

**UNIVERSIDAD NACIONAL
TORIBIO RODRÍGUEZ DE MENDOZA DE AMAZONAS**



ESCUELA DE POSGRADO

**TESIS PARA OBTENER EL GRADO ACADÉMICO DE MAESTRO
EN GESTIÓN PARA EL DESARROLLO SUSTENTABLE**

**ANÁLISIS DEL GENOMA PLASTIDIAL Y
MITOCONDRIAL DEL CACAO FINO DE AROMA
(*Theobroma cacao* L.) DE AMAZONAS, PERÚ**

Autor:

Bach. Daniel Tineo Flores

Asesor:

Ph.D. Danilo Edson Bustamante Mostajo

Co - Asesora:

Ph.D. Martha Steffany Calderón Ríos

Registro: (...)

CHACHAPOYAS – PERÚ

2022

DEDICATORIA

A mis padres y todas las personas que han contribuido en el desarrollo de este proyecto.

AGRADECIMIENTO

A la Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, por la oportunidad brindada para concluir mis estudios de maestría.

Al Instituto de Investigación para el Desarrollo Sustentable de Ceja de Selva (INDES-CES), por la oportunidad brindada en para desarrollar esta investigación.

Al Dr. Segundo Manuel Oliva Cruz, por la oportunidad brindada en el proyecto CINCACAO Al Fondo Nacional de Desarrollo Científico, Tecnológico y de Innovación Tecnológica (FONDECYT) por financiar el proyecto con contrato N° 026-2016 del "Círculo de Investigación para la Innovación y el fortalecimiento de la cadena de valor del cacao nativo fino de aroma en la zona nor oriental del Perú-CINCACAO" proyecto ejecutado por Instituto de Investigación para el Desarrollo Sustentable de Ceja de Selva (INDES - CES), de la Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas.

A mis asesores del proyecto de tesis, Ph.D. Danilo Edson Bustamante Mostajo y Ph.D. Martha Steffany Calderón Ríos.

Al laboratorio FISIODEV, por brindar los espacios y equipos para cumplir con los objetivos de dicha investigación.

**AUTORIDADES DE LA UNIVERSIDAD NACIONAL TORIBIO RODRÍGUEZ DE
MENDOZA DE AMAZONAS**

Ph.D. Jorge Luis Maicelo Quintana
RECTOR

Dr. Oscar Andrés Gamarra Torres
VICERRECTOR ACADÉMICO

Dra. María Nelly Luján Espinoza
VICERRECTORA DE INVESTIGACIÓN

Dr. Efraín Manuelito Castro Alayo
DIRECTOR DE LA ESCUELA DE POSGRADO

JURADO EVALUADOR DE LA TESIS



Ph.D Ligia Magali García Rosero

PRESIDENTE



Dr. Segundo Manuel Oliva Cruz.

SECRETARIO



Mg. Segundo Grimaldo Chávez Quintana.

VOCAL

CONSTANCIA DE ORIGINALIDAD DE LA TESIS



UNTRM

REGLAMENTO GENERAL
PARA EL OTORGAMIENTO DEL GRADO ACADÉMICO DE
BACHILLER, MAESTRO O DOCTOR Y DEL TÍTULO PROFESIONAL

ANEXO 3

CONSTANCIA DE ORIGINALIDAD

Los suscritos, miembros del Jurado Evaluador del Proyecto de Tesis ()/Tesis (X)/Tesis en formato de artículo científico () titulado:

*Análisis del genoma plastidial y mitocondrial del cacao
fino de aroma (Theobroma cacao L.) de Amazonas, Perú.*

presentado por el Aspirante Boch Daniel Tines Pires para obtener el Grado Académico de Maestro (X)/Doctor () en Ecología para el Desarrollo Sostenible de la Escuela de Posgrado de la UNTRM, hacemos constar que después de revisar la originalidad del Proyecto de Tesis ()/Tesis (X)/Tesis en formato de artículo científico () con el software de prevención de plagio **Turnitin**, verificamos:

- a) De acuerdo con el informe de originalidad, el Proyecto de Tesis ()/Tesis (X)/Tesis en formato de artículo científico () tiene 16 % de similitud, que es menor al 25% permitido en la UNTRM.
- b) La persona responsable de someter el trabajo al software de prevención de plagio **Turnitin** fue: Ph.D. Ligia Magali García Poseo y pertenece al área () / oficina () / dependencia (X) de Facultad de Ingeniería Zootecnista, Agonegocios y Biotecnología



SE ADJUNTA:

- Resultado del informe del software **Turnitin**.

Chachapoyas, 10 de junio del 2022

Ligia Magali García Poseo
PRESIDENTE

Nombres y apellidos: Ligia Magali García Poseo

DNI: CE 001691738

Segundo Germán
VOCAL

Nombres y apellidos: Segundo Germán

DNI: 44011651 Chary Avilona

Segundo Manuel Oliva Cruz
SECRETARIO

Nombres y apellidos: Segundo Manuel Oliva Cruz

DNI: 00324749

OBSERVACIONES:

.....
.....

REPORTE DE TURNITIN

ANÁLISIS DEL GENOMA PLASTIDIAL Y MITOCONDRIAL DEL CACAO FINO DE AROMA (*Theobroma cacao* L.) DE AMAZONAS, PERÚ

INFORME DE ORIGINALIDAD

16%	15%	7%	6%
INDICE DE SIMILITUD	FUENTES DE INTERNET	PUBLICACIONES	TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	Submitted to Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas Trabajo del estudiante	3%
2	repositorio.untrm.edu.pe Fuente de Internet	2%
3	repositorio.serfor.gob.pe Fuente de Internet	1%
4	amoebidia.bcm.umontreal.ca Fuente de Internet	1%
5	hdl.handle.net Fuente de Internet	1%
6	link.springer.com Fuente de Internet	<1%
7	pesquisa.bvsalud.org Fuente de Internet	<1%
8	core.ac.uk Fuente de Internet	<1%



ACTA DE SUSTENTACIÓN DE LA TESIS



UNTRM

REGLAMENTO GENERAL
PARA EL OTORGAMIENTO DEL GRADO ACADÉMICO DE
BACHILLER, MAESTRO O DOCTOR Y DEL TÍTULO PROFESIONAL

ANEXO 5

ACTA DE SUSTENTACIÓN DE TESIS

En el Auditorio de la Escuela de Posgrado de la UNTRM - Chachapoyas, el día 2 de Julio del año 2022, siendo las 15:00 horas, el Aspirante Boch Daniel Tineo Flores, cuyo asesor es Ph.D. Danilo Edson Bustamante Mostajo, defiende en sesión pública presencial la Tesis titulada: Análisis del genoma plasmidial y mitocondrial del cacao fino de aroma (Theobroma cacao L.) de Amazonas, Perú.

para obtener el Grado Académico de Maestro (X) / Doctor () en Exestión para el Desarrollo sustentable, a ser otorgado por la Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas; ante el Jurado Evaluador, conformado por:


Presidente: Ph. D. Ligia Mogali García Posero
Secretario: Dr. Segundo Manuel Oliva Cruz
Vocal: Mg. Segundo Brimbaldo Chavez Quintana


Luego de la sustentación y absueltas las preguntas del Jurado Evaluador se procedió a la calificación individual y secreta, teniendo el resultado de:


Aprobada (X)/Desaprobada () por Unanimidad (X)/Mayoría ().

Otorgada la calificación, el Secretario del Jurado Evaluador lee la presente Acta en esta misma sesión pública. A continuación, se levanta la sesión.

Siendo las 17:00 horas del mismo día y fecha, el Jurado Evaluador concluye el acto de sustentación de la Tesis.


PRESIDENTE
Nombres y apellidos: Ligia Mogali García Posero
DNI: CE001691738


VOCAL
Nombres y apellidos: Segundo Brimbaldo Chavez Quintana
DNI: 42001657


SECRETARIO
Nombres y apellidos: Segundo Manuel Oliva Cruz
DNI: 05374749

ÍNDICE GENERAL

DEDICATORIA.....	ii
AGRADECIMIENTO	iii
AUTORIDADES DE LA UNIVERSIDAD NACIONAL TORIBIO RODRÍGUEZ DE MENDOZA DE AMAZONAS	iv
JURADO EVALUADOR DE LA TESIS.....	v
CONSTANCIA DE ORIGINALIDAD DE LA TESIS	vi
REPORTE DE TURNITIN	vii
ACTA DE SUSTENTACIÓN DE LA TESIS.....	viii
ÍNDICE GENERAL.....	ix
ÍNDICE DE TABLAS.....	x
ÍNDICE DE FIGURAS.....	xi
RESUMEN.....	xii
ABSTRACT.....	xiii
I. INTRODUCCIÓN.....	14
II. MATERIALES Y MÉTODOS.....	17
2.1. Área de estudio.....	17
2.2. Colecta de muestras de cacao fino de aroma	17
2.3. Análisis genómico del cacao fino de aroma.....	18
2.3.1. <i>Extracción de ADN genómico</i>	18
2.3.2. <i>Cuantificación de ADN genómico extraído</i>	20
2.3.3. <i>Preparación de ADN genómico y secuenciamiento masivo</i>	21
2.3.4. <i>Ensamblaje del ADN genómico secuenciado</i>	22
2.3.5. <i>Anotación de genes plastidiales</i>	22
2.4. Análisis de las relaciones evolutivas del cacao fino de aroma.....	22
2.4.1. <i>Análisis comparativo de genomas plastidiales</i>	22
2.4.2. <i>Análisis filogenéticos de los genomas plastidiales y mitocondriales</i>	22
2.5. Implicancia de los análisis moleculares en la producción sostenible	23
III. RESULTADOS.....	24
3.1. Genomas plastidiales del cacao fino de aroma.....	24
3.1.1. <i>Secuenciación y ensamblaje de genomas plastidiales</i>	24
3.1.2. <i>Anotación de genes de los genomas plastidiales</i>	25
3.1.3. <i>Generación de mapas genómicos plastidiales</i>	30
3.2. Genomas mitocondriales del cacao fino de aroma	32
3.3. Relaciones evolutivas del cacao fino de aroma.....	33
3.3.1. <i>Análisis comparativo de genomas plastidiales</i>	33
3.3.2. <i>Análisis filogenómicos</i>	34
3.3. Implicancia de las investigaciones del cacao en la producción sostenible	35
IV. DISCUSIÓN	42
V. CONCLUSIONES.....	49
VI. RECOMENDACIONES.....	51
VII. REFERENCIAS BIBLIOGRÁFICAS.....	52

ÍNDICE DE TABLAS

Tabla 1. Lista de voucher de genomas plastidiales de especies de <i>Theobroma</i> utilizadas en los análisis filogenómicos.....	23
Tabla 2. Tamaño de los reads limpios de muestra de Cacao Fino de Aroma de la Región Amazonas secuenciada por Illumina.	24
Tabla 3. Tamaño de los genomas plastidiales del Cacao Fino de Aroma de la Región Amazonas.	25
Tabla 4. Características de los genomas plastidiales de Cacao Fino de Aroma y la especie más cerca, <i>T. grandiflorum</i>	25
Tabla 5. Comparación del tamaño de genes y su funcionalidad con el intercambio de electrones y fotosíntesis de los genomas plastidiales de Cacao Fino de Aroma.	27
Tabla 6. Número de deleciones e inserciones a lo largo de los 4 genomas plastidiales del Cacao Fino de Aroma.	29
Tabla 7. Porcentaje de publicaciones según el tipo de base de datos e idioma de publicación.....	37
Tabla 8. Posibles supuestos de la importancia de los estudios de genómica sobre el cacao peruano.....	40

ÍNDICE DE FIGURAS

Figura 1. Mapa de localización del área de estudio donde se obtuvo el material vegetal de Cacao Fino de Aroma de Amazonas.	17
Figura 2. Colecta de muestras del Cacao Fino de Aroma.	18
Figura 3. Proceso de extracción de ADN genómico del Cacao Fino de Aroma.	20
Figura 4. Proceso de Cuantificación de ADN genómico del Cacao Fino de Aroma.	21
Figura 5. Mapa del genoma plastidial del Cacao Fino de Aroma (INDES34_6a).....	30
Figura 6. Mapa del genoma plastidial del Cacao Fino de Aroma (INDES34_6b).....	31
Figura 7. Mapa del genoma plastidial del Cacao Fino de Aroma (INDES67_3a).....	31
Figura 8. Mapa del genoma plastidial del Cacao Fino de Aroma (INDES67_3b).....	32
Figura 9. Genoma mitocondrial parcial del Cacao Fino de Aroma de Amazonas..	33
Figura 10. Gráfico de identidad porcentual obtenido de mVISTA para los cuatro genomas plastidiales de Cacao Fino de Aroma (INDES34_6a, INDES34_6b, INDES67_3a, INDES67_3b) y <i>T. cacao</i> (HQ336404) como referencia..	34
Figura 11. Filogenia de máxima verosimilitud de variedades de la especie <i>Theobroma cacao</i> basadas en genomas plastidiales completos..	35
Figura 12. Número de artículos publicados en relación al cacao peruano, según el área de estudio.....	36
Figura 13. Número de publicaciones relacionados a Cacao Fino de Aroma que se han realizado en todo el Perú, y en la región Amazonas durante los años 1981-mayo del 2022.	37
Figura 14. Mapa de países con mayor número de investigaciones relacionadas al Cacao de Perú.....	38
Figura 15. Mapa de instituciones nacionales e internacionales con mayor número de investigaciones relacionados al cacao.	39
Figura 16. Mapa de red de co-ocurrencia de palabras en los títulos de los artículos asociados a cacao del Perú..	40

RESUMEN

El cacao se encuentra entre los diez productos agrícolas más importantes del mundo. El presente estudio decodificó el genoma plastidial de Cacao Fino de Aroma de Amazonas utilizando técnicas de secuenciamiento masivo. En este estudio se obtuvieron dos genomas plastidiales completos de cada una de las muestras INDES34 e INDES67, mientras que los genomas mitocondriales fueron ensamblados parcialmente (~20%). La presencia de más de un genoma plastidial por cada ejemplar de cacao, confirmaría el primer reporte de “heteroplasmia” en plantas de cacao. El tamaño de los 4 genomas plastidiales varía desde 160,589 pb hasta 160,620 pb y contienen el mismo porcentaje de GC (36.9). Están compuestos por 130 genes (37 tRNA, 8 rRNA y 85 genes codificantes de proteínas), de los cuales 17 genes están duplicados en la IR (6 genes codificantes de proteínas, 4 rRNA y 7 tRNA). Los análisis filogenómicos revelaron que los genomas plastidiales INDES67_3a, INDES67_3b e INDES34_6a tienen mayor afinidad con los grupos genéticos de Nanay e Iquitos; mientras que INDES34_6b tiene mayor afinidad con el grupo genético Contamana. Este estudio resalta la importancia de los estudios genómicos para identificar distinciones entre variedades de cacao con características agronómicas demandadas por los agricultores. Adicionalmente, se confirma que las investigaciones en Cacao están más centradas en estudios bioquímicos. Sin embargo, hay pocos estudios relacionados al uso de herramientas moleculares, pero que han repercutiendo indirectamente en la mejora de los estándares de calidad de vida de la población.

Palabras claves: ADN, Cacao Fino de Aroma, genoma plastidial, genoma mitocondrial, *Theobroma cacao*.

ABSTRACT

Cocoa is among the ten most important agricultural products in the world. The present study decoded the plastid genome of Cacao Fino de Aroma de Amazonas using massive sequencing techniques. In this study, two complete plastid genomes were obtained from each of the INDES34 and INDES67 samples, while the mitochondrial genomes were partially assembled (~20%). The presence of more than one plastid genome for each cocoa specimen would confirm the first report of "heteroplasmy" in cocoa plants. The size of the 4 plastid genomes varies from 160,589 bp to 160,620 bp and they contain the same percentage of GC (36.9). They are composed of 130 genes (37 tRNA, 8 rRNA, and 85 protein-coding genes), of which 17 genes are duplicated in the IR (6 protein-coding genes, 4 rRNA, and 7 tRNA). The phylogenomic analyzes revealed that the plastid genomes INDES67_3a, INDES67_3b and INDES34_6a have greater affinity with the genetic groups of Nanay and Iquitos; while INDES34_6b has a higher affinity with the Contamana genetic group. This study highlights the importance of genomic studies to identify distinctions between cocoa varieties with agronomic characteristics demanded by farmers. Additionally, it is confirmed that research on Cocoa is more focused on biochemical studies. However, there are few studies related to the use of molecular tools, but they have had an indirect impact on improving the quality of life standards of the population.

Keywords: DNA, Fine Aroma Cocoa, mitochondrial genome, phylogenomics, plastid genome, *Theobroma cacao*.

