UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN CIENCIAS BIOLÓGICAS

INSTITUTO DE INVESTIGACIONES EN ECOSISTEMAS Y SUSTENTABILIDAD - IIES

DOMESTICACIÓN Y DIVERSIDAD GENÉTICA DE *Psidium guajava* L. (MYRTACEAE) EN EL CONTINENTE AMERICANO: UNA APROXIMACIÓN FILOGEOGRÁFICA

TESIS

QUE PARA OPTAR POR EL GRADO DE:

DOCTORA EN CIENCIAS

PRESENTA:

EDNA ARÉVALO MARÍN

TUTOR PRINCIPAL DE TESIS: DR. ALEJANDRO CASAS FERNÁNDEZ
        INSTITUTO DE INVESTIGACIONES EN ECOSISTEMAS Y SUSTENTABILIDAD, UNAM
COMITÉ TUTOR: DR. CHARLES ROLAND CLEMENT
        INSTITUTO NACIONAL DE PESQUISAS DA AMAZONIA, INPA
COMITÉ TUTOR: DR. EDUARDO RUÍZ SÁNCHEZ
        CENTRO UNIVERSITÁRIO DE CIENCIAS BIOLÓGICAS Y AGROPECUARIAS, UdG
COMITÉ TUTOR: DR. ANTONIO GONZÁLEZ RODRÍGUEZ
        INSTITUTO DE INVESTIGACIONES EN ECOSISTEMAS Y SUSTENTABILIDAD, UNAM

MORELIA, MICHOACÁN ENERO, AÑO 2024

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN CIENCIAS BIOLÓGICAS

INSTITUTO DE INVESTIGACIONES EN ECOSISTEMAS Y SUSTENTABILIDAD - IIES

DOMESTICACIÓN Y DIVERSIDAD GENÉTICA DE *Psidium guajava* L. (MYRTACEAE) EN EL CONTINENTE AMERICANO: UNA APROXIMACIÓN FILOGEOGRÁFICA

TESIS

QUE PARA OPTAR POR EL GRADO DE:

DOCTORA EN CIENCIAS

PRESENTA:

EDNA ARÉVALO MARÍN

TUTOR PRINCIPAL DE TESIS: DR. ALEJANDRO CASAS FERNÁNDEZ
INSTITUTO DE INVESTIGACIONES EN ECOSISTEMAS Y SUSTENTABILIDAD , UNAM
COMITÉ TUTOR: DR. CHARLES ROLAND CLEMENT
INSTITUTO NACIONAL DE PESQUISAS DA AMAZONIA, INPA
COMITÉ TUTOR: DR. EDUARDO RUÍZ SÁNCHEZ
CENTRO UNIVERSITÁRIO DE CIENCIAS BIOLÓGICAS Y AGROPECUARIAS, UdG
COMITÉ TUTOR: DR. ANTONIO GONZÁLEZ RODRÍGUEZ
INSTITUTO DE INVESTIGACIONES EN ECOSISTEMAS Y SUSTENTABILIDAD, UNAM

MORELIA, MICHOACÁN ENERO, AÑO 2024

M. en C. Ivonne Ramírez Wence
Directora General de Administración Escolar, UNAM
P r e s e n t e

Me permito informar a usted que en la reunión ordinaria del Comité Académico del Posgrado en Ciencias Biológicas, celebrada el día **29 de mayo de 2023** se aprobó el siguiente jurado para el examen de grado de **DOCTORA EN CIENCIAS** de la estudiante **ARÉVALO MARÍN EDNA** con número de cuenta **517494264** con la tesis titulada **"DOMESTICACIÓN Y DIVERSIDAD GENÉTICA DE PSIDIUM GUAJAVA L. (MYRTACEAE) EN EL CONTINENTE AMERICANO: UNA APROXIMACIÓN FILOGEOGRÁFICA",** realizada bajo la dirección del **DR. ALEJANDRO CASAS FERNÁNDEZ,** quedando integrado de la siguiente manera:

Presidente:	DRA. ALEJANDRA CITLALLI MORENO LETELIER
Vocal:	DR. LEV ORLANDO JARDÓN BARBOLLA
Secretario:	DR. EDUARDO RUÍZ SÁNCHEZ
Vocal:	DRA. ALEJANDRA VÁZQUEZ LOBO YURÉN
Vocal:	DRA. XITLALI AGUIRRE DUGUA

Sin otro particular, me es grato enviarle un cordial saludo.

A T E N T A M E N T E
"POR MI RAZA HABLARÁ EL ESPÍRITU"
Ciudad Universitaria, Cd. Mx., a 24 de octubre de 2023

COORDINADOR DEL PROGRAMA

DR. ADOLFO GERARDO NAVARRO SIGÜENZA

c. c. p. Expediente del alumno

AGNS/LTC/EARR/ltc

AGRADECIMIENTOS INSTITUCIONALES

AGRADECIMIENTOS PERSONALES

Gracias por el cariño, la compañía y el apoyo a Hernán, Gatinho von Sauerkraut, Sakura Kurosawa, Grigio Piccolino (*in memoriam*) y Morcillo de Entrerríos.

Agradezco a mi mami, Liliana Marín, a mis hermanos Lely y Vladi y a mi sobrino Samu por apoyarme en todo momento y alegrarse siempre de mis logros. A mi familia Marín Bedoya por estar siempre presente.

A la familia Alvarado-Sizzo por acogerme en su seno familiar durante todo este tiempo.

Agradezco al Dr. Alejandro Casas, por aceptarme y darme la oportunidad de desarrollar mi investigación doctoral en su laboratorio, por la ayuda académica y personal, por su calidad humana y por confiar en mis capacidades.

Agradezco al Dr. Charles R. Clement por sus interminables cuestionamientos, por alentarme a realizar y finalizar este trabajo, por hacerme parte de su equipo desde la distancia, por sus orientaciones y, sobre todo, por sus *des-orientaciones.* ¡¡Y hasta aquí porque…KISS!!

Agradezco al Dr. Eduardo Ruíz Sánchez, por siempre aclarar mis dudas, compartir y debatir diferentes ideas a cualquier hora del día y cualquier día de la semana, por orientarme siempre que estuve perdida y por ayudarme en mis debates con CRC.

Un agradecimiento especial al Dr. Leslie Landrum (ASU) y a la Dra. Myrtle P. Shock (UFOPA), por todas sus valiosas contribuciones para el desarrollo de este trabajo. Ustedes son investigadores de gran conocimiento en cada una de sus respectivas áreas y con una calidad humana realmente inspiradora.

Agradezco a la Dra. Andrea Cosacov (UNC), por abrirme las puertas de su laboratorio y sumar su valioso conocimiento para el éxito de este trabajo.

A la Dra. Alicia Castillo por el apoyo extra para la culminación de este trabajo.

Al Dr. Francisco Mora por siempre estar presto a compartir su conocimiento ecológico, estadístico y del uso adecuado de R.

TABLA DE CONTENIDO

RESUMEN

La guayaba es un cultivo económicamente importante en países de la región neotropical, especialmente en Brasil, Perú, Colombia, Las Antillas, Centroamérica y México. Sin embargo, la distribución actual de su cultivo incluye regiones de India, Pakistán, China y Egipto. A pesar de su importancia a nivel global, preguntas sobre su origen, historia evolutiva bajo domesticación y su dispersión en América siguen sin ser abordadas. Ante este panorama los objetivos de este trabajo se centraron en: 1) Examinar y sistematizar información biológica, molecular, arqueológica e histórica sobre la guayaba para analizar su posible origen y difusión; 2) Determinar el patrón filogeográfico de *Psidium guajava* en su rango de distribución con el fin de brindar apoyo o revisar las hipótesis acerca de su historia; 3) Evaluar la distribución espacial de la estructura y diversidad genética de *P. guajava* a escala continental, con el fin de determinar las consecuencias del proceso de domesticación sobre los patrones de variación genética observados.

Este estudio se enfocó en la recuperación, revisión y sistematización de la información disponible sobre evidencia ecológica, taxonómica, genética, arqueológica e histórica disponible sobre la guayaba, y con base en esta se desarrollaron propuestas hipotéticas para orientar las investigaciones sobre evolución y domesticación de *P. guajava* (Capítulo I). Para avanzar en el estudio de la historia evolutiva e identificar los procesos que moldearon la distribución histórica de la guayaba en América, se emplearon enfoques filogeográficos y demográficos usando datos de secuencias de cloroplasto y nucleares (ITS y gen MeNu39 de una sola copia) de 241 accesiones que cubren casi todo el rango de distribución de la especie en el continente americano (Capítulo II). Respecto a la historia reciente de la especie, se estudió la distribución espacial, la estructura y la historia de la diversidad genética de *P. guajava* en las Américas. Se analizaron 215 individuos de 11 países de las regiones mesoamericana, andina y amazónica mediante el uso de 25 loci de microsatélites nucleares específicos para la especie, aplicando análisis bayesianos de estructura y diferenciación genética (Capítulo III).

Se encontró un amplio número de haplotipos y ribotipos. La mayoría los haplotipos y ribotipos más frecuentes estaban ampliamente distribuidos. Sin embargo, se recuperaron linajes privados para Sudamérica y Mesoamérica. Las poblaciones de guayaba de América del Sur mostraron valores más altos de diversidad genética y un mayor número de haplotipos privados, principalmente en las muestras amazónicas de Perú y Brasil, así como en las muestras de los Andes peruanos. Estas son interesantes áreas, probables escenarios de diversificación de guayaba. El análisis bayesiano de agrupamiento filogeográfico y ecológico (BPEC) también respaldó estos hallazgos.

