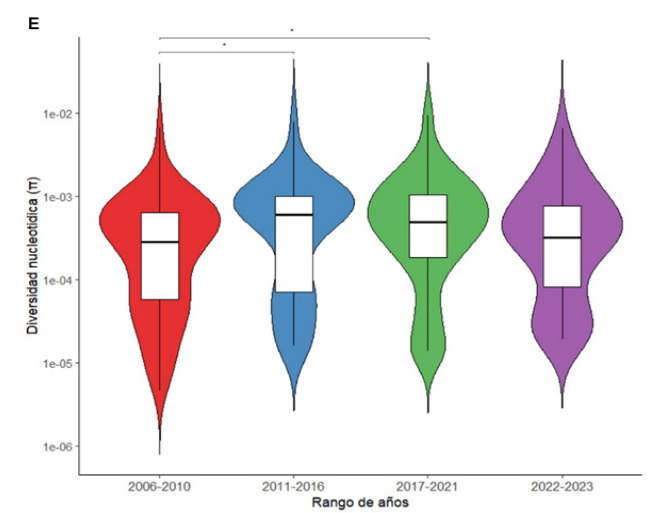
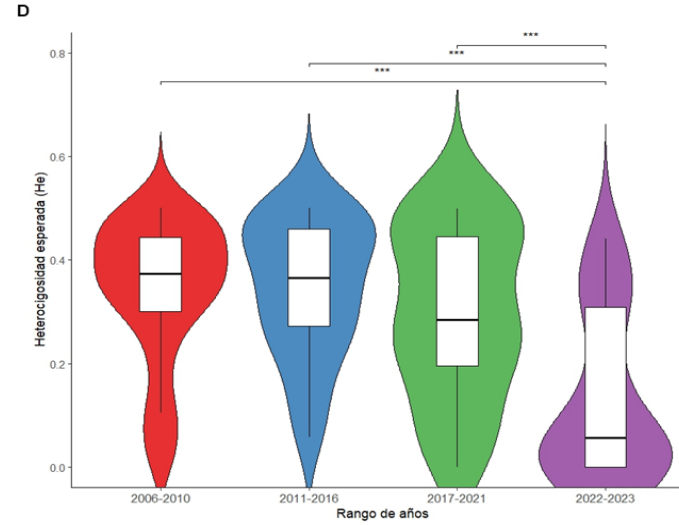
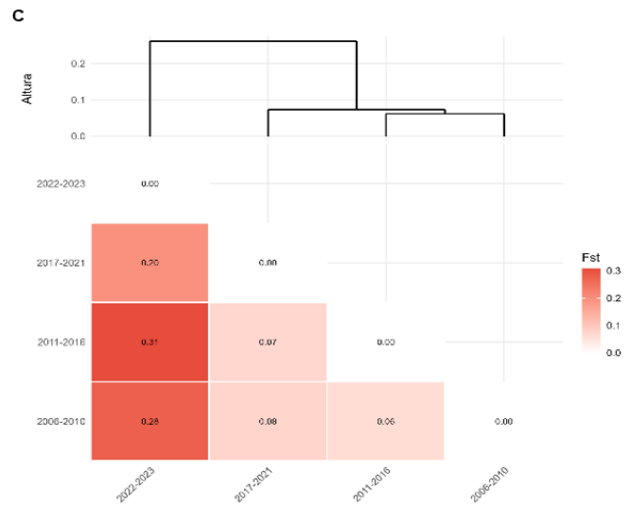
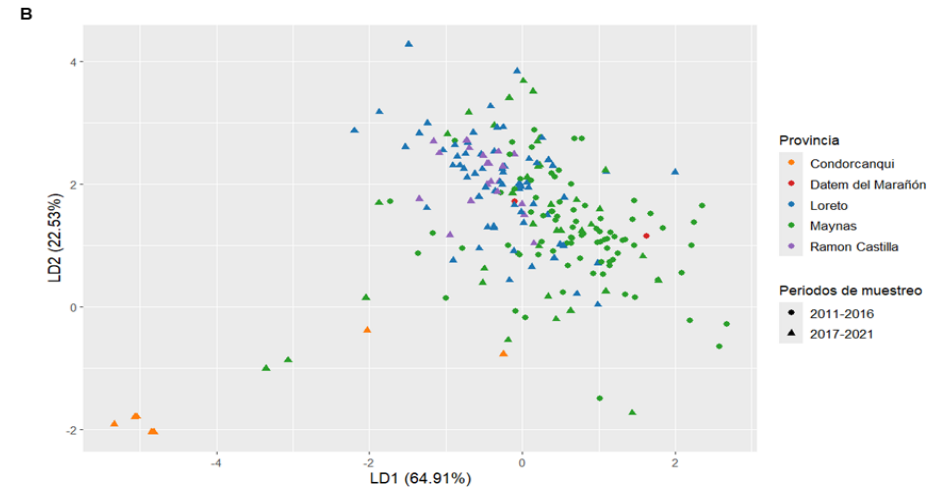
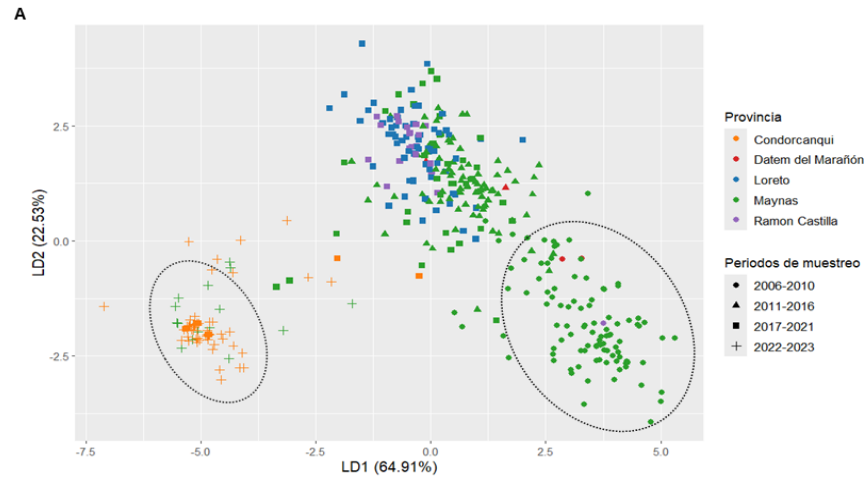