I. INTRODUCCIÓN

El género *Theobroma* L. pertenece a la familia Malvaceae y agrupa 22 especies (Bayer et al., 1999), siendo el cacao (*Theobroma cacao* L.) la principal especie económica (Gopaulchan et al., 2019). El cacao se originó en la cuenca amazónica sudamericana y se presenta como un árbol tropical del sotobosque (Bartley, 2005). Este cacao se cultiva como plantaciones comerciales, cuyo crecimiento y producción se ha adaptado en África, Asia y América, y constituye una fuente importante de ingresos para muchos pequeños productores en esas regiones (Dantas y Guerra, 2010; Da Silva et al., 2016). En la actualidad, el cacao se encuentra entre los diez productos agrícolas más importantes del mundo (Utro et al., 2012). Los granos de cacao (semillas) son la fuente de chocolate, además de carbohidratos, grasas, proteínas, minerales naturales, flavonoides y vitaminas (World Cocoa Foundation, 2014; Boza et al., 2014).

El árbol del cacao fue domesticado a partir de ancestros salvajes en América Central (Motamayor et al., 2002) y actualmente se cultiva con fines comerciales en más de 50 países de regiones tropicales de América Central y del Sur, Asia y África (Wickramasuriya y Dunwell, 2018). Tradicionalmente, se designaron tres variedades de cacao; Forastero, Criollo y Trinitario, los cuales guardan relación con su extensión geográfica natural desde el sur de Bolivia hasta México (Toxopeus, 1985; Cheesman, 1944; Motamayor et al., 2002; 2008). Si bien existe cierta confusión sobre el significado de estos nombres en el registro histórico, la industria reconoce a la variedad Forastero como vigorosa con granos morados oscuros con un sabor relativamente amargo y, a menudo, un sabor ácido (Toxopeus, 1985; Cheesman, 1944; Loor et al. 2009). La variedad Criollo es menos vigorosa y produce granos blancos o ligeramente pigmentados con un fuerte aroma deseable y un ligero amargor (Toxopeus, 1985; Argout et al., 2011, 2017). Por otro lado, las variedades Trinitarias se caracterizan por su alto rendimiento y resistencia a enfermedades, los granos tienen un sabor mucho más sutil (Cheesman, 1944; Bekele et al., 2007; Motilal et al., 2010), y se cree que son híbridos derivados de las variedades Forastero y Criollo (Motamayor et al., 2003; Motilal y Sreenivasan, 2012; Yang et al., 2013).

Sin embargo, en un estudio con marcadores microsatélites se han reconocido diez grupos genéticos distintos de cacao: Amelonado (en Brasil, Costa Rica y Ghana); Contamana e Iquitos (en Perú y Brasil); Criollo (en Ecuador, Venezuela, Panamá, Costa Rica y México); Nacional y Curaray (sólo en Ecuador); Guiana (sólo en Brasil); Marañón (en Perú y Bolivia); Nanay (sólo en Perú); y Purús (en Brasil y Bolivia) (Motamayor et al., 2008). Recientes estudios han sugerido que podrían existir poblaciones adicionales en Bolivia (Zhang et al., 2012), Perú (Motamayor et al., 2010) y Colombia (Osorio-Guarín et al., 2017). Es probable que se encuentren otros grupos genéticos únicos de cacao a medida que se exploren más áreas silvestres en América del Sur (Gopaulchan et al., 2019). Actualmente, los análisis de polimorfismos de un solo nucleótido (SNP) tienen un mayor potencial para registrar la diversidad genética y descubrir diferencias en los genomas de las variedades vegetales (Gopaulchan et al., 2019). Sin embargo, en varios países del tercer mundo, donde se produce el cultivo de cacao, el genotipado de alto rendimiento todavía no es económicamente factible (Da Silva et al., 2016). Estas secuencias / regiones polimórficas podrían ser adecuadas para futuros estudios evolutivos y filogenéticos del género *Theobroma* y la familia Malvaceae, tanto para resolver incertidumbres taxonómicas para distinguir taxones estrechamente relacionadas con el cultivo de cacao (Abdullah et al., 2020). Además, el diseño de los marcadores podría ser adecuado para resolver cultivares estrechamente compatibles genéticamente y taxones silvestres para programas de mejoramiento (Abdullah et al., 2020). De esta manera, las versiones recientes del genoma de *T. cacao*, representan un modelo nuevo y simple para estudiar los procesos evolutivos, genética estructural y funcional, bioquímica, y genómica comparativa para el árbol del cacao (Argout et al., 2011, 2017).

En el Perú, la región Amazonas es el séptimo productor de cacao fino de aroma, el cual posee características distintivas de aroma y sabor muy apreciadas en el mercado internacional (MINAGRI, 2018). En 2015, mediante la Ordenanza Regional N° 368, el Gobierno Regional de Amazonas estableció la denominación de origen del “Cacao Amazonas Perú” por sus características bromatológicas, las condiciones ambientales donde se cultiva y porque la región Amazonas forma parte del gran Centro de Diversidad Genética de la especie de cacao (El Peruano, 2015; MINAGRI, 2018). Actualmente, se conocen estudios de huella volátil (Valle-Epquín et al., 2020); ácidos grasos (Oliva-

Cruz, et al., 2021a), compuestos fenólicos, aromáticos y físico-químicos (Castro-Alayo et al., 2019; Ordoñez et al., 2020) y estudios fenotípicos (Oliva-Cruz y Maicelo, 2020; Oliva-Cruz et al., 2021b). Sin embargo, este cacao fino de aroma aún no ha sido caracterizado con el uso de herramientas moleculares (Ordoñez et al., 2020).

Recientemente, estudios preliminares en la identificación de genotipos del Cacao Fino de Aroma han permitido caracterizar cinco variedades genéticas de cacao en la Región Amazonas (Bustamante et al., 2022). Sin embargo, la información sobre la estructura y funcionalidad del genoma del cacao fino de aroma de Amazonas aún es desconocida. Esta caracterización genómica estructural y funcional permitirá entender la funcionalidad de los genes del cacao (Yang et al., 2014). La genómica estructural involucra la naturaleza física del genoma e incluye la secuenciación y el mapeo de los genomas; asimismo, la genómica funcional implica el estudio de la expresión y función del genoma (Wang et al., 2019). Por lo tanto, es importante estudiar la localización y función de los genes candidatos mediante enfoques de genómica funcional y estructural para conocer procesos biológicos de interacción, por ejemplo, entre el árbol de cacao y sus enfermedades tal como *Phytophthora* o moniliasis (Micheli et al., 2010; Argout et al., 2011). Asimismo, estas metodologías son claves para acelerar el desarrollo de cultivares promisorios mediante la aplicación de nuevos enfoques biotecnológicos (Fister et al., 2016).

De esta manera, el presente estudio busca decodificar los genomas plastidiales y mitocondriales del cacao fino de aroma (*T. cacao*) estableciendo la localización y función de sus respectivos genes. Asimismo, los ensamblajes de estos genomas permitirán determinar las relaciones evolutivas entre las distintas variedades de cacao. Este estudio también busca determinar las implicancias de las investigaciones del cacao fino de aroma bajo un enfoque de producción sostenible en la región Amazonas.

II. MATERIALES Y MÉTODOS

2.1. Área de estudio

La región Amazonas fue el centro de investigación, específicamente las muestras de cacao fueron colectadas en la provincia de Bagua (420 m.s.n.m.) (Figura 1).

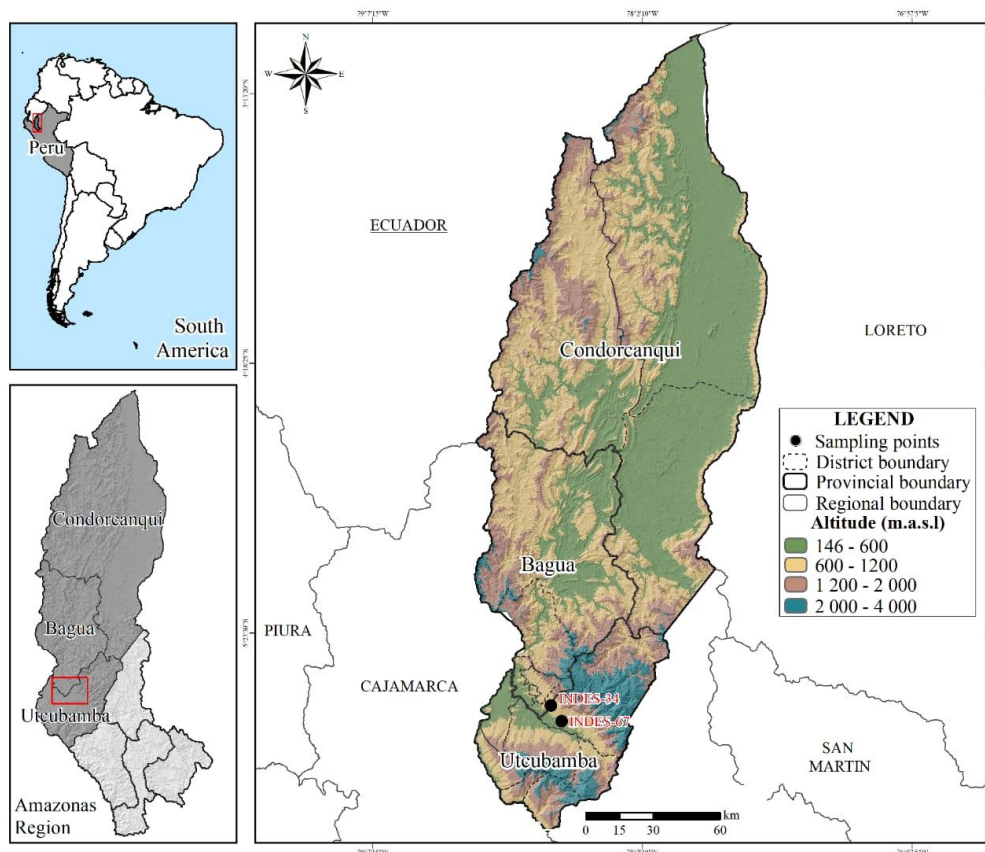


Figura 1. Mapa de localización del área de estudio donde se obtuvo el material vegetal de Cacao Fino de Aroma de Amazonas.

2.2. Colecta de muestras de cacao fino de aroma

El Servicio Nacional Forestal y de Fauna Silvestre (SERFOR) otorgó un permiso de investigación científica en flora silvestre para la colecta de Cacao Fino de Aroma (MIDAGRI-SERFOR-DGGSPFFS, con código de autorización N° AUT-IFL-2020-0051). Se colectaron muestras de dos genotipos enmarcados en “cacao fino de aroma de Amazonas”, identificados por Bustamante et al. (2022), provenientes de la provincia de Bagua (INDES34 = 5°40'52.1688", 78°24'13.9934" y INDES67 = 5°44'36.2040", 78°21'37.8225") en la Región Amazonas (Figura 1). Muestras de tejido de

aproximadamente 50 mm² de las hojas más tiernas de cacao fueron tomadas para los análisis moleculares y se colocaron en tubos Eppendorf Safelock de 1,5 ml preetiquetados. Estas muestras también se depositaron en el herbario de la Universidad Nacional Toribio Rodríguez de Mendoza (KUELAP) y asignados a los códigos: KUELAP-638 (INDES34) y KUELAP-659 (INDES67) (Thiers, 2016). Para cada sitio de colección, se registró la fecha, la hora y las coordenadas GPS. Se tomaron fotografías para registrar los lugares de muestreo y las características del lugar (Figura 2).

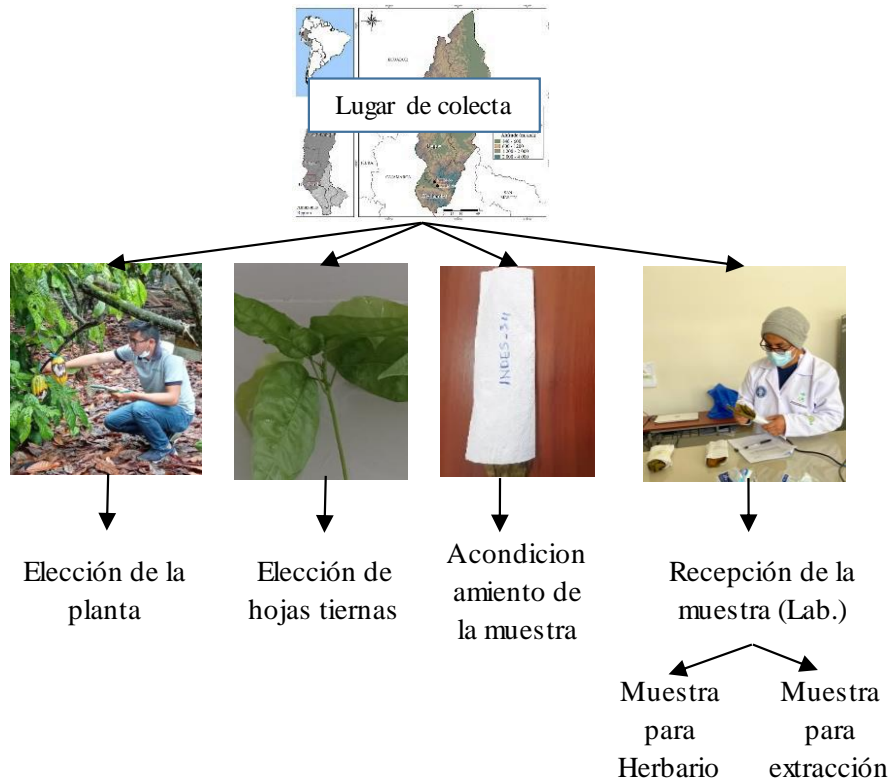


Figura 2. Colecta de muestras del Cacao Fino de Aroma.

2.3. Análisis genómico del cacao fino de aroma

2.3.1. Extracción de ADN genómico

La extracción de ADN se realizó en el laboratorio de Biología Molecular y Genómica de la Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas. El ADN genómico total fue extraído usando el kit NucleoSpin (Macherey-Nagel, Düren, Alemania) según el siguiente protocolo:

- Se seleccionó 50 mm² de tejido vegetal (hojas jóvenes de cacao). Las mismas que fueron sumergidas en nitrógeno líquido y pulverizadas en un disruptor manual.
- A la muestra pulverizada se agregó 400 µL del Buffer PL1 y 10 µL de RNasa a cada muestra, se agitó durante 15 a 30 segundos a 1500 x g. y se incubó las muestras a 65 °C durante 60 minutos.
- Luego se centrifugó las muestras durante 1 minuto a 11,000 x g.
- Se adicionó el lisado en la columna con anillo violeta y se centrifugó durante 2 minutos a 11,000 x g.
- Luego se agregó 450 µL de Buffer PC y mezclar bien, pipeteando 5 veces
- Se colocó una columna NucleoSpin Plant II (anillo verde) en un nuevo tubo de recogida (2 mL), y se adicionó un máximo de 700 µL de la muestra y posteriormente se centrifugó durante 1 minuto a 11,000 x g, se desechó el flujo.
- 1er Lavado: Se Agregó 400 µL Buffer PW1 a la NucleoSpin Plant II Columna, se centrifugó durante 1 minuto a 11,000 x g y se desechó el flujo.
- 2do Lavado: Se Agregó 700 µL Buffer PW2 a la NucleoSpin Plant II Columna se centrifugó durante 1 minuto a 11,000 x g y se desechó el flujo.
- 3er Lavado: Se Agregó 200 µL más de Buffer PW2 a la Columna NucleoSpin Plant II y se centrifugó durante 2 minutos a 11,000 x g.
- Elución: Se colocó la columna NucleoSpin Plant II en un tubo nuevo de microcentrífuga de 1,5 mL. Se añadió 50 µL de Buffer PE (65 °C) en la membrana, incubando la columna durante 5 minutos a 65 °C y se centrifugó durante 3 minutos a 145,000 x g, luego eludir el ADN. Este paso se repitió dos veces hasta completar un volumen total de 100 µL.