Nuestros análisis demográficos históricos revelaron expansiones demográficas principalmente para el linaje sudamericano. El tiempo de divergencia estimado usando un reloj molecular, empleando marcadores de cloroplasto y núcleo, indican que la guayaba podría haberse originado en el Mioceno (~13 - ~8 Mya), divergiendo rápidamente en dos linajes principales: el linaje de América del Sur, el cual diversificó hace ~11 - ~8 Ma., y el linaje Mesoamericano diversificando hace ~5 - ~1 Ma. Por lo tanto, su origen no parece estar asociado a un área geográfica precisa, sino podría considerarse como una especie de amplia distribución. Adicionalmente, esta evidencia demuestra que el rango de distribución natural de la especie sería el resultado, primariamente, de eventos paleoclimáticos y biogeográficos y no de actividades antropogénicas.

El abordaje de genética poblacional nos permitió identificar grupos genéticos específicos para América del Sur y Mesoamérica. Las guayabas amazónicas de Brasil y Perú y de los Andes peruanos mostraron estructura. En contraste, las muestras de guayaba de Colombia y Venezuela se superponen con las de las Antillas. De igual forma, definimos un grupo Mesoamericano (Centroamérica + México). Estos patrones de diferenciación fueron consistentes en todas las pruebas utilizadas. Los grupos más contrastantes, según Structure, fueron los de los Andes peruanos, la Amazonía brasileña y el grupo mexicano. En el caso de la guayaba, el valor F_{ST} observado puede atribuirse al flujo genético limitado entre las poblaciones muestreadas, que abarcan diferentes regiones del Neotrópico. Sin embargo, las poblaciones del norte de Sudamérica, Centroamérica y Mesoamérica muestran valores de diferenciación más bajos y algunos individuos los tienen menos definidos (mixtos). Este escenario sugiere un patrón de diferenciación Norte-Sur (mayor diferenciación en el Sur), dado el aislamiento entre las poblaciones de guayaba en general, especialmente las de las regiones amazónica y andina. Por último, según el escenario demográfico más favorable, la primera ruta de difusión de la guayaba probablemente fue a través del Amazonas brasileño hacia los Andes peruanos. Por tanto, la guayaba pudo haberse extendido por los Andes, desde el Amazonas hasta la costa peruana.

Se concluye que el estudio sobre la historia evolutiva de frutales perennes debe abordarse desde múltiples enfoques que permitan dilucidar claramente los efectos de los procesos naturales y de los procesos antrópicos en especies ampliamente distribuidas. Con base en los resultados obtenidos, se recomienda llevar a cabo investigaciones con un muestreo focalizado en las poblaciones andinas y amazónicas e incluso con poblaciones del extremo Sur de la distribución (en Argentina y Paraguay). Así mismo, es necesario ampliar los estudios arqueobotánicos usando estrategias

paleogenómicas que permitan ampliar la evidencia sobre la historia de uso de la guayaba en las poblaciones mesoamericanas y suramericanas.

ABSTRACT

Guava is an economically important crop in tropical countries worldwide. In the Neotropics its cultivation is particularly relevant in Brazil, Peru, Colombia, the Antilles, Central America, and Mexico. But its current distribution area includes regions of China, Egypt, India, and Pakistan. Despite its global significance, questions regarding its origin, evolutionary history under domestication, and dispersal in the Americas remain unaddressed. Given this scenario, the aims of this study were: 1) To examine and systematize biological, molecular, archaeological, and historical information about guava to analyze its possible origin and diffusion; 2) to determine the phylogeographic patterns and gene flow of *Psidium guajava* within its distribution range for testing historical hypotheses; 3) to evaluate the spatial variation of genetic diversity and genetic structure of *P. guajava* at the scale of the whole Neotropics, to understand the consequences of the domestication process on the patterns of neutral genetic variation.

This study started with recovering, revieing, and systematizing of available information concerning the natural and historical aspects about the species. A thorough examination of ecological, taxonomic, genetic, archaeological, and historical evidence regarding guava was conducted, and this work led to the formulation of hypothetical proposals to guide research on the evolutionary and domestication processes of *P. guajava* (Chapter I). In an endeavor to elucidate the evolutionary history and identify the events that shaped the historical distribution of guava in the Americas, phylogeographic and demographic approaches were conducted though plastid and nuclear sequence data (ITS and single-copy gene MeNu39) obtained from 241 accessions encompassing nearly the entire guava's distribution across the Americas (Chapter II). The recent history of guava, its spatial distribution, structure, and demographic history in the Americas were also investigated. A comprehensive analysis involving 215 individuals from 11 countries within the Mesoamerican, Andean, and Amazonian regions was conducted by analyzing 25 nuclear microsatellite loci through structure and Bayesian differentiation analyses (Chapter III).

Many haplotypes and ribotypes were found, with the most frequent haplotypes and ribotypes being those widely distributed in the Americas. However, private lineages were recovered for South America and Mesoamerica. Guava populations from South America showed higher values of genetic diversity and a higher number of private haplotypes, mainly in the Amazonian samples from Peru and Brazil as well as in samples from the Peruvian Andes. These areas are interesting likely settings of guava diversification. Bayesian phylogeographic and ecological clustering analysis (BPEC) also supported these findings. Our historical demographic analyses revealed demographic expansions mainly for the South American lineage. The estimated

molecular time based on plastid and nuclear markers suggests that guava could have originated in the Miocene (~13 - ~8 Mya), diverging rapidly into two main lineages: the South American lineage, which diversified between ~11 and ~8 Mya ago, and the Mesoamerican lineage diversifying ~5 - ~1 Mya. Therefore, its origin does not appear to be associated with a precise geographic area. Guava could be considered as a widely distributed species. Thus, evidence suggests that current distribution of guava is foremost the result of paleoclimatic and biogeographic events rather than anthropogenic activities.

The population genetics approach allowed identifying specific genetic groups for South America and Mesoamerica. The Amazonian guavas from Brazil and Peru, and those from the Peruvian Andes showed genetic structure and high genetic diversity. The guava samples from Colombia and Venezuela grouped together with those from the Antilles. It was also defined a Mesoamerican group (Central America + Mexico). These differentiation patterns were consistent across all the tests used.

The most contrasting groups determined by Structure analysis were the Peruvian Andes, the Brazilian Amazon, and the Mexican group. The F_{ST} value estimated may be attributed to limited gene flow among the sampled populations, which span different regions of the Neotropics. However, populations from northern South America, Central America and Mesoamerica show lower differentiation values, and some individuals have less well defined (mixed) differentiation values. This scenario suggests a North-South differentiation pattern (higher differentiation in the South), given the isolation between the guava populations analyzed, especially those of the Amazon and Andean regions. Finally, according to the most favorable demographic scenario, the first route of guava diffusion was probably through the Brazilian Amazon to the Peruvian Andes. Therefore, the Amazonia is a probable primary area of guava domestication. Thus, guava could have spread through the Andes, from the Amazon to the Peruvian coast, and from there to the north of the continent.

It is possible to conclude that understanding the evolutionary history of perennial fruit trees requires to be approached through an interdisciplinary perspective that clearly elucidate the effects of natural processes and anthropogenic influences on widely distributed useful plants. Based on the results of this study, it is recommended further research with a focused sampling strategy targeting the Andean and Amazonian populations, including populations from the southernmost distribution range (Argentina and Paraguay). Furthermore, it is necessary to expand the archaeobotanical studies by employing paleogenomic strategies that help to elucidate the history of guava utilization in Mesoamerican and South American populations.

INTRODUCCIÓN GENERAL

Áreas de origen: biogeografía

Establecer el área de origen y explicar los patrones biogeográficos observados para una especie, puede ser una tarea difícil de conseguir. El patrón de la distribución espacial actual de las plantas se basa principalmente en la dinámica de los procesos evolutivos, biogeográficos, climáticos, y ecológicos (Villaseñor, 2016). En términos generales, poblaciones pequeñas con limitado flujo de genes y diversas presiones de selección y/o deriva, presentan patrones genéticos y espaciales de diferenciación poblacional en contraposición a poblaciones más grandes, con flujo génico y presiones de selección constante, las cuales muestran patrones uniformes de variación en diferentes regiones geográficas. A su vez, la cantidad de variación está determinada por el tamaño efectivo de la población, los eventos históricos, los sistemas de apareamiento y la estructura de la población (Menken et al., 1995). Por lo tanto, los patrones de distribución de la variabilidad genética son el resultado de una variedad de parámetros poblacionales (Menken et al., 1995), combinados con eventos de dispersión y/o vicarianza (Avise, 2000). En el caso de las plantas cultural y económicamente importantes, su rango de distribución natural puede haberse extendido como consecuencia de la transferencia de material debido a las actividades humanas (Schaefer et al., 2009). En este contexto, reunir datos sobre la historia natural y cultural de las especies con algún grado de domesticación es de gran importancia para comprender el alcance de la dispersión natural y antropogénica (Schaefer et al., 2009) y los efectos de la selección humana sobre diversidad y estructura genética poblacional.