Anexo 6. Análisis Discriminante de Componentes Principales (DAPC) agrupados *a priori* por provincia y Diferenciación genética (F_{ST}). Patrones de estructuración genética de las muestras analizadas mediante DAPC, agrupadas *a priori* por provincia e identificadas según distrito y periodo de muestreo, utilizando el conjunto completo de SNPs bialélicos (A), junto con los valores pareados de F_{ST} calculados entre agrupaciones a nivel de distrito para este mismo conjunto de marcadores (B). Patrones de estructuración genética de las muestras analizadas mediante DAPC, identificadas según distrito y periodo de muestreo, utilizando el barcode *ITM-Peru* (C), junto con los valores pareados de F_{ST} calculados entre agrupaciones a nivel de distrito para este conjunto de marcadores (D).

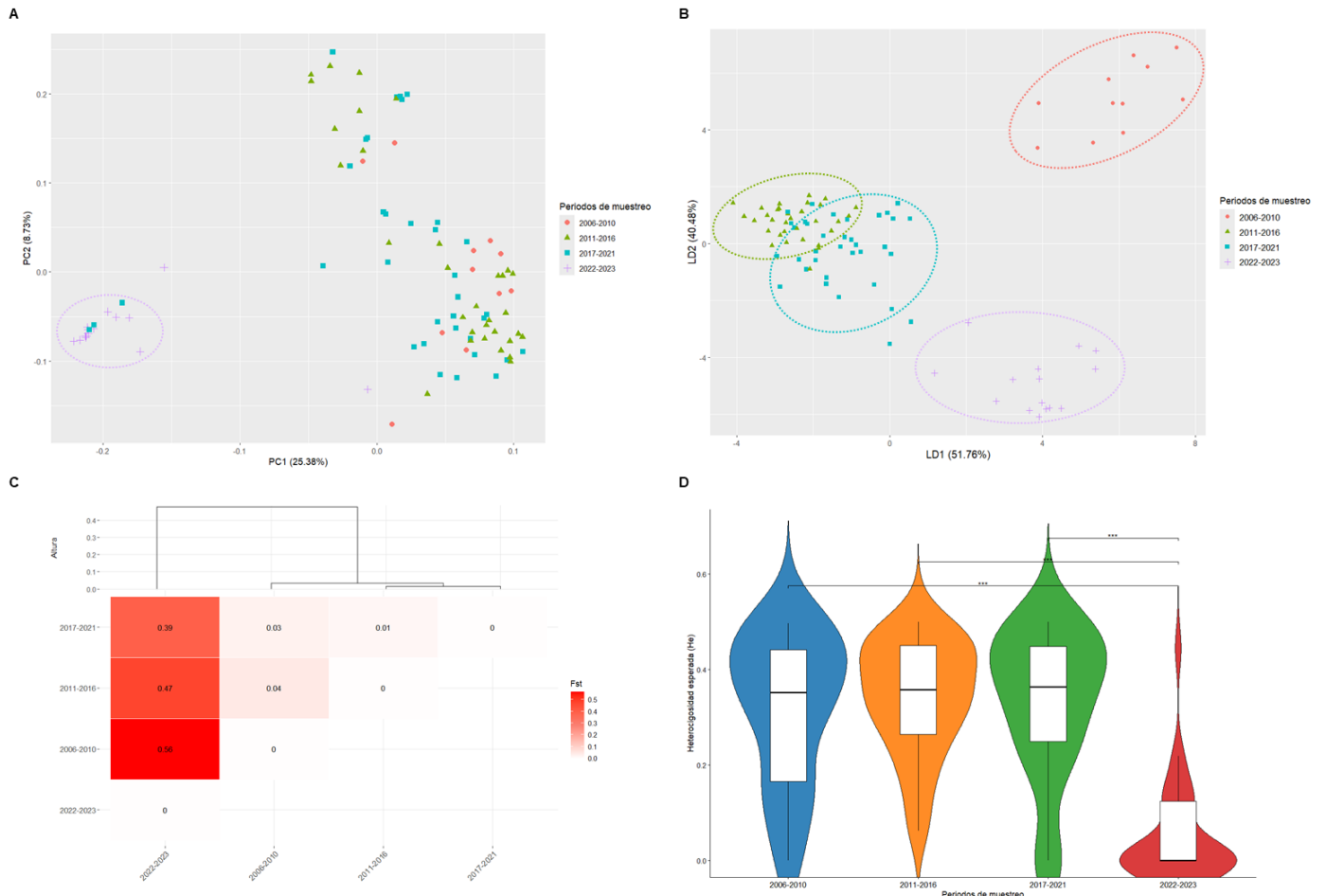


Anexo 7. Análisis de estructuración genética analizado por periodos de tiempo.

Análisis Discriminante de Componentes Principales (DAPC) de todos los aislados, agrupados a priori por periodo de muestreo, utilizando el conjunto completo de SNPs bialélicos (A). DAPC restringido a los aislados correspondientes a los periodos 2011–2016 y 2017–2021, también basado en el conjunto completo de SNPs bialélicos (B). A partir de las agrupaciones temporales parcialmente observadas en el DAPC, se presentan los valores pareados de F_{ST} entre periodos de muestreo (C). Se muestran además los valores de heterocigosidad esperada utilizando el barcode *ITM-Peru* (D) y los valores de diversidad nucleotídica estimados para las agrupaciones temporales (E). Las barras horizontales indican las comparaciones significativas entre grupos, donde * $p < 0.05$; ** $p < 0.01$ y *** $p < 0.001$