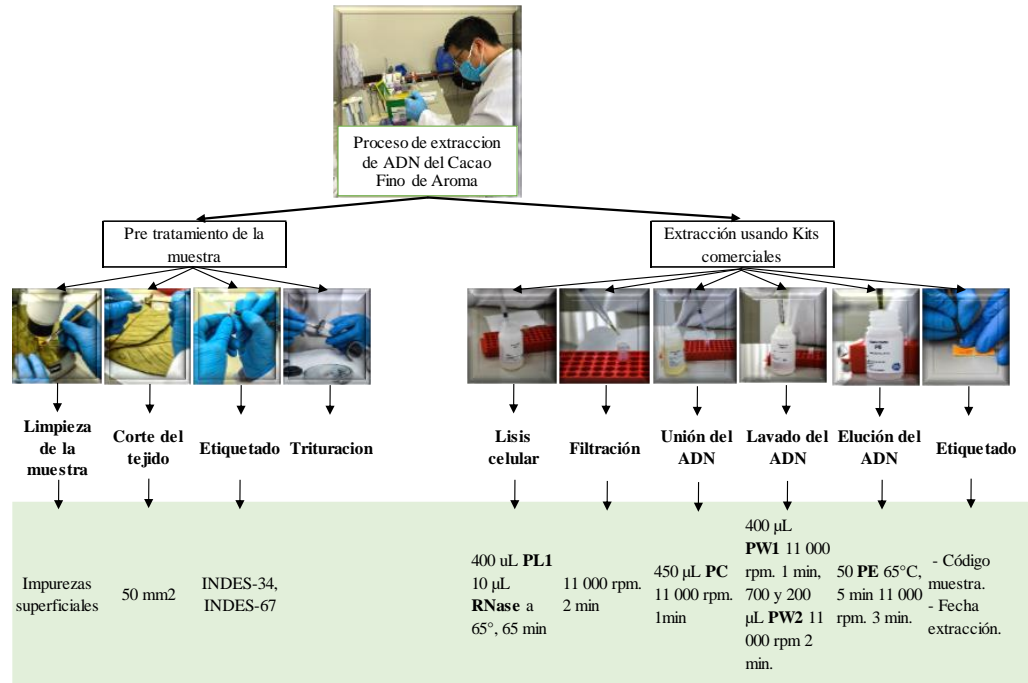


Figura 3. Proceso de extracción de ADN genómico del Cacao Fino de Aroma.

2.3.2. Cuantificación de ADN genómico extraído

Una vez extraído el ADN, se procedió a cuantificar la concentración de ADN usando un Qubit 3 (Applied Scientific, Estados Unidos) y el kit de cuantificación Qubit™ dsDNA BR Assay Kit (cat. Q32850) de la siguiente manera:

- Inicialmente el equipo fue calibrado. Seguidamente se utilizó tubos de 0.6 mL para estándares y muestras. El Qubit® y el dsDNA BR requiere 2 estándares.
- Se preparó la solución de trabajo Qubit® diluyendo el reactivo Qubit® dsDNA BR. Se añadió 199 μ L de la solución de trabajo Qubit® a los tubos de ensayo individuales y 1 μ L de ADN genómico, obteniendo un volumen final de 200 μ L.
- Luego se homogenizó agitando en un Vórtex por un lapso de 2–3 segundos.
- Las muestras se incubaron a temperatura ambiente durante 5 minutos.
- Finalmente, cada muestra fue leída en la cámara de muestra de forma individual, y anotados las lecturas de cada muestra en ng/ μ L.

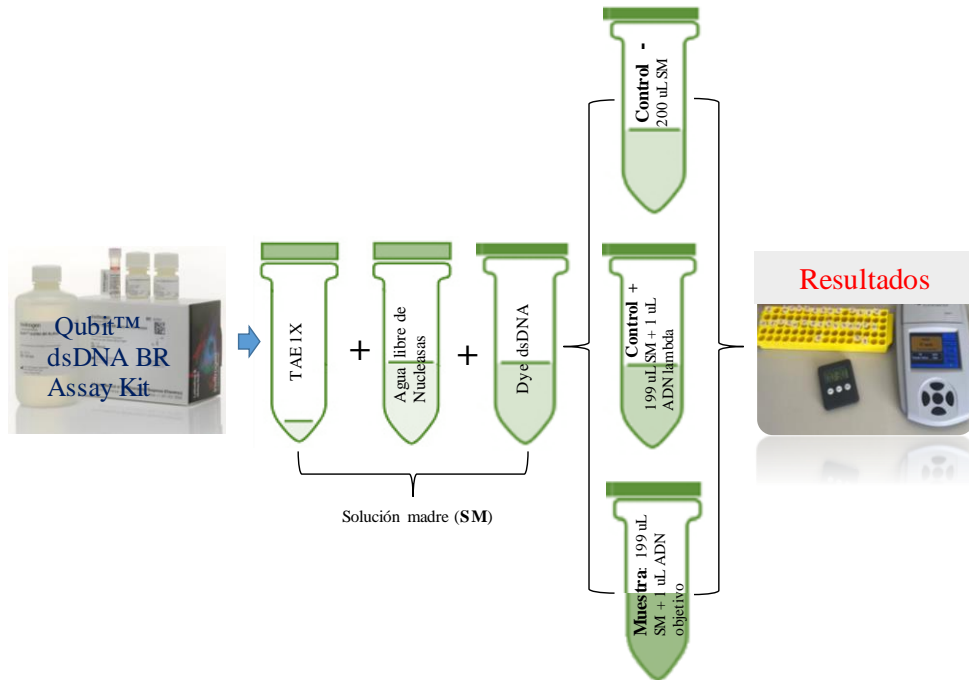


Figura 4. Proceso de Cuantificación de ADN genómico del Cacao Fino de Aroma.

2.3.3. Preparación de ADN genómico y secuenciación masiva

El ADN genómico fue enviado hacia myGenomics, LLC (Alpharetta, GA EE. UU.) para su secuenciación comercial. Brevemente, el ADN genómico fue sometido a mediciones de densidad óptica en NanoDrop y Qubit (ThermoFisher Scientific, Waltham, MA EE. UU.). La calidad del ADN fue evaluada en electroforesis en gel de agarosa y Agilent TapeStation para confirmar la pureza y la concentración antes de la preparación de las librerías. El ADN genómico fue fragmentado y ligado a adaptadores únicos utilizando el kit de preparación de la biblioteca Swift 2S Turbo DNA (Swift Bioscience, Ann Harbor, MI, EE. UU.). La librería resultante fue probada para la distribución del tamaño y la concentración usando Qubit y tapestation. Las librerías se secuenciaron en NextSeq 500 (Illumina, San Diego, CA) de acuerdo con la operación estándar. Se generaron reads pareadas de 150 nucleótidos (nt) y se verificó la calidad de los datos utilizando FastQC (Babraham Institute, Cambridge, Reino Unido).

2.3.4. Ensamblaje del ADN genómico secuenciado

Los genomas plastidiales fueron ensamblados utilizando la función Map to Reference con la configuración predeterminada (Custom / Sensitivity, con 10 interacciones) en Geneious Prime (Biomatters, Ltd, Auckland, Nueva Zelanda) utilizando una secuencia de referencia de *T. cacao* (accesión de GenBank HQ336404) (Jansen et al., 2011). Los contigs generados para las muestras de Cacao Fino de Aroma (INDES34, INDES67) se unieron al superponer regiones de solapamiento (Argout et al., 2017). La exactitud y circularidad de los genomas se confirmaron mediante el mapeo de los reads y contigs utilizando la misma herramienta de mapeo a la referencia en Geneious Prime, v. 2020.0.3 (www.geneious.com). Asimismo, los genomas mitocondriales fueron ensamblados utilizando la función Map to Reference con la configuración predeterminada (Custom / Sensitivity, con 10 interacciones) en Geneious Prime utilizando *Gossypium hirsutum* L, como secuencia de referencia (accesión de GenBank NC_028254 (Tang et al., 2015).

2.3.5. Anotación de genes plastidiales

Los genomas plastidiales y mitocondriales fueron anotados usando MFannot (Beck & Lang 2010), Geneious Prime, NCBI ORFfinder, y tRNAscan-SE 1.21 (Schattner et al., 2005). Asimismo, se usó los softwares RNAmmer 1.2 Server y CPGAVA2 para corroborar la anotación.

2.4. Análisis de las relaciones evolutivas del cacao fino de aroma

2.4.1. Análisis comparativo de genomas plastidiales

La identidad de las secuencias entre los genomas plastidiales de Cacao Fino de Aroma fueron comparadas utilizando el software mVISTA (Frazer et al., 2004) con *T. cacao* (HQ336404) como referencia.

2.4.2. Análisis filogenéticos de los genomas plastidiales y mitocondriales

Los genomas plastidiales fueron alineados con otros genomas de cacao descargados de la base de datos (Genbank) utilizando MAFFT (Katoh y Standley,

2013). Los análisis filogenéticos se ejecutaron con RAxML-NG (Kozlov et al., 2019) usando el modelo GTR + gamma y 1000 réplicas. Los árboles filogenéticos se visualizaron con TreeDyn 198.3 en Phylogeny.fr (Dereeper et al., 2008).

Tabla 1. Lista de voucher de genomas plastidiales de especies de *Theobroma* utilizadas en los análisis filogenómicos.

Species	Voucher	Código de colecta	País
<i>T. cacao</i>	KUELAP-638	INDES34	Perú
<i>T. cacao</i>	KUELAP-659	INDES67	Perú
<i>T. cacao</i>	JQ228387	-	Trinidad and Tobago
<i>T. cacao</i>	JQ228386	-	Trinidad and Tobago
<i>T. cacao</i>	JQ228381	-	Trinidad and Tobago
<i>T. cacao</i>	KY085907	-	USA
<i>T. cacao</i>	HQ244500	-	Perú
<i>T. cacao</i>	JQ228380	-	Ghana
<i>T. cacao</i>	JQ228389	-	Puerto Rico
<i>T. cacao</i>	JQ228383	-	Trinidad and Tobago
<i>T. cacao</i>	JQ228382	-	Perú
<i>T. cacao</i>	JQ228385	-	Suriname
<i>T. cacao</i>	JQ228379	-	USA
<i>T. cacao</i>	HQ336404	-	USA
<i>T. grandiflorum</i>	NC_054233	-	China
<i>T. grandiflorum</i>	JQ228388	-	Puerto Rico

2.5. Implicancia de los análisis moleculares en la producción sostenible

Se recopilaron artículos (indexados en SciELO, Scopus, Web of Science publicados en temas de Cacao Fino de Aroma utilizando palabras clave referidas a "*Theobroma*", "cacao" a lo largo del país. La base de datos fue agrupada en el componente regional "Ambiente y Desastre Naturales" que está relacionado con la producción de cacao y el desarrollo sostenible dentro del Plan Regional de Desarrollo Concertado (PRDC) (GOREA, 2014). Se realizaron matrices de tabulación cruzada y gráficos para analizar la relación entre el PRDC y las investigaciones relacionados a cacao. Finalmente, se construyeron mapas de red utilizando el método de "recuento completo" del software VOSviewer v.1.6.17 (Van Eck & Waltman 2010), para explorar la co-ocurrencia (palabras en los títulos de los artículos) de las investigaciones sobre genética y genómica en cacao fino de aroma de Amazonas.

III. RESULTADOS

3.1. Genomas plastidiales del cacao fino de aroma

3.1.1. Secuenciación y ensamblaje de genomas plastidiales

Se generaron un total de 2.24×10^9 reads de 150 bp luego del secuenciamiento Illumina de extremo único para cada una de las muestras INDES34 e INDES67. De cada muestra se produjo una media de 160 Mb de secuencia de alta calidad. La secuenciación Illumina de 151 pb de ADN plastidial ha generado 2,998,427 - 2,428,060 de lecturas limpias para las dos muestras de *Theobroma* muestreadas (Tabla 2).

Tabla 2. Tamaño de los reads limpios de muestra de Cacao Fino de Aroma de la Región Amazonas secuenciada por Illumina.

Especie	Código/Voucher	Localidad	Reads limpios
<i>T. cacao</i>	INDES34	Lluhuana	2,998,427
<i>T. cacao</i>	INDES67	Naranjos Altos	2,428,060

Luego del ensamblaje *de novo* y usando genoma de referencia se obtuvieron cuatro genomas plastidiales completos, dos genomas de la muestra INDES34 y dos genomas de la muestra INDES67. Estos genomas fueron etiquetados como INDES34_6a, INDES34_6b, INDES67_3a e INDES67_3b (Tabla 3). Las secuencias de nucleótidos determinadas de los cuatro genomas plastidiales contienen el mismo porcentaje de GC (36.9). La longitud de los genomas varía desde 160,589 pb hasta 160,620 pb, con una diferencia máxima de 16 pb respecto al genoma de referencia (HQ336404) (Tabla 3). Todos tienen una estructura genómica similar a la mayoría de las angiospermas, y los mapas genómicos plastidiales se presentan en las Figuras 5-8.

Tabla 3. Tamaño de los genomas plastidiales del Cacao Fino de Aroma de la Región Amazonas.

Especie	Voucher	Código	Longitud (pb)	Diferencia tamaño (pb)	% GC
<i>T. cacao</i>	HQ336404 (ref)	-	160,604		36.9
<i>T. cacao</i>	KUELAP-638	INDES34_6a	160,619	+15	36.9
		INDES34_6b	160,620	+16	36.9
<i>T. cacao</i>	KUELAP-659	INDES67_3a	160,617	+13	36.9
		INDES67_3b	160,589	+15	36.9

3.1.2. Anotación de genes de los genomas plastidiales

Los cuatro genomas plastidiales mostraron el mismo contenido génico con 130 genes que incluyen 37 tRNA (ARN de transferencia), 8 rRNA (ARN ribosomales) y 85 genes codificantes de proteínas. Entre ellos, 17 genes fueron duplicados en la IR (Región invertida repetitiva) del genoma plastidial de cada muestra (Tabla 4) que comprendían seis genes codificantes de proteínas, 4 (rRNA) y 7 tRNA. El LSC (Copia Individual Larga), el IR y SSC (Copia Individual Corta) tuvieron longitudes similares (Tabla 4).

Tabla 4. Características de los genomas plastidiales de Cacao Fino de Aroma y la especie más cerca, *T. grandiflorum*.

Características	<i>T. cacao</i> (INDES 34_6a)	<i>T. cacao</i> (INDES34 _6b)	<i>T. cacao</i> (INDES6 7_3a)	<i>T. cacao</i> (INDES6 7_3b)	<i>T. cacao</i> (HQ336 404)	<i>T. grandiflorum</i> (JQ228388)
Tamaño (pb)	160,619	160,620	160,617	160,589	160,604	160,619
Longitud del LSC (pb)	89,410	89,407	89,410	89,389	89,395	89,333
Longitud del SSC (pb)	21,872	21,876	21,884	21,877	20,187	20,194
Longitud del IR (pb)	24,688	24,688	24,661	24,661	25,511	25,546
Número de genes	130	130	130	130	130	130
Número de genes duplicados	17	17	17	17	17	17
Genes codificadores de proteínas	85	85	85	85	85	85

Contenido GC (%)	36.9	36.9	36.9	36.9	36.9	36.8
Número de tRNA	37	37	37	37	37	37
Número de rRNA	8	8	8	8	8	8
Número de proteínas codificantes (CDS % pb)	49.08	49.08	49.08	49.08	49.08	49.08

Los cuatro genomas plastidiales contienen el mismo número de genes (Tabla 4). Sin embargo, varían en cuanto a la posición de inicio y término. INDES34_6a, INDES67_3a e INDES67_3b son lo que más similitudes presentan en cuanto a las posiciones de los genes frente a INDES34_6b (Tabla 5). Asimismo, 22 genes están relacionados con el proceso fotosintético, 8 genes relacionados con del intercambio de protones y 18 genes están relacionados con el intercambio de electrones (Tabla 5).

Tabla 5. Comparación del tamaño de genes y su funcionalidad con el intercambio de electrones y fotosíntesis de los genomas plastidiales de Cacao Fino de Aroma.