La domesticación y los centros de origen de plantas domesticadas

La domesticación es un proceso co-evolutivo entre los seres humanos y otros organismos (Casas & Caballero, 1995; Clement, 1999b; Rindos, 1984). En plantas, mediante este proceso la selección humana actúa sobre los fenotipos de las poblaciones silvestres, promovidas, manejadas o cultivadas, resultando en un conjunto de modificaciones morfológicas, fisiológicas y genéticas que conllevan a una divergencia entre el taxón domesticado y su pariente silvestre (Purugganan & Fuller, 2009; Rindos, 1984). Estas modificaciones, reconocidas como rasgos o síndromes de

6

domesticación, surgen como resultado de las preferencias humanas por ciertas características en los fenotipos silvestres, siendo la selección humana inconsciente y consciente las principales fuerzas que llevan a la domesticación y posterior diversificación de las especies vegetales. Diversos autores han propuesto que el proceso de selección inconsciente dirigió los primeros pasos de la domesticación a través del movimiento de individuos, especialmente en especies anuales, desde sus hábitats silvestres originales hacia entornos antropogénicos (Zohary, 2004). Este proceso continúa ocurriendo en la actualidad en las comunidades rurales y es posible estudiar cómo se lleva a cabo y cuáles son sus consecuencias ecológicas y evolutivas (Clement et al., 2021). En la dispersión hacia esos nuevos ambientes, los organismos domesticados suelen perder y desarrollar nuevas adaptaciones para sobrevivir a las también nuevas condiciones ecológicas (Zohary, 2004). Eventualmente, las actividades de selección humana conscientes pudieron dirigirse a favorecer la reproducción (casi exclusiva) de aquellos fenotipos que se ajustan a las necesidades y deseos de las personas (Abbo et al., 2018; Svizzero, 2018; Weiss et al., 2006). Este proceso suele ser un esfuerzo multigeneracional que implica una cuidadosa selección y cultivo de plantas a lo largo del tiempo (Hancock, 2012; Harlan, 1975; Hawkes, 1998).

Las primeras investigaciones sobre dónde ocurrió la domesticación de plantas estuvieron regidas por la noción de centros de origen (Kantar et al., 2017). En perspectiva, Alphonse de Candolle (1886), quien concentró parte de sus esfuerzos al estudio del origen de organismos domesticados, reuniendo múltiples fuentes de evidencia. Afirmó en su hipótesis inicial que el centro de origen para una planta cultivada debería coincidir con la distribución natural y original de sus parientes silvestres. Durante sus investigaciones, notó una distribución desigual de especies cultivadas y propuso a China, el Suroeste Asiático y América tropical como tres principales centros de origen. Posteriormente, Nicolai Vavilov (1992) usando los criterios de una alta diversidad varietal, la coexistencia de ancestros silvestres con sus parientes domesticados y una larga historia de uso de los cultivos definió ocho centros de origen, puntualizados como áreas geográficamente pequeñas y asociadas con una "cuna de la agricultura" (Fuller et al., 2023). Estableció una diferenciación entre centros primarios, descritos como áreas de aparición inicial de la planta domesticada, y centros

secundarios de diversidad, los cuales albergan una alta diversidad de variedades domesticadas, pero se encuentran alejados de los centros donde se inició la domesticación. Con base en la propuesta de Vavilov, Harlan (1971) percibió que la distribución de plantas cultivadas y sus parientes silvestres era aún más compleja, demostrando que los centros de diversidad no constituyen necesariamente "centros de origen". Fundamentó su propuesta en evidencias contemporáneas e históricas, tanto botánicas como culturales, y propuso áreas denominadas "centros" y "no centros" de origen. Los "no centros", de acuerdo con este autor, son áreas extensas y difusas, a diferencia de los centros.

Con base en los avances de estos autores, el concepto sobre centros de origen de organismos domesticados se ha ido refinando y precisano por generaciones posteriores de científicos (Gepts, 1998; Hawkes, 1983, 1998; Zohary, 1970). Doebley et al. (2006) reconocen 10 centros geográficos de domesticación que difieren en su tamaño, número y diversidad de especies domesticadas, su potencial como fuente de recursos alimenticios y su desarrollo independiente. Por su parte, Fuller et al. (2023) identifican 13 a 24 centros geográficos de origen considerando la subdivisión de los centros originales propuestos por Vavilov para reflejar su naturaleza multinucleada. Afirman además que estos centros de diversidad de plantas domesticadas típicamente, pero no exclusivamente, corresponden a regiones de alta biodiversidad, donde podría haber surgido la agricultura. Destacan la asincronía del proceso de domesticación en estos centros de origen dado que evidencias genómicas, genéticas y arqueológicas han revelado que, aunque diversas especies pueden haberse originado en una región amplia, este proceso pudo suceder en diferentes momentos y lugares (Fuller et al., 2023). Durante las últimas tres décadas, los estudios sobre la historia evolutiva de las especies domesticadas han incorporado resultados genéticos, genómicos y arqueológicos, los cuales han evidenciado la complejidad del proceso evolutivo de las plantas cultivadas. En algunos casos este proceso se ha caracterizado por múltiples eventos independientes de domesticación en regiones geográficas extensas o difusas (Fuller et al., 2023; Kantar et al., 2017). Estas características revelan una historia de domesticación mucho más intrincada, con centros de origen difíciles de determinar, limitando así la universalidad del concepto (Kantar et al., 2017).

El Neotrópico como área de origen y de domesticación de plantas

El Neotrópico alberga una gran cantidad de especies vegetales que han sido domesticadas por los pueblos indígenas durante miles de años por lo que, en la actualidad, es considerado como una importante área de origen y diversificación de plantas de importancia cultural (Clement et al., 2021; Meyer et al., 2012). Tres partes del Neotrópico: Mesoamérica, los Andes y la Amazonía son fundamentales ya que muchos de los cultivos alimentarios más importantes del mundo, como el maíz, las papas, los chiles y los tomates, tienen su origen en estas áreas (Clement et al., 2021). Sin duda, la diversidad climática, geomorfológica, ecológica, geográfica y su historia biogeográfica han influido en la existencia de ambientes contrastantes entre estas tres zonas (Hughes et al., 2013). Estas condiciones proporcionaron una amplia gama de hábitats para diferentes especies de plantas, que eventualmente representaron alguna utilidad para las sociedades humanas. Sin embargo, un factor clave que hace del Neotrópico un área importante de domesticación es la gran diversidad vegetal y de culturas que se encuentra en la región (Clement et al., 2021). Los pueblos indígenas mesoamericanos, andinos y amazónicos desarrollaron técnicas agrícolas sofisticadas, incluyendo la rotación de cultivos, el riego y el manejo del suelo (Casas & Blancas, 2022; Pearsall & Piperno, 1998; Piperno, 2011; Piperno & Pearsall, 1998), que les permitieron cultivar estas plantas a gran escala. Esto condujo al desarrollo de sistemas agrícolas complejos que pudieron sustentar grandes poblaciones y producir excedentes de alimentos (Clement et al., 2021; Piperno, 2011). Además, los pueblos indígenas que habitaron (y habitan) en el Neotrópico tenían un profundo conocimiento del mundo natural y eran hábiles para reconocer, usar, manejar, seleccionar y cultivar plantas con características deseables (Casas et al., 2007; Clement et al., 2021; Hammer, 1984). Desarrollaron muchas de las variedades de plantas que todavía se usan hoy en día, incluidas variedades que son resistentes a plagas y enfermedades tienen rendimientos adecuados y se adaptan a diferentes condiciones de cultivo.

La domesticación de plantas neotropicales no solo tuvo un profundo impacto en las sociedades humanas contribuyendo al desarrollo de civilizaciones complejas, sino que, además, ha tenido un gran impacto en la producción mundial de alimentos puesto

que varios de los cultivos desarrollados en el Neotrópico ahora se cultivan en todo el mundo (Clement et al., 2021).

Domesticación de especies perennes: algunas particularidades

Durante la domesticación, algunos alelos que confieren fenotipos domesticados son el resultado de mutaciones que ocurrieron en el pariente silvestre y/o surgieron durante el proceso de domesticación (Jaenicke-Despres et al., 2003). Algunos alelos representan variación preexistente que, bajo el proceso de domesticación, fueron favorecidos. La velocidad de fijación de estos alelos depende en gran medida de la intensidad de la selección humana, sin dejar de lado la acción continua y progresiva del resto de las fuerzas evolutivas [endogamia, flujo genético, deriva génica (Casas et al., 2016)]. En el caso de las plantas, la domesticación a menudo se ha visto como un proceso rápido debido a fuertes presiones de selección direccional (Fuller, 2012). Esta suposición fue un paradigma útil para explicar los eventos y patrones evolutivos (síndromes de domesticación, la distribución geográfica de la diversidad genética, etc.) que se encuentran en varias plantas domesticadas, especialmente en plantas anuales (Purugganan, 2019; Vaughan et al., 2007), las cuales tienen rasgos de historia de vida que favorecen este proceso tales como generaciones que no se traslapan, tamaños efectivos de las poblaciones grandes y tiempos generacionales relativamente cortos. Sin embargo, arqueólogos y genetistas han sugerido que la domesticación no es lineal ni rápida (Clement et al., 2021; Fuller et al., 2012; Purugganan, 2019). Esta afirmación es especialmente cierta en cultivos perennes, ya que suelen ser especies leñosas de ciclo de vida largo con fases juveniles prolongadas y, por lo tanto, largos tiempos generacionales (Diez et al., 2015; Miller & Gross, 2011). Bajo domesticación, los árboles perennes a menudo se propagan tanto asexualmente (mediante diversas técnicas clonales o de injerto) como sexualmente (Diez et al., 2015; McClure et al., 2014; Miller & Gross, 2011). Esto tiene varias consecuencias genéticas. Muchos de los individuos utilizados en la propagación clonal son producto de cruzas. Por lo tanto, un número importante de heterócigos son mantenidos en las diferentes generaciones (Gaut et al., 2015). Esta combinación de reproducción clonal y exocruza promueve la generación de nuevas variaciones tanto sexuales como somáticas (Gaut et al., 2015; Miller & Gross,

2011) en las poblaciones perennes bajo manejo o cultivo. Por otra parte, las plantas perennes domesticadas tienen menos ciclos sexuales sobre los que puede actuar la selección, en comparación con lo que ocurre en las plantas anuales, lo que repercute en tasas de evolución más lentas (Miller & Gross, 2011).