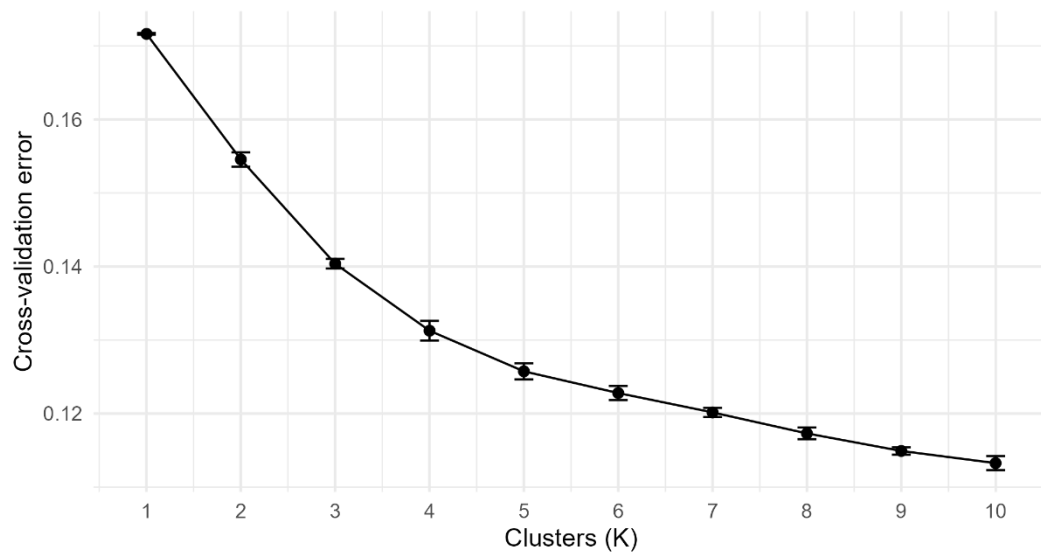
Anexo 8. Análisis genético de las muestras de Mazán. Análisis de Componentes Principales de las muestras de Mazán, visualizados según periodos de muestreo, sin evidencia de estructuraciones genéticas (A). Análisis Discriminante de Componentes Principales de las mismas muestras agrupadas *a priori* por periodo de muestreo, con evidencia de estrcuturación genética asociada al tiempo (B); ambos análisis realizados a partir del conjunto completo de SNPs bialélicos. Valores pareados de F_{ST} entre periodos de muestreo calculados a partir del conjutno completo de SNPs (C) y las estimaciones de hetericogidad esperada calculadas a partir del barcode del *ITM-Peru* (D).