Función	Gen	Producto	<i>T. cacao</i> (INDES34_6a)		<i>T. cacao</i> (INDES34_6b)		<i>T. cacao</i> (INDES67_3a)		<i>T. cacao</i> (INDES67_3b)		Tamaño (pb)
			Inicio	Fin	Inicio	Fin	Inicio	Fin	Inicio	Fin	
Genes relacionados al proceso fotosintético	<i>psbA</i>	photosystem II protein D1	1613	552	1613	552	1613	552	1613	552	1061
	<i>psbK</i>	photosystem II protein K	7380	7565	7379	7564	7380	7565	7383	7568	185
	<i>psbI</i>	photosystem II protein I	7908	8084	7907	8083	7908	8084	7911	8087	176
	<i>psbM</i>	photosystem II protein M	31053	30949	31053	30949	31053	30949	31056	30952	104
	<i>psbD</i>	photosystem II protein D2	34673	35734	34673	35734	34673	35734	34691	35752	1061
		photosystem II CP43 chlorophyll									
	<i>psbC</i>	apoprotein	35682	37103	35682	37103	35682	37103	35700	37121	1421
	<i>psbZ</i>	photosystem II protein Z	37799	37987	37799	37987	37799	37987	37817	38005	188
	<i>psaB</i>	photosystem I P700 apoprotein A2	41810	39606	41810	39606	41810	39606	41828	39624	2204
	<i>psaA</i>	photosystem I P700 apoprotein A1	44088	41836	44088	41836	44088	41836	44106	41854	2252
	<i>psaI</i>	photosystem I subunit VIII	63780	63893	63780	63893	63780	63893	63780	63893	113
	<i>ycf4</i>	photosystem I assembly protein ycf4	64294	64869	64294	64869	64294	64869	64294	64869	575
	<i>psbJ</i>	photosystem II protein J	68936	68814	68935	68813	68936	68814	68936	68814	122
	<i>psbL</i>	photosystem II protein L	69307	69083	69306	69082	69307	69083	69307	69083	224
		photosystem II cytochrome b559 beta subunit									
	<i>psbF</i>	photosystem II cytochrome b559 alpha subunit	69345	69226	69344	69225	69345	69226	69345	69226	119
	<i>psbE</i>	photosystem II cytochrome b559 alpha subunit	69606	69355	69605	69354	69606	69355	69606	69355	251
	<i>psaJ</i>	photosystem I subunit IX	71687	71815	71686	71814	71687	71815	71687	71815	128
		photosystem II CP47 chlorophyll									
	<i>psbB</i>	apoprotein	77436	78962	77434	78960	77436	78962	77415	78941	1526
<i>psbT</i>	photosystem II protein T	79145	79252	79143	79250	79145	79252	79124	79231	107	
<i>psbN</i>	photosystem II protein N	79444	79313	79442	79311	79444	79313	79423	79292	131	
<i>psbH</i>	photosystem II phosphoprotein	79550	79771	79548	79769	79550	79771	79529	79750	221	
<i>psaC</i>	photosystem I subunit VII	122908	122663	122906	122661	122908	122663	122880	122635	245	
	ribulose 1,5-bisphosphate										
<i>rbcL</i>	carboxylase/oxygenase large subunit	59370	60824	59371	60825	59370	60824	59370	60824	1454	
Genes relacionados al intercambio de	<i>petN</i>	cytochrome b6/f complex subunit VIII	29646	29792	29646	29792	29646	29792	29649	29795	146
	<i>petA</i>	cytochrome f	66744	67706	66744	67706	66744	67706	66744	67706	962
	<i>petL</i>	cytochrome b6/f complex subunit VI	70482	70577	70481	70576	70482	70577	70482	70577	95
	<i>petG</i>	cytochrome b6/f complex subunit V	70755	70868	70754	70867	70755	70868	70755	70868	113
	<i>petB</i>	cytochrome b6	79900	81359	79898	81357	79900	81359	79879	81338	1459
	<i>petD</i>	cytochrome b6/f complex subunit IV	81567	82811	81566	82808	81567	82811	81546	82790	1244

	<i>ccsA</i>	cytochrome c heme attachment protein	119735	120700	119734	120699	119735	120700	119707	120672	965
	<i>accD</i>	acetyl-CoA carboxylase carboxyltransferase beta subunit	61546	63024	61546	63024	61546	63024	61546	63024	1478
Genes relacionados al intercambio de electrones	<i>ndhJ</i>	NADH-plastoquinone oxidoreductase subunit J	52877	52401	52878	52402	52877	52401	52887	52411	476
	<i>ndhK</i>	NADH-plastoquinone oxidoreductase subunit K	53695	53018	53696	53019	53695	53018	53705	53028	677
	<i>ndhC</i>	NADH-plastoquinone oxidoreductase subunit 3	54193	53756	54194	53757	54193	53756	54203	53766	437
	<i>ndhB</i>	NADH-plastoquinone oxidoreductase subunit 2	102410	100195	102407	100192	102410	100195	102389	100174	2215
	<i>ndhF</i>	NADH-plastoquinone oxidoreductase subunit 5	117154	114923	117151	114920	117154	114923	117126	114895	2231
	<i>ndhD</i>	NADH-plastoquinone oxidoreductase subunit 4	122170	121019	122169	121018	122170	121019	122142	120991	1151
	<i>ndhE</i>	NADH-plastoquinone oxidoreductase subunit 4L	123460	123155	123458	123153	123460	123155	123432	123127	305
	<i>ndhG</i>	NADH-plastoquinone oxidoreductase subunit 6	124271	123699	124269	123697	124271	123699	124243	123671	572
	<i>ndhI</i>	NADH-plastoquinone oxidoreductase subunit I	125241	124678	125239	124676	125241	124678	125213	124650	563
	<i>ndhA</i>	NADH-plastoquinone oxidoreductase subunit 1	127500	125272	127498	125270	127500	125272	127472	125244	2228
	<i>ndhH</i>	NADH-plastoquinone oxidoreductase subunit 7	128683	127502	128681	127500	128683	127502	128655	127474	1181
	<i>ndhB</i>	NADH-plastoquinone oxidoreductase subunit 2	147620	149835	147625	149840	147618	149833	147590	149805	2215
	<i>atpA</i>	ATP synthase CF1 alpha subunit	11826	10306	11824	10304	11826	10306	11829	10309	1520
	<i>atpF</i>	ATP synthase CF0 subunit I	13273	11899	13271	11897	13273	11899	13276	11902	1374
	<i>atpH</i>	ATP synthase CF0 subunit III	14165	13872	14164	13871	14165	13872	14168	13875	293
	<i>atpI</i>	ATP synthase CF0 subunit IV	16052	15309	16052	15309	16052	15309	16055	15312	743
	<i>atpE</i>	ATP synthase CF1 epsilon subunit	56931	56530	56932	56531	56931	56530	56931	56530	401
<i>atpB</i>	ATP synthase CF1 beta subunit	58424	56928	58425	56929	58424	56928	58424	56928	1496	

Los 4 genomas plastidiales (INDES34_6a, INDES34 _6b, INDES67_3a INDES67_3b) presentaron una gran cantidad de deleciones e inserciones en las regiones no codificantes (intrones). INDES34 _6b e INDES67_3a presentan el mismo número de inserciones (18) y deleciones (20), pero en diferentes posiciones a lo largo del genoma. INDES34_6a y HQ336404 (Referencia), también presentan el mismo número de inserciones (19) y deleciones (19) en diferentes posiciones. INDES67_3b presenta el mayor número de deleciones (24) y la menor cantidad de inserciones (14) (Tabla 6). Adicionalmente, solo se observó un evento de inserción/delección (± 1 pb) en el intrón *trnK* (posición 4491-1908*) que codifica el marco de lectura abierto (ORF) *matK* (Tabla 6).

Tabla 6. Número de deleciones e inserciones a lo largo de los 4 genomas plastidiales del Cacao Fino de Aroma.

Posición de nucleótidos	INDES 34_6a	INDES 34_6b	INDES67_3a	INDES67_3b	<i>T. cacao</i> (HQ336404)
4676-4684	-9	-9	-9	-7	-9
5043-5044	1	-1	1	1	1
9757-9759	1	-1	1	1	1
4491-1908 *	-1	1	-1	-1	1
13453-13455	-2	-2	-2	-2	2
13511-13515	5	5	5	5	-5
13604-13606	-1	1	-1	-1	1
14698-14701	-1	1	-1	-1	1
17231-17233	-1	1	-1	-1	1
28584-28585	1	-1	1	1	-1
33374-33388	-15	-15	-15	15	15
38291-38297	-7	-7	-7	-7	7
39234-39239	6	6	6	6	-6
46931-46933	1	1	1	1	-1
47241-47243	-1	1	-1	-1	1
52324-52331	8	8	8	-8	-6
55348-55357	10	10	10	-10	-10
59122-59124	-1	-1	-1	-1	1
61294-61296	1	-1	1	1	1
68041-68043	1	-1	1	1	-1
68411-68415	-5	-5	-5	-5	5
72015-72025	11	11	11	-11	-11
72255-72264	10	10	10	-10	-10
75710-75712	1	-1	1	1	-1
81543-81545	-1	1	-1	-1	1
81903-81905	1	-1	1	1	1
82306-82308	1	-1	1	1	-1
114198-114205	7	7	7	-7	-7
117490-117510	-1	-1	-1	-1	-2
117827-117830	-1	1	-1	-1	1

118123-118135	-1	-1	-1	-1	1
118655-118668	-1	1	-1	-1	-1
119525-119542	-5	-5	-5	-5	5
122697-122702	1	-1	1	1	1
125048-127636	1	1	1	1	-1
129307-129315	-3	-2	-3	-3	-1
131260-131270	-6	-4	-6	-6	-6
136100-136110	2	2	-2	-2	2
N° de deleciones	19	20	20	24	19
N° de inserciones	19	18	18	14	19

(-): eliminación de nucleótidos o fragmentos de ADN a lo largo del genoma. (*): inserción/delección en los exones o genes.

3.1.3. Generación de mapas genómicos plastidiales

Todos los genes anotados fueron representados en un mapa genómico. Cada mapa genómico plastidial fue generado con GeSeq (Tillich et al., 2017). El mapa genómico reveló que todos los genes se transcriben en el sentido horario. Además, se muestra la densa disposición de los genes y la gran región intergénica única (LSC).

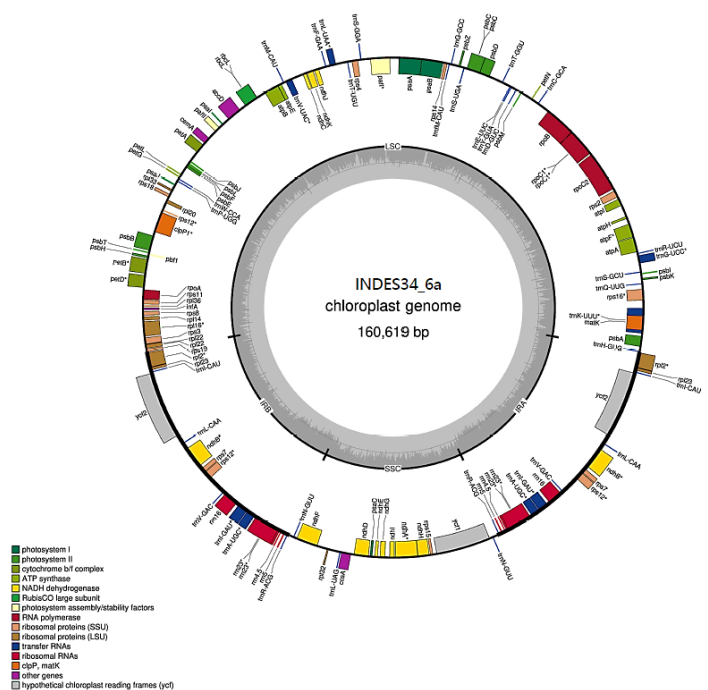


Figura 5. Mapa del genoma plastidial del Cacao Fino de Aroma (INDES34_6a). Los genes codificantes de proteínas y del ARN ribosómico y de transferencia están etiquetados en el mapa. Los genes están codificados por colores según las categorías funcionales.

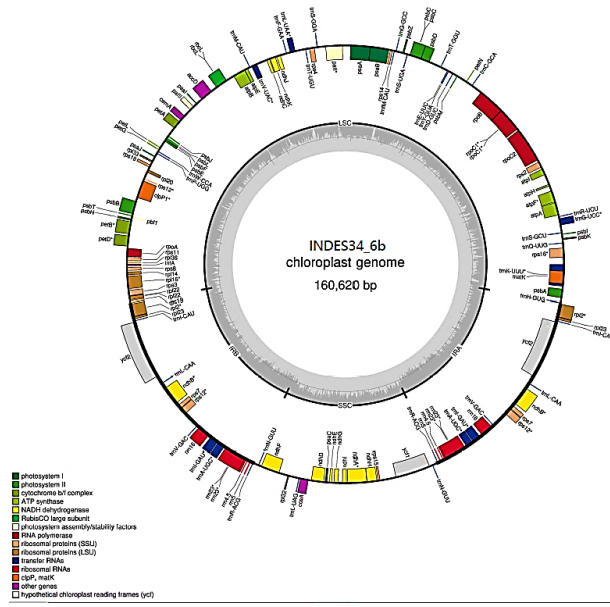


Figura 6. Mapa del genoma plastidial del Cacao Fino de Aroma (INDES34_6b). Los genes codificantes de proteínas y del ARN ribosómico y de transferencia están etiquetados en el mapa. Los genes están codificados por colores según las categorías funcionales.

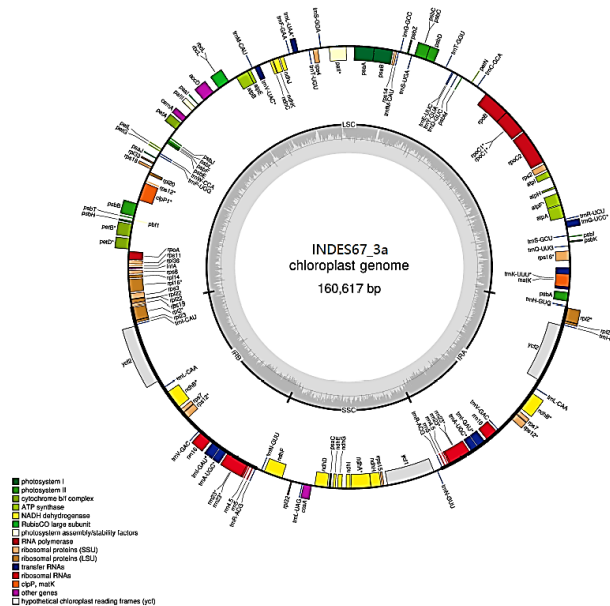


Figura 7. Mapa del genoma plastidial del Cacao Fino de Aroma (INDES67_3a). Los genes codificantes de proteínas y del ARN ribosómico y de transferencia están etiquetados en el mapa. Los genes están codificados por colores según las categorías funcionales.

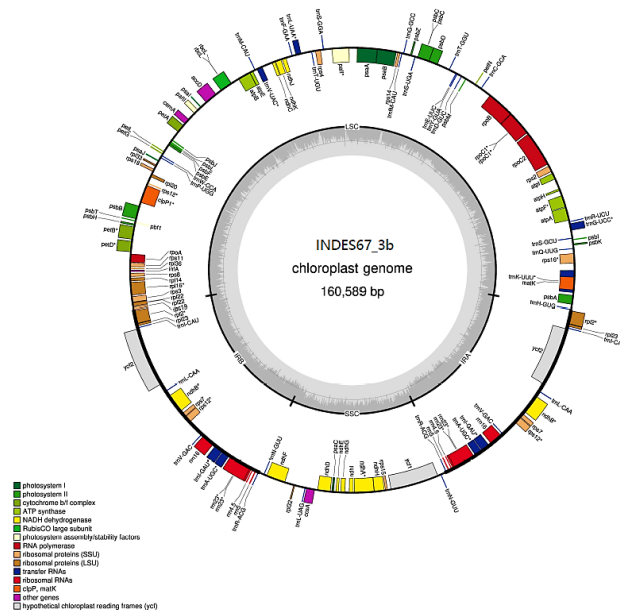


Figura 8. Mapa del genoma plastidial del Cacao Fino de Aroma (INDES67_3b). Los genes codificantes de proteínas y del ARN ribosómico y de transferencia están etiquetados en el mapa. Los genes están codificados por colores según las categorías funcionales.

3.2. Genomas mitocondriales del cacao fino de aroma

Los genomas mitocondriales de Cacao Fino de Aroma fueron ensamblados parcialmente (~20%). Estos ensamblajes parciales fueron interrumpidos por más de 150 gaps (regiones vacías que no pudieron ser completados) usando el genoma de referencia (NC_027406 = *Gossypium hirsutum*). La complejidad del proceso de ensamblaje de genomas mitocondriales de plantas, la gran cantidad y variación del tamaño de los gaps fue desde 10 hasta 70,000 pares de bases (Figura 9). Esto confirma la no circularidad del genoma mitocondrial de cacao. Estos genomas mitocondriales parciales no pueden ser utilizados en los siguientes análisis filogenéticos y evolutivos.

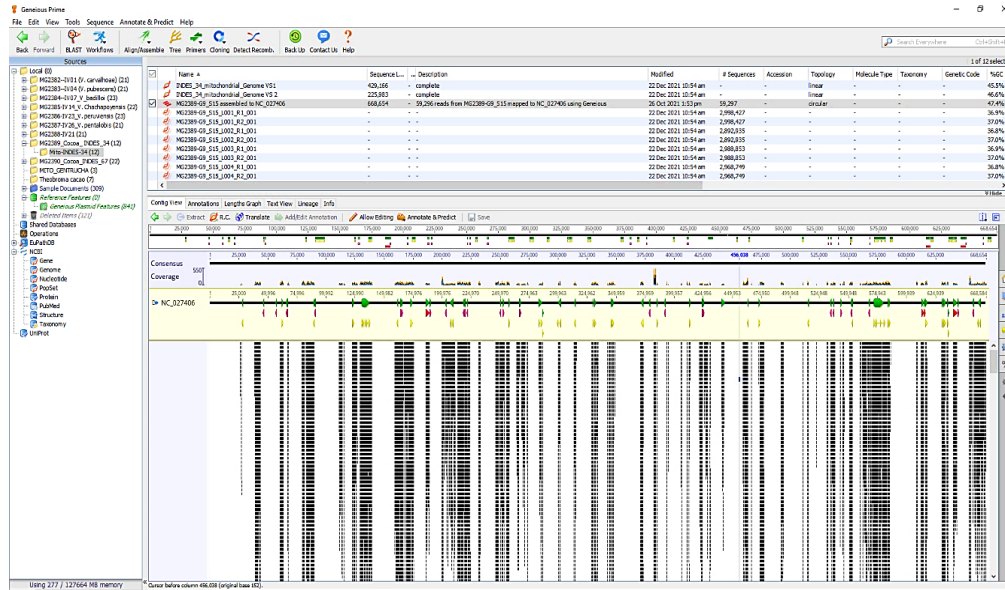


Figura 9. Genoma mitocondrial parcial del Cacao Fino de Aroma de Amazonas. **Reads** (barras en color negro) cubriendo zonas parciales del genoma de referencia flanqueado por gaps (flechas).