Otro factor importante en el aumento o mantenimiento de niveles altos de variación genética es el proceso de introgresión, ya que este es una fuente potencial de nuevas variaciones durante los procesos de domesticación. Este proceso podría ser más frecuente que lo esperado en plantas perennes, ya que muchas poblaciones de especies cultivadas crecen en simpatría con sus parientes silvestres y no existen barreras para la hibridación. En tales casos, será necesario descartar explicaciones como la convergencia, la retención de la condición ancestral (alelos ancestrales) y los múltiples eventos de domesticación (Doebley, 1992).

La guayaba (*Psidium guajava* L.)

La tribu pantropical Myrteae (sensu Wilson et al., 2005) dentro de la familia Myrtaceae, comprende ~2500 especies dentro de 51 géneros (Govaerts et al., 2019; Lucas et al., 2007). *Psidium,* perteneciente a la tribu Myrteae, es el cuarto género más rico en especies de Myrtaceae (Govaerts et al., 2008). Actualmente se incluyen en este género al menos 90 especies, aunque este número puede incrementarse a 122 (Proença et al., 2022). Las mirtáceas fueron componentes destacados en las floras del sur de la Patagonia durante el Paleógeno temprano (Panti, 2016). Cuatro fósiles patagónicos datados entre 57 y 37 millones de años atrás (Paleoceno/Eoceno) han sido asignados a *Psidium* (Panti, 2016): *Psidium membranaceum* Engelhardt (1891), *P. araciforme* Berry (1938), *P. licciardoi* Hünicken (1967) y *Psidium* sp. (Panti, 2016). A pesar de las similitudes evidentes en la forma de la lámina y el patrón de venación con las hojas de las especies de *Psidium* actuales, aún persiste incertidumbre respecto a estas asignaciones.

El género *Psidium* presenta características complejas como lo son sus altas tasas de diversificación (Vasconcelos et al., 2017), su amplia distribución neotropical (Landrum, 2017; Richardson & Rejmánek, 2011), niveles variables de ploidía (Costa & Forni-Martins, 2007), y probables eventos de hibridación entre especies (Landrum et al.,

1995). Pocos estudios se han desarrollado con el objetivo de esclarecer sus relaciones evolutivas. Sin embargo, estudios filogenéticos basados en regiones de ADN nuclear y del cloroplasto han confirmado la monofilia del género con *Myrrhinium* como su grupo hermano (Flickinger et al., 2020; Lucas et al., 2005; Nadra et al., 2018). Así mismo, los análisis filogenéticos y de biogeografía histórica desarrollados por Vasconcelos et al. (2017), identificaron ocho clados bien soportados dentro del linaje neotropical principal en la tribu Myrteae, uno de los cuales se denomina grupo *Psidium* (con categoría taxonómica informal), y está conformando por los géneros *Psidium*, *Mosiera*, *Myrrhynium* y una especie del género polifilético *Calyptrogenia*. Los patrones biogeográficos identificados en dicho estudio mostraron a América del Sur como el área más probable de distribución ancestral tanto para el linaje neotropical principal como para el grupo *Psidium*, con evidencias de movimientos desde América del Sur hacia Centro y Norte América. Adicionalmente, en un estudio reciente Proença et al. (2022) proporcionaron una revisión de la biología, morfología y ecología de *Psidium*, proponiendo un árbol filogenético, una clasificación infragenérica y una lista de especies. Con base en los resultados de este estudio, *Psidium* es considerado un género monofilético con cuatro clados principales reconocidos como secciones. La sección *Psidium,* a la que pertenece *P. guajava* (guayaba), es hermana del resto de las secciones del género y está ampliamente distribuida.

Psidium presenta una distribución naturalmente neotropical que se extiende desde el norte de México hasta el centro-este de Argentina, incluyendo las Antillas (Landrum, 2017; Proença et al., 2022). Los biomas del Bosque Costero Atlántico, Cerrado y Caatinga en Brasil, con más de 50 especies, son reconocidos como el centro de diversidad del género (Landrum, 2017, 2021). Por lo tanto, las especies de *Psidium* centro y norteamericanas son consideradas como arribos relativamente recientes en escaka geológica, y estrechamente relacionadas con las especies suramericanas (Landrum, 2017), a diferencia de las especies endémicas del Caribe, consideradas como colonizaciones geológicamente antiguas (Landrum, 2017).

Las especies de *Psidium* producen frutos comestibles, los cuales se consumen de forma directa o son procesados por los humanos en diversos alimentos por lo que es reconocido como un género de económicamente importante (Landrum, 2017). *Psidium*

guajava L. La guayaba, es la especie del género más ampliamente reconocida y cultivada (Landrum, 2017; Proença et al., 2022). Las hojas y frutos de la guayaba son comúnmente usados como alimento y medicina en el continente americano (Gutiérrez et al., 2008; Landrum, 2017; Patiño, 2002). Sin embargo, hoy, además de algunos países neotropicales como Brasil y México, también se reconoce a la India, Pakistán, China y Egipto como importantes países cultivadores y exportadores de esta especie, lo que demostra la importancia económica global de la planta (Altendorf, 2018, 2019).

Los registros históricos de cronistas europeos del siglo XVI mencionan la presencia de guayabas tanto en Mesoamérica como en Sudamérica, en poblaciones silvestres y cultivadas, como se documenta en el trabajo de Patiño en 2002. También se reportó la presencia de guayabas en las Antillas (Patiño, 2002). No obstante, hallazgos arqueológicos más antiguo, datan la presencia de la guayaba hace aproximadamente entre 9490 y 6505 años antes del presente (cal. AP) en la región del suroeste de la Amazonia (Watling et al., 2018). También se han encontrado evidencias de la guayaba en asentamientos humanos en la costa peruana que datan de alrededor de 7000 años cal. AP (Cárdenas, 1999); Colombia, las Antillas y Centro América (ca. 2600 cal. AP; (Hammond et al., 1995; Kosztura, 2020; Miksicek, 1991; Newsom & Wing, 2004); en bosques secos de México (ca. 670 cal. AP; Smith, 1967).

La distribución actual de la guayaba se extiende desde México y las Antillas hasta el noroeste de Argentina (Landrum, 2017; Proença et al., 2022). Habita principalmente en áreas alteradas, como bordes de caminos, pastizales y huertos, desde cerca del nivel del mar hasta los 1600 m de altitud (Landrum, 2017). También es posible encontrar guayabas ferales en áreas naturales abiertas y sobre las márgenes de pequeños ríos (observación pers.). Es una especie heliófila. Sus semillas necesitan un fotoperíodo largo (>10 h) con temperaturas que oscilan entre 20 y 30°C para alcanzar > 90% de germinación (Sugahara & Takaki, 2004). La exposición de los árboles maduros a la luz solar directa garantiza una fructificación abundante con un mayor crecimiento de las ramas (Paull & Duarte, 2012), preferiblemente en temperaturas medias 23º a 28°C (Menzel, 1985). Hoy en día, las semillas de guayaba pueden ser dispersadas de manera efectiva por aves y mamíferos (p. ej., murciélagos, monos, ungulados,

marsupiales) (Gressler et al., 2006; Torres & Gutiérrez, 2018) y por algunos animales domésticos como las vacas (Herrera, 2013; Somarriba, 1985, 1986), y por los humanos.

Aunque la guayaba es un cultivo económicamente importante, el origen de la especie, su proceso de dispersión en América, su grado de domesticación y la historia evolutiva bajo domesticación son en su mayoría conjeturas. Varios autores han sugerido algunas áreas en Suramérica como el norte de América del Sur (Aranguren et al., 2008; Risterucci et al., 2005), el norte de la Amazonia (Hastorf, 2006), o el noreste de Brasil (Clement, 1999a; Pearsall, 2008) como el área de origen para la especie, mientras que otros han propuesto Mesoamérica (Brand, 1939; Nakasone & Paull, 1998). De igual forma, hasta la fecha, ninguna investigación se ha enfocado en el estudio sobre el proceso de domesticación de la guayaba.