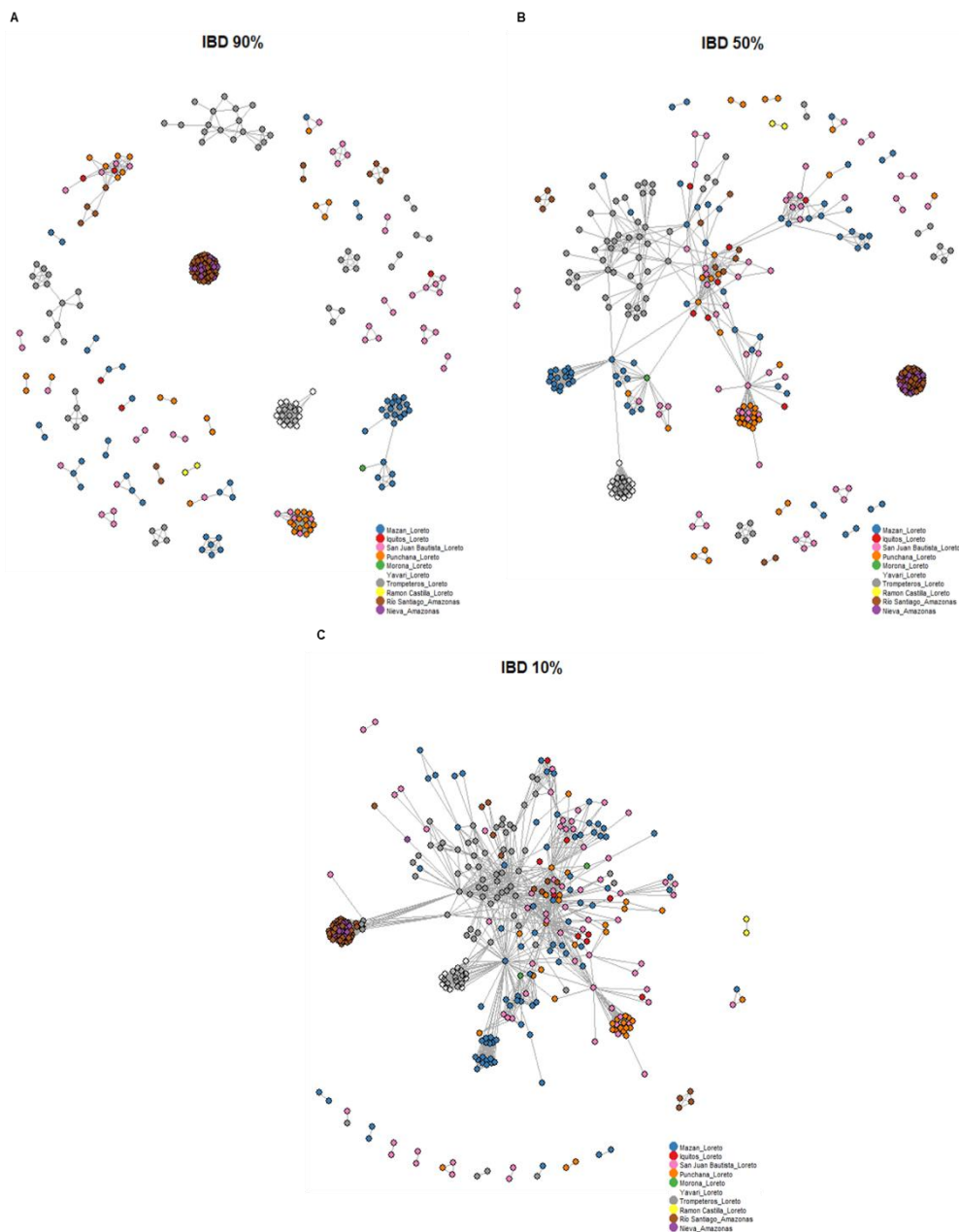
Anexo 9. Análisis de Varianza Molecular (AMOVA) por rango de años utilizando los 802 SNPs bialélicos. El AMOVA se realizó agrupando las muestras de Loreto y Amazonas en cuatro periodos: 2006–2010, 2011–2016, 2017–2021 y 2022–2023. Los resultados indican una diferenciación genética significativa entre los periodos temporales, no obstante, dado que las muestras provienen de distintos lugares, esta diferencia podría deberse a la variación espacial de las poblaciones y no exclusivamente a cambios temporales en su composición genética.