3.3. Relaciones evolutivas del cacao fino de aroma

3.3.1. Análisis comparativo de genomas plastidiales

El alineamiento de todos los genomas plastidiales muestra una alta similitud (mayor al 50% de identidad) de secuencia total (Figura 9). La divergencia total de secuencias estimada por la p-distancia entre los cuatro genomas plastidiales fue de 0,0028. La p-distancia por pares entre las cuatro especies fue entre 0,000006 y 0,00028. Los genomas plastidiales de Cacao Fino de Aroma (INDES34_6a, INDES34_6b, INDES67_3a e INDES67_3b) comparten genes idénticos codificantes en un orden similar. La región SSC de estos genomas plastidiales están en la misma orientación en relación con *T. grandiflorum*. Además, los cuatro genomas tienen pequeñas variaciones en el límite IR, LSC y SSC. Las muestras INDES34_6a e INDES34_6b tienen el mismo límite de IR (24,688 pb) pero varían en 4 pb para la región SSC y 3 pb en LSC. Por otro lado, INDES67_3a e INDES67_3b también comparten el mismo límite de IR (24,661 pb), pero con una variación de 7 pb en la región de SSC y 21 pb en LSC. Asimismo, INDES67_3a e INDES34_6a comparten el mismo límite en LSC (89,410 pb). Los cuatro genomas plastidiales obtenidos tienen el mismo número

de genes, tRNA, rRNA, pero no comparten el mismo límite de las regiones IR, LSC y SSC, tienen una diferencia aproximada de 823 pb en IR, 1690 pb en SSC y 77 pb en LSC con relación a *T. cacao* (HQ336404) y *T. grandiflorum* (JQ228388) (Tabla 4).

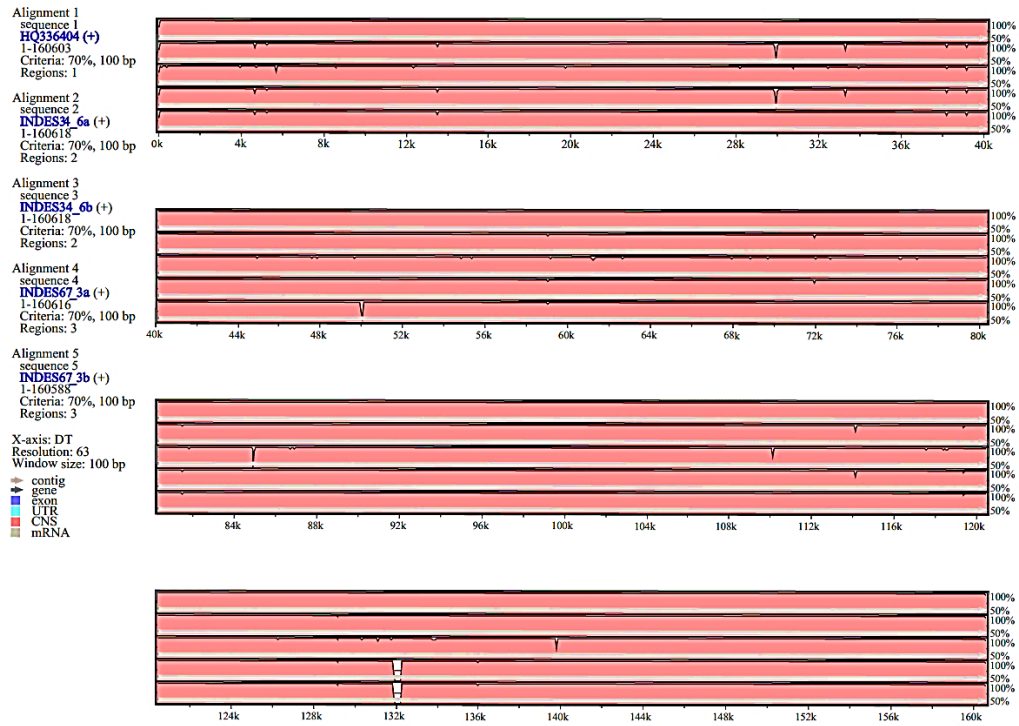


Figura 10. Gráfico de identidad porcentual obtenido de mVISTA para los cuatro genomas plastidiales de Cacao Fino de Aroma (INDES34_6a, INDES34_6b, INDES67_3a, INDES67_3b) y *T. cacao* (HQ336404) como referencia. La escala vertical indica el porcentaje de identidad entre 50% al 100%.

3.3.2. Análisis filogenómicos

La matriz de datos para los análisis filogenómicos contuvo genomas plastidiales completos, incluyendo 12 genomas de *T. cacao* y dos genomas de *T. grandiflorum* como grupo externo. El conjunto de datos genómicos estuvo compuesto de 160,799 posiciones de nucleótidos. Los análisis ML de los 14 genomas plastidiales dieron como resultado un árbol con valores de confianza relativamente altos en cada una de las ramas (Figura 11). El genoma plastidial INDES34_6b resultó ser una variedad hermana a otras provenientes de Perú

(JQ228382, HQ244500), Trinidad y Tobago (JQ228383) y USA (KY085907). Sin embargo, el genoma plastidial INDES34_6a que proviene de la de la misma planta de cacao (INDES34) resultó ser la misma variedad que INDES67 (INDES67_3a, INDES67_3b), la cual es una variedad hermana a otros genomas de cacao provenientes de Suriname (JQ228385), Trinidad y Tobago (JQ228386, JQ228387) y USA (JQ228379) (Figura 11).

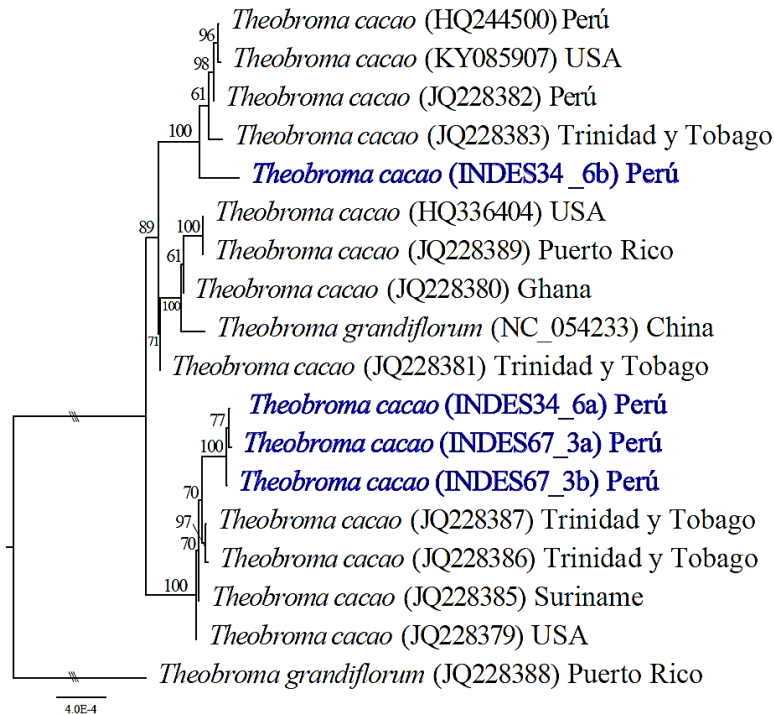


Figura 11. Filogenia de máxima verosimilitud de variedades de la especie *Theobroma cacao* basadas en genomas plastidiales completos. Los números asociados a los nodos son los valores de soporte bootstrap. La escala indica el número de sustituciones de nucleótidos por sitio.

3.3. Implicancia de las investigaciones del cacao en la producción sostenible

Se identificaron artículos relacionados con cacao. Así, 143 artículos indexados en SciELO, Scopus y Web of Science relacionados a estudios en cacao peruano fueron incluidos para el análisis bibliométrico. Los resultados muestran que la mayor parte de estudios se han realizado en el área de “bioquímica” (componentes bioactivos, metales pesados, análisis físicos-químicos del chocolate), seguido del área de “genética” (diversidad genética del cacao, germoplasma, citogenética), asimismo

tenemos el área de “manejo agronómico” que incluye los sistemas de siembra, y está más relacionado con el área de “fitopatología” y “entomología”. Finalmente, tenemos el área de “genómica”, donde no se han desarrollado muchas investigaciones tanto a nivel nacional como en la región Amazonas (Figura 12).

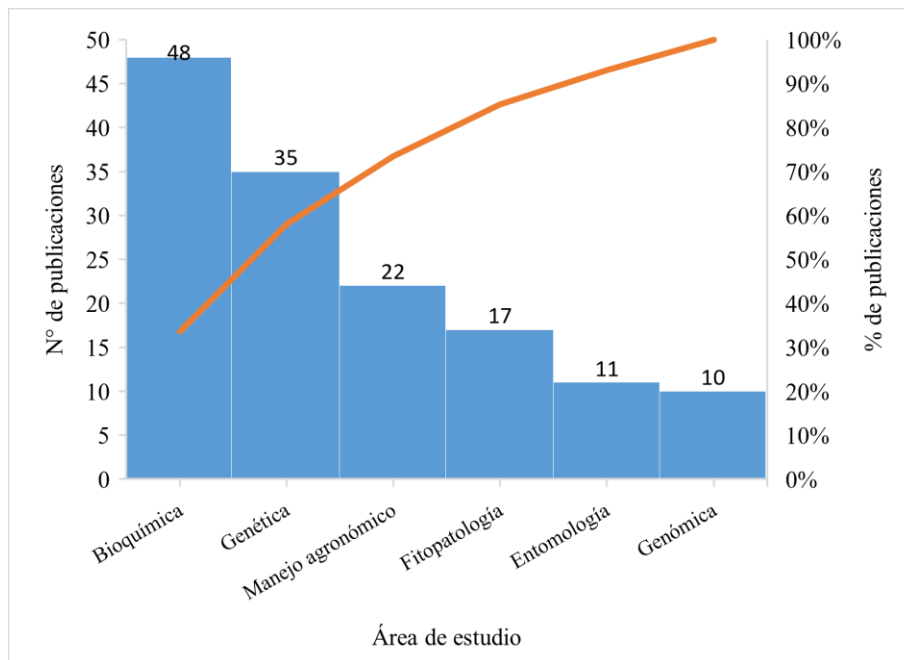


Figura 12. Número de artículos publicados en relación al cacao peruano, según el área de estudio.

La figura 13 muestra una tendencia de crecimiento entre los años 1981 al 2022, con una tasa de crecimiento anual promedio de 11.44 %. Las primeras publicaciones referente a cacao en Perú se realizaron en 1981. Este crecimiento no ha sido ascendente durante los siguientes años. Recién a partir del 2015, el crecimiento ha sido más significativo, siendo los años 2020 y 2021, los años con el mayor número de investigaciones en cacao (21 publicaciones) (Línea azul, Figura 13). Adicionalmente, en la región Amazonas se ha realizado un total de 31 investigaciones relacionados a Cacao Fino de Aroma, siendo los años 2020 y 2021 con la mayor investigación en Cacao Fino de Aroma (línea naranja, Figura 13).

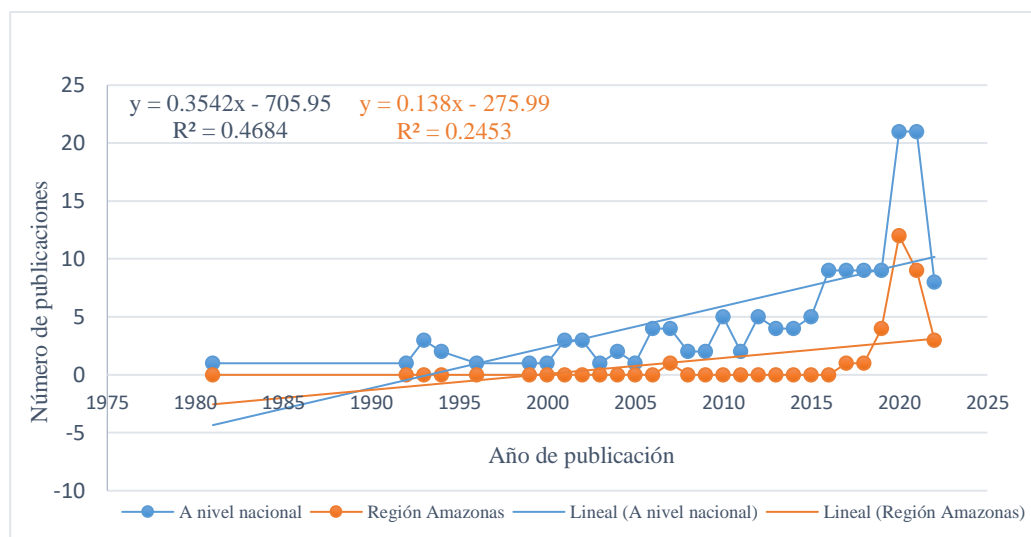


Figura 13. Número de publicaciones relacionados a Cacao Fino de Aroma que se han realizado en todo el Perú, y en la región Amazonas durante los años 1981-mayo del 2022.

El análisis de publicaciones relacionados a cacao de la Región Amazonas en relación a la base de datos muestra que el 98.5 % han sido publicadas en la base de datos Scopus, 7 % en SciELO y 3.5 % en WoS (Tabla 8). De las 143 publicaciones referente a cacao, el 91.6 % han sido publicadas en el idioma inglés y solo el 8.4 % en español (Tabla 8).

Tabla 7. Porcentaje de publicaciones según el tipo de base de datos e idioma de publicación.

Base de datos	%	Idioma de publicación	%
Scopus	89.5	Inglés	91.6
SciELO	7.0	Español	8.4
WoS	3.5		

Del total de investigaciones relacionadas al Cacao Fino de Aroma desarrolladas en Perú, nuestro país ha desarrollado 59 publicaciones de 143, y otras investigaciones han sido liderados por investigadores internacionales, muchos de ellos con y sin ninguna colaboración con instituciones peruanas (Figura 14). Aquí destaca Estados Unidos, con 40 investigaciones, seguido de Francia, Brazil, Trinidad y Tobago con

21, 16 y 14 publicaciones, respetivamente. Según la coloración de la red del mapa y la escala, se muestra que la mayoría de investigaciones han sido lideradas por instituciones internacionales durante los años 1981 hasta el 2015; sin embargo, durante los últimos 7 años, nuestro país ha liderado en investigaciones relacionadas a Cacao (Figura 14).

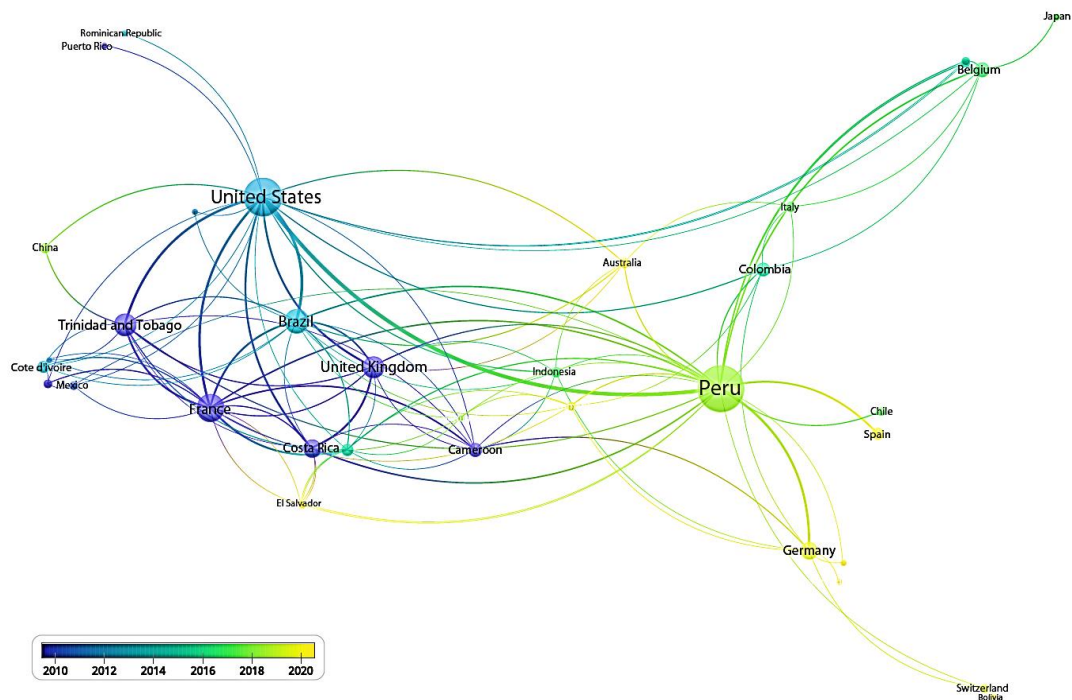


Figura 14. Mapa de países con mayor número de investigaciones relacionadas al Cacao de Perú.