En esta perspectiva, la guayaba representa un sistema interesante para el estudio sobre la historia evolutiva y de domesticación de un frutal perenne distribuido a escala continental. Este trabajo constituye un primer abordaje, el cual tiene como objetivo principal: Inferir la historia evolutiva, el proceso de dispersión y domesticación de *P. guajava* en América. Para ello, esta investigación se divide de la siguiente forma:

El capítulo I lo conforma el artículo publicado **The Taming of *Psidium guajava*: Natural and Cultural History of a Neotropical Fruit** (Arévalo-Marín et al. 2021). En este se examina y sistematiza información biológica, molecular, arqueológica e histórica sobre la guayaba. Se identifican los principales desafíos que hasta ahora han limitado nuestra comprensión sobre la domesticación de la guayaba, y se presentan cinco propuestas hipotéticas que pueden guiar las futuras investigaciones sobre las siguientes preguntas centrales que aún no tienen respuesta: (1) ¿Cuáles son sus parientes más cercanos? (2) ¿Cuál es el área de origen de la guayaba como especie? (3) ¿Cuál(es) es(son) el(los) área(s) de domesticación de la guayaba? (4) ¿Cómo llegó la guayaba a su distribución actual? (5) ¿Qué papel jugaron los conquistadores europeos y sus animales domésticos en la dispersión y el flujo de genes de la guayaba después del contacto?

El capítulo II está constituido por el artículo (aún no enviado) **Evolutionary history of *Psidium guajava* (Myrtaceae) in the Americas inferred by chloroplast and nuclear markers**, en el cual se realizó una descripción general del contexto

evolutivo de *P. guajava* en las Américas. Se investiga el área geográfica de origen de la especie, relaciones genealógicas y tiempos de divergencia, e historia demográfica de la guayaba. Los objetivos específicos de este estudio fueron: 1) proporcionar información detallada sobre el patrón de diversidad genética de la guayaba en las Américas, 2) identificar si existe una divergencia genética entre las muestras de guayaba de las regiones de América del Sur, Mesoamérica y las Antillas, 3) dilucidar las relaciones filogenéticas y los patrones filogeográficos entre los linajes de guayaba, y 4) determinar si la distribución de *P. guajava* está relacionada principalmente con eventos biogeográficos o procesos antropogénicos.

El capítulo III incluye el artículo (aún no enviado) **Genetic differentiation of guava across the Americas**. En este se caracterizar la estructura genética de la guayaba en casi toda su distribución geográfica con base en marcadores SSR, para responder las siguientes preguntas: (a) ¿cuál es el nivel de variabilidad genética de la guayaba entre los genotipos muestreados? (b) ¿Cómo se estructura esta diversidad? (c) ¿Existe aislamiento por distancia entre poblaciones? (d) ¿Existe alguna evidencia de eventos de domesticación?

Finalmente, en la Discusión General de esta tesis se analizan de forma integral los patrones naturales, biogeográficos y antropogénicos que han determinado la distribución espacial y genética de la especie. Asimismo, se plantean algunas preguntas de investigación interesantes que podrían ser desarrolladas en investigaciones futuras.

Métodos para el estudio sobre el origen y el proceso domesticación en especies perennes

Filogeografía

La domesticación puede ser entendida como un proceso de diferenciación poblacional, guiada principalmente por la selección humana, que puede resultar en el favorecimiento de algunos linajes e influenciar diferentes grados de divergencia entre los organismos domesticados y silvestres y entre los organismos domesticados entre sí. Entender los patrones geográficos de la estructura genealógica de una especie en su distribución y distinguir entre eventos históricos y actuales de flujo génico, puede ayudar a discernir

entre el papel del ser humano y los eventos naturales que influyeron en el movimiento de una especie en el espacio (Parra & Casas, 2016). Para esclarecer esta historia, la filogeografía es una herramienta de investigación especialmente útil, pues aporta un cuerpo de conocimiento adecuado para el estudio de las áreas de origen y rutas de difusión. Esta disciplina estudia los principios que rigen las distribuciones espaciales de los linajes genealógicos especialmente dentro y entre especies estrechamente relacionadas (Avise, 2009; Avise et al., 1987). Utiliza herramientas similares a las que se emplean en la reconstrucción de árboles o especies buscando reconstruir genealogías de genes. Con ello busca generar aproximaciones históricas que faciliten la interpretación de los procesos evolutivos que influyeron sobre las poblaciones y que dieron lugar a la diversidad genética observada y su estado actual de distribución (Parra & Casas, 2016). No obstante, la evolución intraespecífica de genes no puede ser representada por árboles bifurcados ya que la inferencia cladística de relaciones filogenéticas requiere que las variantes genéticas no formen linajes reticulados (Schaal et al., 1998). Una alternativa es la construcción de redes de haplotipos las cuales representan de manera más acertada la presencia de haplotipos ancestrales, divergencias, reticulaciones y homoplasia (Aguirre-Dugua & González-Rodríguez, 2016).

Esta disciplina se basa también en los principios de coalescencia y de métodos de cómputo bayesiano aproximado para evaluar formal y estadísticamente las predicciones filogeográficas (De Maio et al., 2015; Domínguez, 2007). La reconstrucción de las relaciones entre los alelos es realizada desde el presente hacia el pasado hasta lograr una convergencia en su alelo ancestral; revelando la historia demográfica de las poblaciones a través de la estimación de factores como tamaño efectivo poblacional, cantidad de flujo génico, escenarios de divergencia y los efectos de la deriva génica y la selección (Avise, 2009; Avise et al., 1987; De Maio et al., 2015).

Métodos Poblacionales

Conocer los niveles de variación genética y su distribución dentro y entre poblaciones es importante para comprender el origen y evolución de las poblaciones vegetales. En esta perspectiva, analizar la distribución espacial de la variabilidad genética e identificar

los sitios de alta diversidad en poblaciones silvestres y manejadas, contribuiría a identificar el área en donde se originó la domesticación de una planta cultivada (Doebley, 1992; Doebley et al., 2006). El nivel de diferenciación y variación en las frecuencias génicas evidenciaría el grado de intensidad de manejo y de domesticación. Los análisis de genética de poblaciones se basan en el cálculo de frecuencias alélicas y genotípicas de las poblaciones bajo estudio, para estimar posteriormente niveles de diversidad genética y grado de diferenciación entre las poblaciones de una especie (Rendón & Núñez Farfán, 1998). Debido a que los seres humanos frecuentemente interactuaron con subconjuntos de la diversidad genética existente en las poblaciones de plantas, determinando cuellos de botella, se considera que los procesos de deriva génica han sido importantes en la evolución bajo domesticación. Generalmente se espera que la diversidad alélica en el cultivo sea un subconjunto de la que se encuentra en la población silvestre de la cual se derivó. Por lo tanto, si las poblaciones del progenitor silvestre existen en la ubicación geográfica en la que se produjo la domesticación, el origen geográfico de la domesticación se puede localizar en una determinada población o región, identificando los cuellos de botella referidos (Gross & Olsen, 2010; Meyer et al., 2012; Meyer & Purugganan, 2013).

Los estadísticos tradicionales de la genética de poblaciones tales como el índice de fijación (F), los estadísticos de F de Wright, heterocigosidad, los alelos privados, son el primer paso para comprender cómo está repartida la diversidad genética. Posteriormente, los métodos de clasificación y asignación permiten evaluar cómo se distribuye la variedad genética en el espacio, proveyendo información importante para el estudio sobre origen y difusión de la domesticación de plantas cultivadas. Los primeros se caracterizan por asignar los individuos a categorías predefinidas (Manel et al., 2005). En tales casos se incluyen los análisis discriminantes (AD) y los análisis de componentes principales (ACP). Estos análisis permiten resumir la variación presente en los loci analizados (ACP) o caracterizar la variación genética dentro y entre grupos (AD). Los métodos bayesianos usan la información genética para estimar la probabilidad de que un individuo pertenezca a un grupo o población, con base en su genotipo y la frecuencia alélica, sin asumir algún tipo de población predefinida.

Referencias

Abbo, S., Nelson, M. N., Warkentin, T., & Weeden, N. F. (2018). *Domestication of Pea (Pisum sativum L.): The Case of the Abyssinian Pea.* *9*(April), 1–11. https://doi.org/10.3389/fpls.2018.00515

Aguirre-Dugua, X., & González-Rodríguez, A. (2016). Phylogeographical approaches to the study of plant domestication, with special emphasis on perennial plants. In *Ethnobotany of Mexico* (pp. 319–366). Springer.

Altendorf, S. (2018). Minor tropical fruits: Mainstreaming a niche market. In *Food Outlook Biannual Report on Global Food Markets* (Issue July, pp. 67–74). http://www.fao.org/3/a-I8080e.pdf

Altendorf, S. (2019). *Major tropical fruits market review 2018.* http://www.fao.org/3/ca5692en/CA5692EN.pdf

Aranguren, Y., Briceño, A., & Fermin, G. (2008). Assessment of the variability of Venezuelan guava landraces by microsatellites. In W. Rohde & G. Fermin (Eds.), *II International Symposium on Guava and Other Myrtaceae* (Vol. 849, pp. 147–154).

Avise, J. C. (2000). *Phylogeography: the history and formation of species.* Harvard university press.

Avise, J. C. (2009). Phylogeography: Retrospect and prospect. *Journal of Biogeography*, *36*(1), 3–15. https://doi.org/10.1111/j.1365-2699.2008.02032.x

Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A., & Saunders, N. C. (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, *18*(1), 489–522.

Brand, D. D. (1939). The Origin and Early Distribution of New World Cultivated Plants. *Agricultural History*, *13*(2), 109–117.

Candolle, A. (1886). *Origin of cultivated plants.* Trench, Paul.

Cárdenas, M. (1999). El Periodo Precerámico en el valle de Chao. *Boletín de Arqueología PUCP*, *3*, 141–169.