Por agrupaciones de rango de años							
Fuente de Variación	Df	Suma de Cuadrados	Media de Cuadrados	Componente de varianza	Porcentaje de variación	Indice de fijación	P-value
Entre rangos de años	3	1128.46	376.15343	3.598813	15.41223	0.1541223	0.001
Dentro de rangos de años	398	7861.121	19.75156	19.751561	84.58777		
Total	401	8989.581	22.41791	23.350374	100		

Anexo 10. Determinación del número óptimo de agrupamientos (K) mediante validación cruzada utilizando 802 SNPs bialélicos. Los valores del error de validación cruzada (Cross-validation error) obtenidos para diferentes números de clusters (K = 1–10) en el análisis de ADMIXTURE. Se observa una disminución progresiva del error al incrementar el número de grupos, con una estabilización a partir de K = 5



Anexo 11. Redes de identidad por descendencia (IBD) utilizando 802 SNPs bialélicos con diferentes umbrales de similitud. Cada plot emplea tres niveles de conectividad (A) IBD 90 %, (B) IBD 50 % y (C) IBD 10 %. Cada nodo representa un individuo y los colores indican la provincia y región de procedencia, mientras que las conexiones reflejan la proporción de alelos compartidos entre pares de individuos.



Anexo 12. Métricas de centralidad derivadas del árbol de Neighbor-Joining (NJ) construido con 802 SNPs bialélicos, calculadas en StrainHub. La tabla muestra los valores de centralidad de grado (Degree Centrality), grado de entrada (Indegree Centrality), grado de salida (Outdegree Centrality), intermediación (Betweenness Centrality), cercanía (Closeness Centrality) y proporción de nodos fuente (Source Hub Ratio) para cada provincia analizada.

Provincia	Degree Centrality	Indegree Centrality	Outdegree Centrality	Betweenness Centrality	Closeness Centrality	Source Hub Ratio
Condorcanqui	3	2	1	0	0.16666667	0.33333333
Datem del Marañon	1	1	0	0	0.14285714	0
Loreto	4	2	2	3	0.16666667	0.5
Maynas	5	1	4	4	0.25	0.8
Ramon Castilla	1	1	0	0	0.14285714	0

Anexo 13. Representación de la red de transmisión entre provincias basada en el índice Source Hub Ratio obtenido en StrainHub. El tamaño de los nodos indica su centralidad (importancia relativa de cada provincia dentro de la red) y el grosor de las flechas la magnitud del flujo de transmisión inferido.