A lo largo de los últimos ~41 años, 22 instituciones peruanas han realizado investigaciones relacionadas a cacao. Las que más destacan son la UNTRM (18 de 143 publicaciones), seguido por la Universidad Nacional Agraria de la Selva (UNAS) con 9 publicaciones, y el Instituto de Cultivos Tropicales (ICT) con 7 publicaciones. Estas dos últimas instituciones están ubicadas en regiones como Loreto y San Martín, respectivamente, ambas vecinas a la región Amazonas. Adicionalmente, 39 instituciones extranjeras han colaborado con entidades peruanas para realizar investigaciones en Cacao peruano, entre la que más destaca es United States Department of Agriculture (USDA) con 30 publicaciones. Según la coloración de la red del mapa y la escala, hasta el año 2018 se observa que la mayoría de

investigaciones han sido lideradas por instituciones internacionales, sin embargo, durante los últimos 3 años, la UNTRM ha desarrollado un número mayor de investigaciones en Cacao Fino de Aroma (Figura 15).

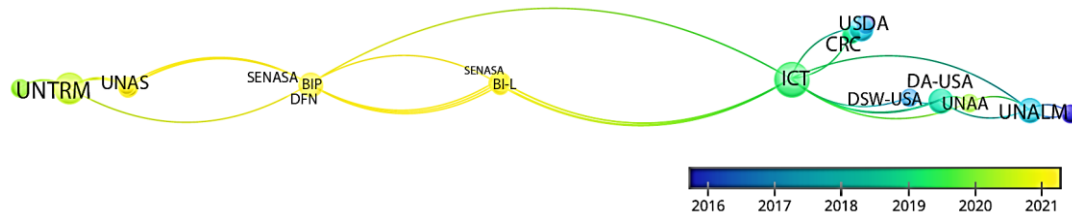


Figura 15. Mapa de instituciones nacionales e internacionales con mayor número de investigaciones relacionados al cacao.

En los títulos de las publicaciones revisadas, se identificaron un total de 905 palabras, de las cuales 206 aparecían más de dos veces. La palabra “*Theobroma cacao*” fue incluida en el análisis; sin embargo, fue excluida para la interpretación por ser el tema central del estudio. Por lo tanto, las palabras con mayor aparición en las publicaciones fueron "genetics" y "genotype" con 16 y 13 ocurrencias respectivamente (Figura 16). Otras palabras con menos ocurrencias fueron "agroforestry", "*Phytophthora palvivora*", "genetic variation", "germplasm", con 10, 10, 9 y 9 ocurrencias, respectivamente. Las palabras "chocolate", "antioxidants", "poliphenols" aparecen con 6 ocurrencias. Sin embargo, las palabras "genome cacao", "plant DNA", aparecen con 3 ocurrencias (Figura 16), una cantidad baja que representa alrededor de 10 publicaciones.

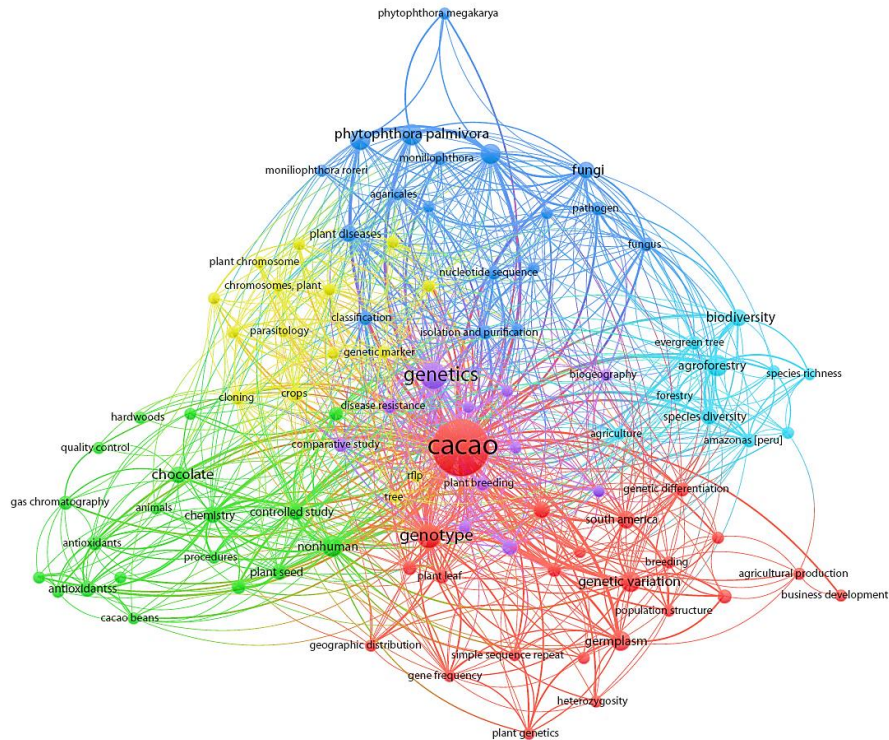


Figura 16. Mapa de red de co-ocurrencia de palabras en los títulos de los artículos asociados a cacao del Perú. El tamaño del nodo representa la cantidad de co-ocurrencias. Distancias menores entre los nodos, mayores serán el número de co-ocurrencias.

Finalmente, se describe algunos supuestos / hechos relevantes, producto del avance de las investigaciones sobre el cacao peruano. Los últimos 3 años son los que más resalta y es más notorio los cambios en la producción del cacao, considerando la transformación, la productividad y la parte socioeconómica de los productores cacaoteros (Tabla 9).

Tabla 8. Posibles supuestos de la importancia de los estudios de genómica sobre el cacao peruano.

Área de estudio	Supuesto/Hechos	Referencias
Social / económico	Incremento en la producción de cacao.	MINAGRI, 2021; Guimac-Cedillo & Verastegui Huanca, 2022

	Mayor competitividad, por la calidad del producto.	MINAGRI, 2021
	Extensión de áreas de cultivo	MINAGRI, 2021; Cedillo & Verastegui-Huanca, 2022
	Mayor ingreso económico de los cacaoteros.	MINAGRI, 2021
	Mayor asociatividad de los cacaoteros, para mejorar la rentabilidad.	MINAGRI, 2021
Agronómica	Mejor manejo del cultivo de cacao (podas de formación, sanitarias, etc.).	Oliva-Cruz et al. 2020
	Mejor manejo de la densidad de las plantaciones	Oliva-Cruz et al. 2020, 2021
Fitopatología / entomología	Nuevos métodos de control de enfermedades (<i>Moniliophthora roreri</i>).	Leiva et al. 2020
	Menor presencia de plagas y enfermedades	Leiva et al. 2020
		Valle-Epquín et al., 2020;
Bioquímica	Mayor transformación del cacao (variedades de chocolates, cacao en polvo, cocteles), diversos estudios químicos	Oliva-Cruz, et al., 2021; Castro-Alayo et al., 2019; Ordoñez et al., 2020
	Mejora de la genética del cacao, seleccionando nuevas variedades.	Estudios en proceso
	Creación de viveros con variedades uniformes de cacao.	Estudios en proceso
Genómica	Mejora del fermento de cacao, usando cultivos iniciadores.	Estudios en proceso
	Edición genética del cacao.	Estudios en proceso

IV. DISCUSIÓN

El presente estudio decodificó los genomas plastidiales de Cacao Fino de Aroma, siendo el primero en la región en utilizar tecnologías de secuenciamiento masivo para la localización y asignación de la función de los genes plastidiales de este importante cultivo. Los genomas plastidiales del Cacao Fino de Aroma (INDES34_6a, INDES34_6b, INDES67_3a, INDES67_3b) presentan su estructura, contenido, organización y características muy similares a la de otros genomas pertenecientes al género *Theobroma*.

En relación a la estructura genómica, todos los genomas presentan el mismo número total de genes (130), genes duplicados (17) y genes codificantes de proteínas (85). Esta composición es muy similar a *T. grandiflorum* (Abdullah et al., 2020). Sin embargo, las diferencias entre los cuatro genomas ensamblados se observan en los tamaños de las regiones LSC, SSC e IR, y a la variación de indels que ocurre en regiones intergénicas (Tabla 4 y 6). La mayoría de aminoácidos son codificados por codones con la misma terminación (A/T), una tendencia presente en genomas con abundancia en porcentajes de AT, el cual es típico en especies de Malvaceae (Argout et al., 2011; Abdullah et al., 2020). Adicionalmente, dentro de los cuatro genomas ensamblados se encontró el gen funcional *infA*, el cual sirve para modular la selección del ARNm generando el complejo de preiniciación, mientras que en la especie hermana de *T. cacao*, llamada *T. grandiflorum*, este gen es un pseudogen (Abdullah et al., 2020). También se identificaron algunas proteínas hipotéticas (E1A91, ES288) a lo largo de los genomas ensamblados. Para determinar su correcta función es necesario estudios del tipo metabólicos y celulares y así entender los cambios evolutivos que se están produciendo en el cacao, puesto que hay indicios de funciones aún desconocidas a lo largo del genoma plastidial del cacao (Abdullah et al., 2020). Este tipo de estudios permitirá entender la ubicación de algunos otros genes que usualmente están presentes con más frecuencia en las regiones subteloméricas (Argout et al., 2011, 2017).

Nuestros resultados también muestran un contenido similar de GC dentro de las especies de *Theobroma* (Argout et al., 2011; Abdullah et al., 2020). Los tamaños

ligeramente variantes de cada genoma plastidial del Cacao Fino de Aroma se deberían a la acumulación de indels generando variación genética (Abdullah, et al 2020; Lee et al., 2020). Esta variación genética también podría deberse a que los genes eucariotas derivados de virus inducen el intercambio genético y diversificación temprana dentro de genomas plastidiales y mitocondriales, en particular, a través de las glicosiltransferasas derivadas de la infección de estos virus (Irwin et al., 2022). Este mecanismo evolutivo es clave para facilitar la manipulación del huésped y la resistencia viral (Koonin & Krupovic, 2018; Zimmerman et al., 2019), donde los virus emplean genes del huésped para su replicación y control celular (Filée et al., 2008; Zimmerman et al., 2019). Asimismo, contrarrestan la inmunidad del huésped, secuestran la maquinaria celular y sorteando los cuellos de botella nutricionales, lo que los convierte en recursos clave para la adaptación (Koonin & Krupovic, 2018; Vardi et al., 2012). Los genes derivados del virus en los genomas eucariotas se han percibido con frecuencia como restos intrascendentes de las interacciones virales (Biémont, 2010). Sin embargo, estos genes pueden ser coaccionados y suplantar componentes celulares existentes, o pueden proporcionar una funcionalidad totalmente nueva en la planta, lo que hace que muchas plantas sean más resistentes a ciertos patógenos y otras más vulnerables (Frank & Feschotte, 2017; Irwin et al., 2022). Estos mecanismos podrían confirmar la presencia de más de un genoma plastidial en plantas de cacao tal como se observan en este estudio.

Los análisis genómicos en el Cacao Fino de Aroma permitieron develar dos versiones de genomas plastidiales para las muestras INDES34 (INDES34_6a y INDES34_6b) e INDES67 (INDES67_3a y INDES67_3b). Esto resulta ser un hecho relevante, ya que se trataría del primer reporte de “heteroplasmia” en plantas de cacao. Múltiples alelos pueden coexistir dentro de una célula vegetal debido al gran número de copias celulares de los genomas plastidiales, una condición conocida como heteroplasmia (Broz et al., 2022). La presencia de dos genomas plastidiales en cacao podría ser una causa del mecanismo de transposición (Gabriel et al., 1996) o la conservación aparente de los ETs (elementos transponibles) en el genoma de T. cacao, donde los retrotransposones muestran niveles altos de heterogeneidad (Sveinsson et al., 2013). Estos elementos transponibles causan grandes diferencias inter e intraespecíficas en

el tamaño y la composición de los genomas de las plantas (Vicient et al., 1999). En el caso del cacao solo se han reportado diferencias intraespecíficas en los ETs lo que contribuye a los tamaños variables del genoma (Argout et al., 2010). Estas diferencias en el tamaño del genoma plastidial se debe en gran medida a los indels (inserciones y deleciones de regiones nucleotídicas) en las regiones LSC (Chen et al., 2017). La confirmación de este proceso evolutivo requiere de análisis genómicos a un número mayor de variedades de cacao para así confirmar si este es un fenómeno usual, dado que la presión de selección sobre los genes y la estructura del genoma permitirá entender el proceso evolutivo del cacao (Yang et al., 2014).

Otra posible hipótesis sobre la heteroplasma en algunas variedades del Cacao Fino de Aroma podría ser el resultado de la transmisión paterna (fuga paterna) del cloroplasto a través del polen (McCauley, 2013); sin embargo, este fenómeno aún no ha sido corroborado en plantas de cacao. Cabe resaltar que se han observado heteroplasmas estructurales pequeñas en algunos genomas plastidiales (Menezes et al., 2018; Saina et al., 2018; Abdullah et al., 2020), y su existencia desafía la conclusión de que las repeticiones invertidas largas (LSC) son un requisito previo para la existencia de la heteroplasma estructural (Wang et al., 2019). Adicionalmente, la presencia de más de un genoma plastidial en el Cacao Fino de Aroma encontrados en este estudio, también podría ser producto de la transferencia horizontal de genes virales de ARN, ya que se han reportado “virus del mosaico del cacao” asociado a árboles comerciales asintomáticos de cacao en Brasil y Puerto Rico (Puig et al., 2020; Muller et al., 2021; Ramos-Sobrinho et al., 2021). A pesar que este virus no ha sido reportado en plantaciones de cacao de Amazonas, no se descarta esta posibilidad. Los mecanismos de la transferencia horizontal de genes (THG) podrían darse por parasitismo (Shirasu et al., 2010), simbiosis (Finan, 2002), patógenos, epífitos, endófitos (Bock, 2010) y por injertos (Gao et al., 2014). Este último es la actividad que más se realiza en los cultivos por parte de los productores en la Región Amazonas (Oliva y Maicelo, 2020), lo que podría promover la THG entre cultivares de cacao. Además de estos métodos de transferencia directa, los genes pueden intercambiarse con un vector puente, como el polen, hongos, bacterias, virus, viroides, plásmidos, transposones e insectos (Danchin et al., 2011; Gao et al., 2014).

Sin embargo, aún no hay reportes para de estos eventos en la Región Amazonas. Esto sugiere realizar estudios enfocados en la identificación de secuencias virales integradas en el genoma del cacao para identificar las posibles vías del probable THG y, por supuesto, la presencia de heteroplasmia y/o poliplasmia. La presencia de material genético viral está muy extendida entre los eucariotas y desempeñan un papel importante en la evolución (Liu et al., 2010). Muchos de estos genes transferidos se conservan y algunos son funcionales en los genomas receptores (Liu et al., 2010; Muller et al., 2021), ya que algunos se replican, se encapsulan en partículas virales y se propagan sistémicamente en las plantas (Catoni et al., 2018).

Los análisis filogenómicos revelaron que los genomas plastidiales INDES67_3a, INDES67_3b e INDES34 _6a tienen mayor afinidad con el grupo genético de Nacional y poseen divergencias genéticas significativas (Figura 11). Este grupo genético es hermano al clado de cacao criollo conformado por muestras de Surinam, Trinidad y Tobago y USA (Motamayor et al., 2008; Lachenaud & Motamayor, 2017). Esta diferenciación entre grupos genéticos implica una fuerte reticulación y flujo genético dentro del cacao (Kane et al., 2012). Esto sugiere que la variedad Nacional es genómicamente hermana a la variedad criolla del cacao. Por otro lado, el genoma plastidial INDES34 _6b forma un clado con las muestras de Trinidad y Tobago, USA y Perú (Figura 11), que en su conjunto tienen mayor afinidad con el grupo genético de Iquitos. Esto resalta la importancia de los estudios a nivel genómico para esclarecer la diversidad del cacao fino de aroma, ya que se pueden identificar y mostrar claras distinciones entre variedades incluso cuando se comparte un ancestro reciente (Kane et al., 2012).