Casas, A., & Blancas, J. (2022). Human Cultures and Plant Diversity in the Mountains of Mexico: An Introduction. In *Ethnobotany of the Mountain Regions of Mexico* (pp. 1–22). Springer.

Casas, A., & Caballero, J. (1995). Domesticación de plantas y el origen de la agricultura en Mesoamérica. *Ciencias*, *040*.

Casas, A., Otero-Arnaiz, A., Perez-Negron, E., & Valiente-Banuet, A. (2007). *In situ* management and domestication of plants in Mesoamerica. *Ann Bot*, *100*(5), 1101–1115. https://doi.org/10.1093/aob/mcm126

Casas, A., Torres-Guevara, J., & Parra-Rondinel, F. (2016). *Domesticación en el continente americano* (Vol. 1). Universidad Nacional Autónoma de Méxic.

Clement, C. R. (1999a). 1492 and the loss of Amazonian crop genetic resources. I. The relation between domestication and human population decline. *Economic Botany*, *53*(2), 188–202. https://doi.org/10.1007/bf02866498

Clement, C. R. (1999b). 1492 and the loss of amazonian crop genetic resources. II. Crop biogeography at contact. *Economic Botany*, *53*(2), 203–216. https://doi.org/10.1007/bf02866499

Clement, C. R., Casas, A., Parra-Rondinel, F. A., Levis, C., Peroni, N., Hanazaki, N., Cortés-Zárraga, L., Rangel-Landa, S., Alves, R. P., Ferreira, M. J., Cassino, M. F., Coelho, S. D., Cruz-Soriano, A., Pancorbo-Olivera, M., Blancas, J., Martínez-Ballesté, A., Lemes, G., Lotero-Velásquez, E., Bertin, V. M., & Mazzochini, G. G. (2021). Disentangling Domestication from Food Production Systems in the Neotropics. *Quaternary*, *4*(1), 4. https://doi.org/10.3390/quat4010004

Costa, I. R., & Forni-Martins, E. R. (2007). Karyotype analysis in South American species of Myrtaceae. *Botanical Journal of the Linnean Society*, *155*(155), 571–580. https://doi.org/10.1111/j.1095-8339.2007.00704.x

De Maio, N., Wu, C.-H., O'Reilly, K. M., & Wilson, D. (2015). New routes to phylogeography: a Bayesian structured coalescent approximation. *PLoS Genetics*, *11*(8), e1005421.

Diez, C. M., Trujillo, I., Martinez-Urdiroz, N., Barranco, D., Rallo, L., Marfil, P., & Gaut, B. S. (2015). Olive domestication and diversification in the Mediterranean Basin. *New Phytologist*, *206*(1), 436–447. https://doi.org/10.1111/nph.13181

Doebley, J. (1992). Molecular systematics and crop evolution. In *Molecular systematics of plants* (pp. 202–222). Springer.

Doebley, J. F., Gaut, B. S., & Smith, B. D. (2006). The molecular genetics of crop domestication. *Cell*, *127*. https://doi.org/10.1016/j.cell.2006.12.006

Domínguez, E. (2007). Filogeografía y vertebrados. *Ecología Molecular*, 441–466. http://www2.inecc.gob.mx/publicaciones/libros/530/cap14.pdf

Flickinger, J. A., Jestrow, B., Oviedo Prieto, R., Santiago-Valentín, E., Sustache-Sustache, J., Jiménez-Rodríguez, F., Campbell, K. C. S. E., & Francisco-Ortega, J. (2020). A phylogenetic survey of Myrtaceae in the Greater Antilles with nomenclatural changes for some endemic species. *Taxon*, *69*(3), 448–480. https://doi.org/10.1002/tax.12263

Fuller, D. Q. (2012). New archaeobotanical information on plant domestication from macro-remains: tracking the evolution of domestication syndrome traits. In P. Gepts, T. R. Famula, R. L. Bettinger, S. B. Brush, A. B. Damania, P. E. McGuire, & C. O. Qualset (Eds.), *Biodiversity in Agriculture: Domestication, Evolution, and Sustainability* (p. 110). Cambriedge University Press.

Fuller, D. Q., Asouti, E., & Purugganan, M. D. (2012). Cultivation as slow evolutionary entanglement: Comparative data on rate and sequence of domestication. *Vegetation History and Archaeobotany*, *21*(2), 131–145. https://doi.org/10.1007/s00334-011-0329-8

Fuller, D. Q., Denham, T., & Allaby, R. (2023). Plant domestication and agricultural ecologies. In *Current Biology* (Vol. 33, Issue 11, pp. R636–R649). Cell Press. https://doi.org/10.1016/j.cub.2023.04.038

Gaut, B. S., Díez, C. M., & Morrell, P. L. (2015). Genomics and the Contrasting Dynamics of Annual and Perennial Domestication. *Trends in Genetics*, *31*(12), 709–719. https://doi.org/10.1016/j.tig.2015.10.002

Gepts, P. (1998). Origin and evolution of common bean: past events and recent trends. *HortScience*, *33*(7), 1124–1130.

Govaerts, R., Sobral, M., Ashton, P., Barrie, F., Holst, B. K., Landrum, L. L., Matsumoto, K., Mazine, F. F., Lughadha, E. N., Proenca, C., Soares-Silva, L. H., Wilson, P. G., & Lucas, E. (2019). *World Checklist of Selected Plant Families*. World Checklist of Myrtaceae. Facilitated by the Royal Botanic Gardens, Kew. http://wcsp.science.kew.org/

Govaerts, R., Sobral, M., Ashton, P., Barrie, F., Holst, B. K., Landrum, L. L., Matsumoto, K., Mazine, F. F., Lughadha, E. N., & Proneça, C. (2008). *World checklist of Myrtaceae.* Royal Botanic Gardens.

Gressler, E., Pizo, M. A., & Morellato, L. P. C. (2006). Polinização e dispersão de sementes em Myrtaceae do Brasil. *Revista Brasileira de Botânica*, *29*(4), 509–530. https://doi.org/10.1590/s0100-84042006000400002

Gross, B. L., & Olsen, K. M. (2010). Genetic perspectives on crop domestication. *Trends in Plant Science*, *15*(9), 529–537. https://doi.org/10.1016/j.tplants.2010.05.008

Gutiérrez, R. M., Mitchell, S., & Solis, R. V. (2008). *Psidium guajava*: a review of its traditional uses, phytochemistry and pharmacology. *J Ethnopharmacol*, *117*(1), 1–27. https://doi.org/10.1016/j.jep.2008.01.025

Hammer, K. (1984). The domestication syndrome. *Die Kulturpflanze, 32*(1), 11–34.

Hammond, N., Clarke, A., & Donaghey, S. (1995). The Long Goodbye: Middle Preclassic Maya Archaeology at Cuello, Belize. *Latin American Antiquity*, *6*(2), 120–128.

Hancock, J. F. (2012). The dynamics of plant domestication. In *Plant evolution and the origin of crop species* (pp. 114–131). CABI Wallingford UK.

Harlan, J. R. (1971). Origins: Centers and Noncenters. *Science*, *174*(4008), 468–474. http://science.sciencemag.org/content/174/4008/468.abstract

Harlan, J. R. (1975). *Crops and Man*. American Society of Agronomy, Madison, Wisconsin.

Hastorf, C. A. (2006). Domesticated food and society in early coastal Peru. In W. Baleé & C. L. Erickson (Eds.), *Time and Complexity in Historical ecology: Studies in the Neotropical Lowlands* (pp. 87–126). Columbia University Press.

Hawkes, J. G. (1983). The diversity of crop plants. *The Diversity of Crop Plants.*

Hawkes, J. G. (1998). Back to Vavilov: Why Were Plants Domesticated in. *The Origins of Agriculture and Crop Domestication: Proceedings of the Harlan Symposium, 10-14 May 1997, Aleppo, Syria*, *21*, 5.

Herrera, M. X. A. (2013). *Posibles dispersores de Psidium guajava en la Isla San Cristóbal , Galápagos - Ecuador*. Universidad San Francisco de Quito.

Hughes, C. E., Pennington, R. T., & Antonelli, A. (2013). Neotropical plant evolution: assembling the big picture. *Botanical Journal of the Linnean Society*, *171*(1), 1–18.

Jaenicke-Despres, V., Buckler, E. S., Smith, B. D., Gilbert, M. T. P., Cooper, A., Doebley, J., & Paabo, S. (2003). Early allelic selection in maize as revealed by ancient DNA. *Science*, *302*(5648), 1206–1208.

Kantar, M. B., Nashoba, A. R., Anderson, J. E., Blackman, B. K., & Rieseberg, L. H. (2017). The Genetics and Genomics of Plant Domestication. *BioScience*, *67*(11), 971–982. https://doi.org/10.1093/biosci/bix114

Kosztura, N. J. M. (2020). *Cultivares y plantas silvestres en las Terras Pretas de la Amazonia colombiana (Guaviare-Colombia) [dissertation/marster''s thesis]*. Universidad Nacional de Colombia. https://repositorio.unal.edu.co/handle/unal/77958

Landrum, L. R. (2017). The genus *Psidium* (Myrtaceae) in the state of Bahia, Brazil. *Canotia*, *13*, 1–101.