Por otro lado, los genomas mitocondriales de las angiospermas son inusualmente dinámicos debido a las actividades de recombinación que implican secuencias repetidas (Gualberto & Newton, 2017). Estas actividades generan formas subgenómicas y una amplia variación genómica incluso dentro de la misma especie (Gualberto & Newton, 2017). Por ejemplo, casi todo el genoma mitocondrial de *Gossypium raimondii* (algodón) se ha transferido al cromosoma 1 (Bi et al., 2016). Adicionalmente, las distintas formas en las que se puede presentar el genoma

mitocondrial (lineal, pseudocircular) limita su ensamblaje y total entendimiento (Mower et al., 2012; McCauley, 2013). A pesar de ello, se demostró que la información evolutiva de los genomas mitocondriales en angiospermas es limitada, resaltando mucho más la importancia de la información evolutiva en alelos de origen plastidial y nuclear (Forsythe et al., 2021).

En los últimos 41 años, la investigación relacionada a cacao ha crecido en un ritmo de 11.44 % anual, siendo los dos últimos años (2020 y 2021), los años con mayor número de investigaciones en cacao para el Perú. Cabe resaltar que estos trabajos fueron publicados en revistas científicas de carácter internacional. La región Amazonas también ha crecido en cuanto a investigaciones relacionadas a cacao a pesar que esta región es considerada la tercera región más pobre del país. Esto puede estar influenciado por la presencia y licenciamiento de las universidades nacionales Universidad Nacional Toribio Rodríguez de Mendoza (UNTRM) y la Universidad Nacional Intercultural Fabiola Salazar Leguía de Bagua (UNIFSLB) (SUNEDU, 2017, 2018; Silva et al. 2022). Las investigaciones en cacao están relacionados a la biodiversidad y ecoturismo (Silva et al., 2022), otros a derivados del cacao, chocolate, antioxidantes, polifenoles y compuestos volátiles (Castro-Alayo et al., 2019; Ordoñez et al., 2020) y estudios fenotípicos (Oliva-Cruz y Maicelo, 2020; Oliva-Cruz et al., 2021b). No existen estudios moleculares en Cacao Fino de Aroma. Esto confirma el limitado uso de herramientas genómicas para el mejoramiento genético o identificación de variedades de cacao promisorios, con mejores características genéticas.

En estos últimos tiempos, los investigadores están enmarcando las contribuciones científicas al desarrollo sostenible (Wuelser & Pohl, 2016), pero aun así existe una escasez de estudios relacionados en genómica y genética en Cacao Fino de Aroma en Amazonas y en el país. Solo hay 10 publicaciones a nivel genómico con material genético de las regiones de San Martín, Loreto y Ucayali (Argout et al., 2010; Micheli et al., 2010; Allegre et al., 2012; Romero Navarro et al., 2017). Estas investigaciones han sido desarrolladas por instituciones extranjeras y ninguna de Perú. Estas investigaciones han repercutido directamente en la mejora de los estándares de

calidad de vida de la población puesto que han permitido entender la expresión de tejidos de cacao relacionados con los principales patógenos (*Moniliophthora perniciosa*, *Phytophthora* spp.) (Micheli et al., 2010) o sometidos a otras condiciones de estrés, y estudios de expresión relacionados con el sabor de la calidad del chocolate (Kane et al., 2012; Micheli et al., 2010). Otras herramientas, como el cultivo *in vitro*, la transformación de plantas y la bioinformática han permitido estudiar la genómica funcional o genes relacionados a la absorción de cadmio (Micheli et al., 2010; Correa et al., 2021). Estas investigaciones han permitido a los productores tomar mejores decisiones sobre nuevos métodos de manejo (podas, control fitosanitario, elección de porta-injertos, etc.) para incrementar su producción. Por ejemplo, nuestro país se ha incrementado la producción de cacao en un 13.8 % en el 2021, aunque esto implique que en Amazonas ha disminuido en un -9 % (MINAGRI, 2021). Esto significa que las investigaciones en cacao benefician directamente la productividad y sostenibilidad de este cultivo de gran importancia para el país. Esto influye en las cadenas productivas del cacao de la región Amazonas por ser una región productora de cacao y contribuye directamente a la mejora de las ganancias de los productores, los intermediarios, y la conservación del germoplasma. La adición de estudios con el uso de herramientas moleculares tales como marcadores moleculares y genómica permitirían seleccionar nuevas variedades de cacao con características agronómicas demandadas por los agricultores, además de poder generar uniformidad genética de variedades de cacao fino de aroma de la región Amazonas (Micheli et al., 2010; Kane et al., 2012).

Por lo tanto, la investigación científica es un pilar fundamental para mejorar la calidad de vida y bienestar de las personas, especialmente en la mejora del Cacao Fino de Aroma. Para que esta premisa se cumpla se debería aplicar mejores regímenes de gestión y programas financieros (Díaz-Valderrama et al., 2020). También deben estar orientadas a largo plazo, sobre todo basados en enfoques genómicos, ya que el uso de las ciencias ómicas ofrecen una oportunidad excepcional para la mejora de los cultivos (Mousavi-Derazmahalleh et al., 2019). De esta manera, se podría garantizar la sostenibilidad de la producción de Cacao Fino de Aroma en Amazonas. Asimismo, la investigación futura debe centrarse en la aplicación de herramientas moleculares

para la obtención de variedades de cacao que puedan tolerar diferentes tipos de estrés (Wickramasuriya & Dunwell, 2018; Mousavi-Derazmahalleh et al., 2019), y garantizar que la agricultura pueda seguir ese mismo ritmo de cambios ambientales (Mousavi-Derazmahalleh et al., 2019). Además, éstas investigaciones deben ser de carácter multidisciplinario, entre las diferentes instituciones del estado, incluyendo al sector privado (Millones-Gómez et al., 2021), con la finalidad de buscar desarrollo sostenible generando resultados favorables en el crecimiento de la sociedad.

V. CONCLUSIONES

- En este estudio, se decodificó los genomas plastidiales de Cacao Fino de Aroma, siendo el primero en el Perú en utilizar tecnologías de secuenciamiento masivo para la localización y función de los genes plastidiales de este importante cultivo de la región Amazonas.
- Los análisis de genómica estructural y funcional en el Cacao Fino de Aroma permitieron develar dos versiones de genomas plastidiales para INDES34 (INDES34_6a y INDES34_6b) e INDES67 (INDES67_3a y INDES67_3b). Estos genomas tienen la misma organización de genes que otras especies del género *Theobroma* destacando al gen *infA* como funcional, mientras que *T. grandiflorum* es un pseudogen.
- Todos los genomas ensamblados muestran el mismo número de genes (130), genes duplicados (17) y genes codificantes de proteínas (85), situación similar a la de *T. grandiflorum*. Las diferencias de estos genomas están en los tamaños de las regiones LSC, SSC e IR, y a la variación de indels que ocurre en regiones intergénicas.
- Las posibles causas de la presencia de heteroplasmia en el Cacao Fino de Aroma de Amazonas se podría entender por el mecanismo de transposición o la conservación aparente de los elementos transponibles, la transmisión paterna del cloroplasto a través del polen, o la transferencia horizontal de genes virales de ARN.
- Los análisis filogenómicos mostraron que los genomas plastidiales de INDES67_3a, INDES67_3b y INDES34_6a tienen mayor afinidad con los grupos genéticos de Nanay e Iquitos, mientras que el genoma plastidial INDES34_6b tiene mayor afinidad con el grupo genético de Contamana.
- A nivel nacional, la mayor parte de investigaciones en Cacao Fino de Aroma no están muy relacionadas con genómica, de lo contrario los estudios en cacao fino de aroma están más relacionados a la biodiversidad y ecoturismo, chocolate, antioxidantes, polifenoles, compuestos volátil y estudios fenotípicos.

- Existen pocas investigaciones relacionadas a genómica del cacao (10 publicaciones) utilizando material genético de Perú, repercutiendo indirectamente en la mejora de los estándares de calidad de vida de la población.
- La investigación futura con el uso de herramientas moleculares permitirá seleccionar nuevas variedades de cacao con características agronómicas demandadas por los agricultores, cuyos resultados influyan en las cadenas productivas del cacao Fino de Aroma de la región Amazonas.
- La investigación en Cacao Fino de Aroma debe estar orientada en enfoques genómicos con la aplicación de modernas herramientas moleculares para la obtención de nuevas variedades de cacao que puedan tolerar el cambio climático, y garantizar que la agricultura en la región Amazonas o el país sea más sostenible.

VI. RECOMENDACIONES

- Se recomienda el desarrollo de estudios genómicos en cacao, para determinar si la presencia de heteroplasmia está presente en otros cultivares de cacao de la región Amazonas y del país.
- Debido a la presencia de heteroplasmia en el cacao fino de aroma, se debe realizar estudios metabólicos o celulares para determinar algunas funciones de proteínas hipotéticas y los cambios evolutivos que se están produciendo, ya que hay indicios de funciones aún desconocidas a lo largo del genoma plastidial.
- Se debe realizar estudios más específicos sobre cómo ha impactado el inicio de la era de la bioinformática y la genómica del cacao, sobre la toma de decisiones y en la mejora de su calidad de vida de los productores cacaoteros de la región Amazonas, bajo el enfoque de desarrollo sostenible.

VII. REFERENCIAS BIBLIOGRÁFICAS

- Abdullah, Waseem, S., Mirza, B., Ahmed, I., & Waheed, M. T. (2020). Comparative analyses of chloroplast genomes of *Theobroma cacao* and *Theobroma grandiflorum*. *Biologia*, 75(5), 761-771.
- Allegre, M., Argout, X., Boccara, M., Fouet, O., Roguet, Y., Bérard, A., ... & Lanaud, C. (2012). Discovery and mapping of a new expressed sequence tag-single nucleotide polymorphism and simple sequence repeat panel for large-scale genetic studies and breeding of *Theobroma cacao* L. *DNA research*, 19(1), 23-35.
- Argout, X., Salse, J., Aury, J. M., Gaultier, M. J., Droc, G., Gouzy, J., ... & Lanaud, C. (2011). The genome of *Theobroma cacao*. *Nature genetics*, 43(2), 101-108.
- Argout, X., Martin, G., Droc, G., Fouet, O., Labadie, K., Rivals, E., ... & Lanaud, C. (2017). The cacao Criollo genome v2. 0: an improved version of the genome for genetic and functional genomic studies. *BMC genomics*, 18(1), 1-9.
- Bayer, C, Fay M. F., Bruijn. A. Y., Savolainen, V., Morton, C. M., Kubitzki, K., Alverson, W. S., & Chase, M. W. (1999). Support for an expanded family concept of Malvaceae within a circumscribed order Malvales: a combined analysis of plastid *atpB* and *rbcL* DNA sequences. *Botanic Journal of the Linnean Society*, 129(4), 267-303.
- Bartley, B. G. (2005). *The genetic diversity of cacao and its utilization*. CABI, Wallingford.
- Beck, N. & Lang, B. (2010). MFannot, organelle genome annotation webserver. Université de Montréal, Quebec, Canada. Available from: <https://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl> (accessed 02 Nov 2021).
- Bekele, F. L., Bidaisee, G. G., & Bhola, J. (2007). A comparative morphological study of two Trinitario groups from the International Cocoa Genbank, Trinidad. *Annu. Cocoa Research Unit, University of the West Indies*, 2006, pp 34-42.

- Bi, C., Paterson, A. H., Wang, X., Xu, Y., Wu, D., Qu, Y., ... & Ye, N. (2016). Analysis of the complete mitochondrial genome sequence of the diploid cotton *Gossypium raimondii* by comparative genomics approaches. *BioMed Research International*, 2016.
- Biémont, C. (2010). A brief history of the status of transposable elements: from junk DNA to major players in evolution. *Genetics*, 186(4), 1085-1093.
- Boza, E. J., Motamayor, J. C., Amores, F. M., Cedeno-Amador, S., Tondo, C. L., Livingstone, D. S., Raymond J. Schnell, R. J., & Gutiérrez, O. A. (2014). Genetic characterization of the cacao cultivar CCN 51: its impact and significance on global cacao improvement and production. *Journal of the American Society for Horticultural Science*, 139(2), 219-229.
- Bock, R. (2010). The give-and-take of DNA: horizontal gene transfer in plants. *Trends in plant science*, 15(1), 11-22.
- Broz, A. K., Keene, A., Györfy, M. F., Hodous, M., Johnston, I. G., & Sloan, D. B. (2022). Sorting of plant mitochondrial and plastid heteroplasmy is extremely rapid and depends on MSH1 activity. *bioRxiv*.
- Bustamante D. E., Motilal, L. A., Calderon, M. S., Mahabir, A., Oliva, M. (2022). Genetic diversity and population structure of fine aroma cacao (*Theobroma cacao* L.) from north Peru revealed by single nucleotide polymorphism (SNP) markers. *Frontiers in Ecology And Evolution*. in press
- Castro-Alayo, E. M., Idrogo-Vásquez, G., Siche, R., & Cardenas-Toro, F. P. (2019). Formation of aromatic compounds precursors during fermentation of Criollo and Forastero cocoa. *Heliyon*, 5(1), e01157.
- Catoni, M., Noris, E., Vaira, A. M., Jonesman, T., Matic, S., Soleimani, R., ... & Accotto, G. P. (2018). Virus-mediated export of chromosomal DNA in plants. *Nature communications*, 9(1), 1-8.
- Correa, J. E., Ramírez, R., Ruíz, O., & Leiva, E. I. (2021). Effect of soil characteristics on cadmium absorption and plant growth of *Theobroma cacao* L. seedlings. *Journal of the Science of Food and Agriculture*, 101(13), 5437-5445.

- CONCYTEC (2016). I Censo Nacional de Investigación y Desarrollo a Centros de Investigación 2016. Lima: Consejo Nacional de Ciencia, Tecnología e Innovación. 5-27.
- CONCYTEC (2021). Investigadores en el Registro Nacional de Ciencia, Tecnología y de Innovación Tecnológica. Consejo Nacional de Ciencias, Tecnología e Innovación Tecnológica. <https://ctivitaec.concytec.gob.pe/renacyt-ui/#/registro/investigadores>
- Chen, Z., Grover, C. E., Li, P., Wang, Y., Nie, H., Zhao, Y., ... & Hua, J. (2017). Molecular evolution of the plastid genome during diversification of the cotton genus. *Molecular Phylogenetics and Evolution*, 112, 268-276.
- Cheesman, E. E. (1944). Notes on the nomenclature, classification and possible relationships of cocoa populations. *Tropical Agriculture*, 2:144-159.
- Haegeman, A., Jones, J. T., & Danchin, E. G. (2011). Horizontal gene transfer in nematodes: a catalyst for plant parasitism? *Molecular Plant-Microbe Interactions*, 24(8), 879-887.
- Da Silva, M. R., Clément, D., Gramacho, K. P., Monteiro, W. R., Argout, X., Lanaud, C., y Lopes, U. (2016). Genome-wide association mapping of sexual incompatibility genes in cacao (*Theobroma cacao* L.). *Tree Genetics and Genomes*, 12(3), 62.
- Dantas, L. G. y M. Guerra. (2010). Chromatin differentiation between *Theobroma cacao* L. and *T. grandiflorum* Schum. *Genetics and Molecular Biology*, 33:94–98.
- Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., ... & Gascuel, O. (2008). Phylogeny. fr: robust phylogenetic analysis for the non-specialist. *Nucleic acids research*, 36(suppl_2), W465-W469.
- Díaz-Valderrama, J. R., Leiva-Espinoza, S. T., & Aime, M. C. (2020). The history of cacao and its diseases in the Americas. *Phytopathology*, 110(10), 1604-1619.
- El Peruano (2015). Declaran de interés regional la obtención de la denominación de origen del “Cacao Amazonas Perú”; Ordenanza Regional N° 368, Gobierno Regional Amazonas/CR. Disponible en:

<https://busquedas.elperuano.pe/normaslegales/declaran-de-interes-regional-la-obtencion-de-la-denominacion-ordenanza-no-368-gobierno-regional-amazonascr-1270354-1/>

- Fister, A. S., Shi, Z., Zhang, Y., Helliwell, E. E., Maximova, S. N., & Guiltinan, M. J. (2016). Protocol: transient expression system for functional genomics in the tropical tree *Theobroma cacao* L. *Plant methods*, 12(1), 1-13.
- Filée, J., Pouget, N., & Chandler, M. (2008). Phylogenetic evidence for extensive lateral acquisition of cellular genes by Nucleocytoplasmic large DNA viruses. *BMC Evolutionary Biology*, 8(1), 1-13.
- Frank, J. A., & Feschotte, C. (2017). Co-option of endogenous viral sequences for host cell function. *Current opinion in virology*, 25, 81-89.
- Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M., & Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic acids research*, 32(suppl_2), W273-W279.
- Finan, T. M. (2002). Evolving insights: symbiosis islands and horizontal gene transfer. *Journal of bacteriology*, 184(11), 2855-2856.
- Forsythe, E. S., Williams, A. M., & Sloan, D. B. (2021). Genome-wide signatures of plastid-nuclear coevolution point to repeated perturbations of plastid proteostasis systems across angiosperms. *The Plant Cell*, 33(4), 980-997.
- Gao, C., Ren, X., Mason, A. S., Liu, H., Xiao, M., Li, J., & Fu, D. (2014). Horizontal gene transfer in plants. *Functional & integrative genomics*, 14(1), 23-29.
- Gabriel, A., Willems, M., Mules, E. H., & Boeke, J. D. (1996). Replication infidelity during a single cycle of Ty1 retrotransposition. *Proceedings of the National Academy of Sciences*, 93(15), 7767-7771.
- Gopaulchan, D., Motilal, L. A., Bekele, F. L., Clause, S., Ariko, J. O., Ejang, H. P., & Umaharan, P. (2019). Morphological and genetic diversity of cacao (*Theobroma cacao* L.) in Uganda. *Physiology and Molecular Biology of Plants*, 25(2), 361-375.
- Gualberto, J. M., & Newton, K. J. (2017). Plant mitochondrial genomes: dynamics and mechanisms of mutation. *Annual Review of Plant Biology*, 68, 225-252.