Landrum, L. R. (2021). *Psidium guajava* L.: taxonomy, relatives and possible origin. In S. Mitra (Ed.), *Guava: botany, production and uses* (pp. 1–21). Cab International. https://doi.org/10.1079/9781789247022.0001

Landrum, L. R., Clark, W. D., Sharp, W. P., & Brendecke, J. (1995). Hybridization between Psidium guajava and P. guineense (Myrtaceae). *Economic Botany*, *49*(2), 153–161.

Lucas, E. J., Belsham, S. R., Nic Lughadha, E. M., Orlovich, D. A., Sakuragui, C. M., Chase, M. W., & Wilson, P. G. (2005). Phylogenetic patterns in the fleshy-fruited Myrtaceae - Preliminary molecular evidence. *Plant Systematics and Evolution*, *251*(1), 35–51. https://doi.org/10.1007/s00606-004-0164-9

Lucas, E. J., Harris, S. A., Mazine, F. F., Belsham, S. R., Nic Lughadha, E. M., Telford, A., Gasson, P. E., Chase, M. W., Lughadha, M. N., Telford, A., Gasson, P. E., Chase, M. W., Telford, A., Gasson, P. E., Chase, M. W., Nic Lughadha, E. M., Telford, A., Gasson, P. E., & Chase, M. W. (2007). Suprageneric Phylogenetics of Myrteae, the Generically Richest Tribe in Myrtaceae. *TAXON*, *56*(4), 1105–1128. https://doi.org/10.2307/25065906

Manel, S., Gaggiotti, O. E., & Waples, R. S. (2005). Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology & Evolution*, *20*(3), 136–142.

McClure, K. A., Sawler, J., Gardner, K. M., Money, D., & Myles, S. (2014). Genomics: A potential panacea for the perennial problem. *American Journal of Botany*, *101*(10), 1780–1790. https://doi.org/10.3732/ajb.1400143

Menken, S. B. J., Smit, E., & Nijs, H. (J) C. M. Den. (1995). Genetical population structure in plants: gene flow between diploid sexual and triploid asexual dandelions (Taraxacum section Ruderalia). *Evolution*, *49*(6), 1108–1118.

Menzel, C. M. (1985). Guava: an exotic fruit with potential in Queensland. *Queensland Agricultural Journal, 111*(2), 93–98.

Meyer, R. S., DuVal, A. E., & Jensen, H. R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *The New Phytologist, 196*(1), 29–48. https://doi.org/10.1111/j.1469-8137.2012.04253.x

Meyer, R. S., & Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nature Reviews Genetics, 14*(12), 840–852.

Miksicek, C. H. (1991). The Ecology and Economy of Cuello. The Natural and Cultural Landscape Preclassic Cuello. In N. Hammond (Ed.), *Cuello and early Maya community in Belize.* Cambridge University Press.

Miller, A. J., & Gross, B. L. (2011). From forest to field: perennial fruit crop domestication. *American Journal of Botany, 98*(9), 1389–1414.

Nadra, M. G., Giannini, N. P., Acosta, J. M., & Aagesen, L. (2018). Evolution of pollination by frugivorous birds in Neotropical Myrtaceae. *PeerJ, 6,* e5426. https://doi.org/10.7717/peerj.5426

Nakasone, H. Y., & Paull, R. E. (1998). *Tropical fruits.* Cab International.

Newsom, L. A., & Wing, E. S. (2004). *On land and sea: Native American uses of biological resources in the West Indies.* University of Alabama Press.

Panti, C. (2016). Myrtaceae fossil leaves from the Río Turbio Formation (Middle Eocene), Santa Cruz Province, Argentina. *Historical Biology, 28*(4), 459–469. https://doi.org/10.1080/08912963.2014.976635

Parra, F., & Casas, A. (2016). Origen y difusión de la domesticación y la agricultura en el Nuevo Mundo. In A. Casas, J. Torres-Guevara, & F. Parra (Eds.), *Domesticación en el continente americano. Volumen 1. Manejo de la biodiversidad y evolución dirigida por las culturas del Nuevo Mundo* (pp. 159–187).

Patiño, V. M. (2002). *Historia y dispersión de los frutales nativos del neotrópico.* Centro Internacional de Agricultura Tropical. https://books.google.com.mx/books?id=U12m7M2VMmsC

Paull, R. E., & Duarte, O. (2012). Guava. In O. Paull, R. E., Duarte (Ed.), *Tropical fruits, Volume 2* (pp. 91–122). CABI International.

Pearsall, D. M. (2008). Plant domestication and the shift to agriculture in the Andes. In H. Silverman & W. H. Isbell (Eds.), *The handbook of South American archaeology* (pp. 105–120). Springer Science & Business Media.

Pearsall, D. M., & Piperno, D. (1998). *The origins of agriculture in the lowland Neotropics.* Academic Press.

Piperno, D. R. (2011). The Origins of Plant Cultivation and Domestication in the New World Tropics. *Current Anthropology, 52*(S4), S453–S470. https://doi.org/10.1086/659998

Piperno, D. R., & Pearsall, D. M. (1998). From Small Scale Horticulture to the Formative Period: The Development of Agriculture. In *The origins of agriculture in the Lowland Neotropics* (pp. 243–320). Academic Press.

Proença, C. E. B., Tuler, A. C., Lucas, E. J., Vasconcelos, T. N. da C., de Faria, J. E. Q., Staggemeier, V. G., de-Carvalho, P. S., Forni-Martins, E. R., Inglis, P. W., da Mata, L. R., & da Costa, I. R. (2022). Diversity, phylogeny and evolution of the rapidly evolving genus *Psidium* L. (Myrtaceae, Myrteae). *Annals of Botany, 129*(4), 367–388. https://doi.org/10.1093/aob/mcac005

Purugganan, M. D. (2019). Evolutionary Insights into the Nature of Plant Domestication. *Current Biology, 29*(14), R705–R714. https://doi.org/10.1016/j.cub.2019.05.053

Purugganan, M. D., & Fuller, D. Q. (2009). The nature of selection during plant domestication. *Nature, 457*(7231), 843–848. https://doi.org/10.1038/nature07895

Rendón, B., & Núñez Farfán, J. (1998). *Genetica evolutiva del proceso de domesticacion en plantas.* Universidad Nacional Autonoma de Mexico. Facultad de Ciencias, Mexico (Mexico).

Richardson, D. M., & Rejmánek, M. (2011). Trees and shrubs as invasive alien species - a global review. *Diversity and Distributions*, *17*(5), 788–809. https://doi.org/10.1111/j.1472-4642.2011.00782.x

Rindos, D. (1984). The Evolution of Domestication. In *The Origins of Agriculture*. https://doi.org/10.1016/B978-0-12-589281-0.50009-9

Risterucci, A. M., Duval, M. F., Rohde, W., & Billotte, N. (2005). Isolation and characterization of microsatellite loci from Psidium guajava L. A. *Molecular Ecology Notes*, *5*(4), 745–748. https://doi.org/10.1111/j.1471-8286.2005.01050.x

Schaal, B. A., Hayworth, D. A., Olsen, K. M., Rauscher, J. T., & Smith, W. A. (1998). Phylogeographic studies in plants: problems and prospects. *Molecular Ecology*, *7*(4), 465–474.

Schaefer, H., Heibl, C., & Renner, S. S. (2009). Gourds afloat: a dated phylogeny reveals an Asian origin of the gourd family (Cucurbitaceae) and numerous oversea dispersal events. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1658), 843–851.

Smith, C. E. (1967). Plant remains. In D. S. Byers (Ed.), *The Prehistory of Tehuacan Valley. Volume One: Environment and Subsistence* (pp. 220–255). University of Texas Press.

Somarriba, E. (1985). Árboles de Guayaba (*Psidium guajava* L.) en Pastizales II: Consumo de Fruta y Dispersión de Semillas.pdf. *Revista Interamericanda de Ciencias Agricolas*, *35*(4), 329–332.

Somarriba, E. (1986). Effects of livestock on seed germination of guava (*Psidiurn guajava* L.). *Agroforestry Systems*, *4*, 233–238.

Sugahara, V. Y., & Takaki, M. (2004). Effect of light and temperature on seed germination in guava (Psidium guajava L. - Myrtaceae). *Seed Science and Technology*, *32*, 759–764. https://doi.org/10.15258/sst.2008.36.3.30

Svizzero, S. (2018). Plant domestication more rapid under optimizing behavior. *Journal of Bioeconomics*, *20*(3), 287–308. https://doi.org/10.1007/s10818-018-9272-4

Torres, M. de L., & Gutiérrez, B. (2018). A Preliminary Assessment of the Genetic Diversity and Population Structure of Guava, Psidium guajava, in San Cristobal. In *Understanding Invasive Species in the Galapagos Islands* (pp. 3–17). https://doi.org/10.1007/978-3-319-67177-2_1

Vasconcelos, T. N., Proenca, C. E., Ahmad, B., Aguilar, D. S., Aguilar, R., Amorim, B. S., Campbell, K., Costa, I. R., De-Carvalho, P. S., Faria, J. E., Giaretta, A., Kooij, P. W., Lima, D. F., Mazine, F. F., Peguero, B., Prenner, G., Santos, M. F., Soewarto, J., Wingler, A., & Lucas, E. J. (2017). Myrteae phylogeny, calibration, biogeography and diversification patterns: Increased understanding in the most species rich tribe of Myrtaceae. *Mol Phylogenet Evol*, *109*, 113–137. https://doi.org/10.1016/j.ympev.2017.01.002

Vaughan, D. A., Balazs, E., & Heslop-Harrison, J. S. (2007). From crop domestication to super-domestication. *Annals of Botany*, *100*(5), 893–901.