- Gumiel, M., Rollano-Peñaloza, O. M., Peralta-Rivero, C., Tejeda, L., Palma, V., Cartagena, P., ... & Peñarrieta, J. M. (2021). Whole Chloroplast Genomes reveals the uniqueness of Bolivian native cacao (*Theobroma cacao*) from the northern part of Bolivia. *bioRxiv*.
- Irwin, N. A., Pittis, A. A., Richards, T. A., & Keeling, P. J. (2022). Systematic evaluation of horizontal gene transfer between eukaryotes and viruses. *Nature microbiology*, 7(2), 327-336.
- Jansen, R. K., Saski, C., Lee, S. B., Hansen, A. K., & Daniell, H. (2011). Complete plastid genome sequences of three rosids (*Castanea*, *Prunus*, *Theobroma*): evidence for at least two independent transfers of *rpl22* to the nucleus. *Molecular biology and evolution*, 28(1), 835-847.
- Jhonsy Silva, Nilton B. Rojas-Briceño, Daniel Tineo, Eli Morales, Jhovana Sopla, Jhordy Perez, Nelson Rodríguez., ... & Danilo E. Bustamante (2022). Contributions of scientific research to regional development in the Amazonas region, northern Peru. *Development Studies Research*, 9:1, 129-141
- Kane, N., Sveinsson, S., Dempewolf, H., Yang, J. Y., Zhang, D., Engels, J. M., & Cronk, Q. (2012). Ultra-barcoding in cacao (*Theobroma* spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. *American Journal of Botany*, 99(2), 320-329.
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30(4):772–780.
- Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: a fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, 35(21):4453–4455.
- Koonin, E. V., & Krupovic, M. (2018). The depths of virus exaptation. *Current opinion in virology*, 31, 1-8.
- Lachenaud, P., & Motamayor, J. C. (2017). The Criollo cacao tree (*Theobroma cacao* L.): a review. *Genetic Resources and Crop Evolution*, 64(8), 1807-1820.

- Lee, C., Ruhlman, T. A., & Jansen, R. K. (2020). Unprecedented intraindividual structural heteroplasmy in *Eleocharis* (Cyperaceae, Poales) plastomes. *Genome biology and evolution*, 12(5), 641-655.
- Leiva, S., Oliva, M., Hernández, E., Chuquibala, B., Rubio, K., García, F., & Torres de la Cruz, M. (2020). Assessment of the potential of trichoderma spp. strains native to Bagua (Amazonas, Peru) in the biocontrol of frosty pod rot (*Moniliophthora roreri*). *Agronomy*, 10(9), 1376.
- Liu, H., Fu, Y., Jiang, D., Li, G., Xie, J., Cheng, J., ... & Yi, X. (2010). Widespread horizontal gene transfer from double-stranded RNA viruses to eukaryotic nuclear genomes. *Journal of virology*, 84(22), 11876-11887.
- Loor, R G., Risterucci, A. M., Courtois, B., Fouet, O., Jeanneau, M., Rosenquist, E., Amores, F., Vasco, A., Medina, M., & Lanaud, C. (2009). Tracing the native ancestors of the modern *Theobroma cacao* L. population in Ecuador. *Tree Genetics & Genomes*, 5:421–433
- Menezes, A., Resende-Moreira, L. C., Buzatti, R. S., Nazareno, A. G., Carlsen, M., Lobo, F. P., ... & Lovato, M. B. (2018). Chloroplast genomes of *Byrsonima* species (Malpighiaceae): comparative analysis and screening of high divergence sequences. *Scientific Reports*, 8(1), 1-12.
- McCauley, D. E. (2013). Paternal leakage, heteroplasmy, and the evolution of plant mitochondrial genomes. *New Phytologist*, 200(4), 966-977.
- Micheli, F., Gultinan, M., Gramacho, K. P., Wilkinson, M. J., De Oliveira Figueira, A. V., de Mattos Cascardo, J. C., ... & Lanaud, C. (2010). Functional genomics of cacao. *Advances in Botanical Research*, 55, 119-177.
- Millones-Gómez, P. A., Yangali-Vicente, J. S., Arispe-Alburquerque, C. M., Rivera-Lozada, O., Calla-Vásquez, K. M., Calla-Poma, R. D., ... & Minchón-Medina, C. A. (2021). Research policies and scientific production: A study of 94 Peruvian universities. *PloS one*, 16(5), e0252410.
- MINAGRI. (2018). Estudio del cacao del Perú y el mundo: Un análisis de la producción y el comercio. <https://www.minagri.gob.pe/portal/monitoreo-agroclimatico/cacao-2018>.

- Mower, J. P., Sloan, D. B., & Alverson, A. J. (2012). Plant mitochondrial genome diversity: the genomics revolution. *Plant Genome Diversity Volume 1*, 123-144.
- Motamayor J. C., Lachenaud P., Da Silva e Mota J. W., Loor, R. G., Martinez, W. J., Graham, J., Kuhn, D. N., Brown, S., y Schnell, R. J. (2010). No más forastero: a new protocol for meaningful cacao germplasm classification. In: *Proceedings of the 16th international cocoa research conference, 2010*. COPAL, Nigeria, pp 179–185.
- Motamayor, J. C., Lachenaud, P., Da Silva, e Mota J. W., Loor, R., Kuhn, D. N., Brown, J. S., y Schnell, R. J. (2008). Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L). *PLoS ONE*. 3(10): e3311.
- Motamayor, J. C., Risterucci, A. M., Heath, M., Lanaud, C. (2003). Cacao domestication II: progenitor germplasm of the Trinitario cacao cultivar. *Heredity*. 91:322–330
- Motamayor, J. C., Risterucci, A. M., Lopez, P. A., Ortiz, C. F., Moreno, A., y Lanaud, C. (2002). Cacao domestication I: the origin of the cacao cultivated by the Mayas. *Heredity*, 89(5), 380-386.
- Motilal, L. A. y Sreenivasan, T. N. (2012). Revisiting 1727: crop failure leads to the birth of Trinitario cacao. *Journal of Crop Improvement*, 26(5), 599–626.
- Motilal, L. A., Zhang, D., Umaharan, P., Mischke, S., Mooleedhar, V., y Meinhardt, L. W. (2010). The relic Criollo cacao in Belize – genetic diversity and relationship with Trinitario and other cacao clones held in the International Cocoa Genebank Trinidad. *Plant Genet Resource*, 5(3):106–115.
- Mousavi-Derazmahalleh, M., Bayer, P. E., Hane, J. K., Valliyodan, B., Nguyen, H. T., Nelson, M. N., ... & Edwards, D. (2019). Adapting legume crops to climate change using genomic approaches. *Plant, Cell & Environment*, 42(1), 6-19.
- Muller, E., Ullah, I., Dunwell, J. M., Daymond, A. J., Richardson, M., Allainguillaume, J., & Wetten, A. (2021). Identification and distribution of

- novel badnaviral sequences integrated in the genome of cacao (*Theobroma cacao*). *Scientific reports*, 11(1), 1-13.
- Tang, M., Chen, Z., Grover, C. E., Wang, Y., Li, S., Liu, G., ... & Hua, J. (2015). Rapid evolutionary divergence of *Gossypium barbadense* and *G. hirsutum* mitochondrial genomes. *BMC genomics*, 16(1), 1-16.
- Toxopeus, H. (1985). Botany, types and populations. In: Wood GAR, Lass RA (eds) *Cocoa*, 4th ed. Longman, London, pp 11–37.
- Oliva, M., & Maicelo, J. L. Q. (2020). Identificación y selección de ecotipos de cacao nativo fino de aroma de la zona Nor oriental del Perú. *Revista de Investigación de Agroproducción Sustentable*, 4(2), 31-39.
- Oliva-Cruz, M., Mori-Culqui, P. L., Caetano, A. C., Goñas, M., Vilca-Valqui, N. C., & Chavez, S. G. (2021a). Total fat content and fatty acid profile of fine-aroma cocoa from northeastern Peru. *Frontiers in Nutrition*, 397.
- Oliva-Cruz, M., Goñas, M., García, L. M., Rabanal-Oyarse, R., Alvarado-Chuqui, C., Escobedo-Ocampo, P., & Maicelo-Quintana, J. L. (2021b). Phenotypic Characterization of Fine-Aroma Cocoa from Northeastern Peru. *International Journal of Agronomy*, 2021.
- Ordoñez, E. S., Quispe, Y., & García, L. F. (2020). Cuantificación de fenoles, antocianinas y caracterización sensorial de nibs y licor de cinco variedades de cacao, en dos sistemas de fermentación. *Scientia Agropecuaria*, 11(4), 473-481.
- Osorio-Guarín, J. A., Berdugo-Cely, J., Coronado, R. A., Zapata, Y. P., Quintero, C., Gallego-Sánchez, G., y Yockteng, R. (2017). Colombia a source of cacao genetic diversity as revealed by the population structure analysis of germplasm bank of *Theobroma cacao* L. *Frontiers in plant science*, 8, 1994.
- Parr, R. L., Maki, J., Reguly, B., Dakubo, G. D., Aguirre, A., Wittcock, R., ... & Thayer, R. E. (2006). The pseudo-mitochondrial genome influences mistakes in heteroplasmy interpretation. *BMC genomics*, 7(1), 1-13.
- Puig, A. S., Ramos-Sobrinho, R., Keith, C., Kitchen, N., Gutierrez, O., Goenaga, R., & Brown, J. K. (2020). First Report of Cacao Mild Mosaic Virus

- Associated with Symptomatic Commercial Cacao (*Theobroma cacao*) Trees in Puerto Rico. *Plant disease*, 104(11), 3089-3089.
- Ramos-Sobrinho, R., Ferro, M. M., Nagata, T., Puig, A. S., Keith, C. V., Britto, D. S., ... & Brown, J. K. (2021). Complete genome sequences of three newly discovered cacao mild mosaic virus isolates from *Theobroma cacao* L. in Brazil and Puerto Rico and evidence for recombination. *Archives of virology*, 166(7), 2027-2031.
- Romero Navarro, J. A., Phillips-Mora, W., Arciniegas-Leal, A., Mata-Quirós, A., Haiminen, N., Mustiga, G., ... & Motamayor, J. C. (2017). Application of genome wide association and genomic prediction for improvement of cacao productivity and resistance to black and frosty pod diseases. *Frontiers in plant science*, 8, 1905.
- Saina, J. K., Gichira, A. W., Li, Z. Z., Hu, G. W., Wang, Q. F., & Liao, K. (2018). The complete chloroplast genome sequence of *Dodonaea viscosa*: comparative and phylogenetic analyses. *Genetica*, 146(1), 101-113.
- Schattner, P., Brooks, A. N., y Lowe, T. M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucl Acids Research*, 33:686–689.
- Shahzadi, I., Mehmood, F., Ali, Z., Malik, M. S., Waseem, S., Mirza, B., ... & Waheed, M. T. (2019). Comparative analyses of chloroplast genomes among three *Firmiana* species: Identification of mutational hotspots and phylogenetic relationship with other species of Malvaceae. *Plant Gene*, 19, 100199.
- Yoshida, S., Maruyama, S., Nozaki, H., & Shirasu, K. (2010). Horizontal gene transfer by the parasitic plant *Striga hermonthica*. *Science*, 328(5982), 1128-1128.
- SUNEDU (Superintendencia Nacional de Educación Superior Universitaria) (2017) Resolución del Consejo Directivo N° 033-2017-SUNEDU/CD: Otorga la Licencia Institucional a la Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, para ofrecer el servicio educativo superior universitario.

- SUNEDU (Superintendencia Nacional de Educación Superior Universitaria) (2018) Resolución del Consejo Directivo N° 095-2018-SUNEDU/CD.: Otorgan la Licencia Institucional a la Universidad Nacional Intercultural Fabiola Salazar Leguía de Bagua, para ofrecer el servicio educativo superior universitario.
- Sveinsson, S., Gill, N., Kane, N. C., & Cronk, Q. (2013). Transposon fingerprinting using low coverage whole genome shotgun sequencing in Cacao (*Theobroma cacao* L.) and related species. *BMC genomics*, 14(1), 1-12.
- Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., & Greiner, S. (2017). GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic acids research*, 45(W1), W6-W11.
- Thiers, B., (2016). Index Herbariorum. A global directory of public herbaria and associated staff. New York Botanical Garden's Virtual Herbarium. Available from: <http://sweetgum.nybg.org/science/ih> (accessed 2021).
- Utro, F., Cornejo, O. E., Livingstone, D., Motamayor, J. C., y Parida L. (2012). ARG-based genome-wide analysis of cacao cultivars. *BMC Bioinformatics* 13(Suppl 19):S17.
- Van Eck, N., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *scientometrics*, 84(2), 523-538.
- Valle-Epquín, M. G., Balcázar-Zumaeta, C. R., Auquiñivín-Silva, E. A., Fernández-Jeri, A. B., Idrogo-Vásquez, G., & Castro-Alayo, E. M. (2020). The roasting process and place of cultivation influence the volatile fingerprint of Criollo cacao from Amazonas, Peru. *Scientia Agropecuaria*, 11(4), 599-610.
- Vicient, C. M., Suoniemi, A., Ananthawat-Jónsson, K., Tanskanen, J., Beharav, A., Nevo, E., & Schulman, A. H. (1999). Retrotransposon BARE-1 and its role in genome evolution in the genus *Hordeum*. *The Plant Cell*, 11(9), 1769-1784.
- Wang, W., & Lanfear, R. (2019). Long-reads reveal that the chloroplast genome exists in two distinct versions in most plants. *Genome biology and evolution*, 11(12), 3372-3381.

- Wickramasuriya, A. M., y Dunwell, J. M. (2018). Cacao biotechnology: current status and future prospects. *Plant biotechnology journal*, 16(1), 4-17.
- World Cocoa Foundation (2014). Cocoa market update. Wuelser, G., & Pohl, C. (2016). How researchers frame scientific contributions to sustainable development: a typology based on grounded theory. *Sustainability science*, 11(5), 789-800.
- Vardi, A., Haramaty, L., Van Mooy, B. A., Fredricks, H. F., Kimmance, S. A., Larsen, A., & Bidle, K. D. (2012). Host–virus dynamics and subcellular controls of cell fate in a natural coccolithophore population. *Proceedings of the National Academy of Sciences*, 109(47), 19327-19332.
- Yang, J. Y., Scascitelli, M., Motilal, L. A., Sveinsson, S., Engels, J. M., Kane, N. C., Dempewolf, H., Zhang, D., Maharaj, K., y Cronk, Q. C. (2013). Complex origin of Trinitario-type *Theobroma cacao* (Malvaceae) from Trinidad and Tobago revealed using plastid genomics. *Tree genetics and genomes*, 9(3), 829-840.
- Zhang, D., Martínez, W. J., Johnson, E. S., Somarriba, E., Phillips-Mora, W., Astorga, C., Mischke, S., & Meinhardt, L. W. (2012). Genetic diversity and spatial structure in a new distinct *Theobroma cacao* L. population in Bolivia. *Genetic Resources and Crop Evolution*, 59(2), 239-252.
- Zimmerman, A. E., Howard-Varona, C., Needham, D. M., John, S. G., Worden, A. Z., Sullivan, M. B., ... & Coleman, M. L. (2020). Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nature Reviews Microbiology*, 18(1), 21-34.