Vavilov, N. I. (1992). On the origin of cultivated plants. In Vladimir Filimonovich Dorofeev (Ed.), *Origin and geography of cultivated plants* (pp. 14–21). Cambridge University Press.

Villaseñor, J. L. (2016). Checklist of the native vascular plants of Mexico. *Revista Mexicana de Biodiversidad*, *87*(3), 559–902.

Watling, J., Shock, M. P., Mongeló, G. Z., Almeida, F. O., Kater, T., De Oliveira, P. E., & Neves, E. G. (2018). Direct archaeological evidence for Southwestern Amazonia as an early plant domestication and food production centre. *Plos One*, *13*(7), e0199868. https://doi.org/10.1371/journal.pone.0199868

Weiss, E., Kislev, M. E., & Hartmann, A. (2006). Autonomous Cultivation Before Domestication. *Science*, *312*(5780), 1608 LP – 1610. https://doi.org/10.1126/science.1127235

Wilson, P. G., O'Brien, M. M., Heslewood, M. M., & Quinn, C. J. (2005). Relationships within Myrtaceae sensu lato based on a matK phylogeny. *Plant Systematics and Evolution*, *251*(1), 3–19. https://doi.org/10.1007/s00606-004-0162-y

Zohary, D. (1970). Centres of diversity and centers of origin. *Centres of Diversity and Centers of Origin.*

Zohary, D. (2004). Unconscious selection and the evolution of domesticated plants. *Economic Botany*, *58*(1), 5–10. https://doi.org/10.1663/0013-0001(2004)058[0005:USATEO]2.0.CO;2

CAPITULO I

The Taming of *Psidium guajava*: Natural and Cultural History of a Neotropical Fruit

Cita completa:

Arévalo-Marín E, Casas A, Landrum L, Shock MP, Alvarado-Sizzo H, Ruiz-Sanchez E and Clement CR (2021) The Taming of *Psidium guajava:* Natural and Cultural History of a Neotropical Fruit. Front. Plant Sci. 12:714763. doi: 10.3389/fpls.2021.714763

Check for updates

The Taming of *Psidium guajava*: Natural and Cultural History of a Neotropical Fruit

Edna Arévalo-Marín[1,2], Alejandro Casas[1]*, Leslie Landrum[3], Myrtle P. Shock[4], Hernán Alvarado-Sizzo[5], Eduardo Ruiz-Sanchez[6] and Charles R. Clement[7]

[1] Instituto de Investigaciones en Ecosistemas y Sustentabilidad (IIES), Universidad Nacional Autónoma de México, Morelia, Mexico, [2] Posgrado en Ciencias Biológicas, Unidad de Posgrado, Universidad Nacional Autónoma de México, Mexico City, Mexico, [3] Natural History Collections, School of Life Sciences, Arizona State University, Tempe, AZ, United States, [4] Programa de Antropologia e Arqueologia, Instituto de Ciências da Sociedade, Universidade Federal do Oeste do Pará, Santarém, Brazil, [5] Laboratorio de Biogeografía y Sistemática, Departamento de Biología Evolutiva, Facultad de Ciencias, Universidad Nacional Autónoma de México, Ciudad Universitaria, Mexico City, Mexico, [6] Departamento de Botánica y Zoología, Centro Universitario de Ciencias Biológicas y Agropecuarias, Universidad de Guadalajara, Zapopan, Mexico, [7] Coordenação de Tecnologia e Inovação, Instituto Nacional de Pesquisas da Amazônia, Manaus, Brazil

Guava (*Psidium guajava* L., Myrtaceae) is a Neotropical fruit that is widely consumed around the world. However, its evolutionary history and domestication process are unknown. Here we examine available ecological, taxonomic, genetic, archeological, and historical evidence about guava. Guava needs full sunlight, warm temperatures, and well-distributed rainfall throughout the year to grow, but tolerates drought. Zoochory and anthropochory are the main forms of dispersal. Guava's phylogenetic relationships with other species of the genus *Psidium* are unclear. A group of six species that share several morphological characteristics are tentatively accepted as the *Psidium guajava* complex. DNA analyses are limited to the characterization of crop genetic diversity within localities and do not account for possible evolutionary and domestication scenarios. A significant amount of archeological information exists, with a greater number and older records in South America than in Mesoamerica, where there are also numerous historical records. From this information, we propose that: (1) the guava ancestor may have originated during the Middle or Late Miocene, and the savannas and semi-deciduous forests of South America formed during the Late Pleistocene would have been the most appropriate ecosystems for its growth, (2) the megafauna were important dispersers for guava, (3) dispersal by humans during the Holocene expanded guava's geographic range, including to the southwestern Amazonian lowlands, (4) where its domestication may have started, and (5) with the European conquest of the Neotropics, accompanied by their domestic animals, new contact routes between previously remote guava populations were established. These proposals could direct future research on the evolutionary and domestication process of guava.

Keywords: guava, semi-domesticated, center of origin of domestication, archeology, genetic analyses, dispersal

25

INTRODUCTION

Guava *(Psidium guajava* L.) is an important perennial fruit tree who's distribution extends from Mexico and the Antilles to Argentina and Uruguay (Landrum, 2017). It is valued for fresh consumption because of the fruits' aroma and sweet flavor, but many of its products (pulp concentrate or jelly) have export potential (**Figure 1**). The consumption of guava fruits around the world and their medicinal properties have made it the most important among the minor tropical fruits, with an estimated world production of 2.3 million tons per year between 2015 and 2017, and the highest production in India, followed by China, Mexico, Egypt, and Brazil (Altendorf, 2018, 2019). Its chemical properties – vitamins C, A, B1, and B2, calcium,

FIGURE 1 | (A) Aspect of a feral, unpruned, guava tree at a cropland limit in Morelos, Mexico. **(B)** Typical guava branchelet with dusty leaves highlighting the venation in Morelos, Mexico. **(C)** Flower bud. **(D)** Detail of the exfoliating bark, cultivated specimen in Aguascalientes, Mexico. **(E)** Flower with a pair of hymenoptera visitors. **(F–H)** The most common variety in Central Mexico 'media china' and its variation. **(I)** Commercial variety in João Pessoa, North East Brazil. **(J,K)** Two commercial varieties in Bogotá, Colombia. All fruits at the same scale (0.5×). **(L)** Sample of guava products from around the world. Guava products are produced and sold locally or internationally. This sample includes products from Mexico, Central America, South America, Africa, Asia, Malaysia, Philippines, Polynesia, the Caribbean, and Middle East. They come in the form of juices, jams, jellies, pastes, pieces of fruits in syrup, soup base, dried fruits, medicinal leaf teas, and candies. Collection held by Arizona State University Herbarium. Photo credits: H. Alvarado-Sizzo **(A,B,D–H)**, L. Landrum **(C,L)**, J. Romero-Murcia **(J,K)**, and M. F. M. Brito **(I)**.

26

phosphorus, lycopene, and phenolic compounds – along with low cultivation costs and tolerance to soil salinity (Gutiérrez et al., 2008; Rajan and Hudedamani, 2019), have also contributed to this. In numerous tropical and subtropical countries where guava is cultivated, it is often quite invasive (Richardson and Rejmánek, 2011; Landrum, 2017). This success as an invasive species is probably due to its ability to combine clonal and sexual propagation during invasion events (Urquía et al., 2019), along with the production of numerous seeds that can remain viable for a long time (Adhiambo et al., 2019).

Even though guava is an economically important crop, its origin, evolutionary history under domestication, and dispersal throughout the Americas are mostly conjectural. Various authors have suggested northern South America (Risterucci et al., 2005; Aranguren et al., 2008b), northern Amazonia (Hastorf, 2006), or northeastern Brazil (Clement, 1999; Pearsall, 2008) as the area of origin, while others have proposed Central America (Brand, 1939; Nakasone and Paull, 1998). The evolutionary history of perennial crops has become a dominant theme of plant domestication research (Miller and Gross, 2011; Gaut et al., 2015; Spengler, 2020), but discovering the geographic origin and area of domestication of any crop, and even identifying its wild ancestor, can be a challenge. To date, no research has focused on the guava domestication process. Therefore, a review can help build a roadmap to guide future research.

We examine and systematize biological, molecular, archeological, and historical information about guava. Based on this review, we identify the main challenges that have so far limited our understanding of guava domestication, and we present five proposals that can guide future research about the following central questions that remain unanswered: (1) What are its closest relatives? (2) What is the area of origin of guava as a species? (3) What is (are) the area(s) of guava domestication? (4) How did guava reach its current distribution? (5) What role did the European conquerors and their domestic animals play in guava dispersal and gene flow?

NATURAL HISTORY OF *PSIDIUM GUAJAVA*

Fossil Evidence of the *Psidium* Lineage

A warming climatic trend during the Paleocene/Eocene favored the presence of several now tropical lineages in South American high latitudes. Myrtaceae were a notable element in southern Patagonian floras during the early Paleogene (Panti, 2016). Four Patagonian fossils dated to 57-37 million years ago (mya) (Paleocene/Eocene) have been assigned to *Psidium* (Panti, 2016): *Psidium membranaceum* Engelhardt (1891), *P. araciforme* Berry (1938), *P. licciardoi* Hünicken (1967), and *Psidium* sp. (Panti, 2016). Despite strong similarities in lamina shape and venation pattern to present-day *Psidium* leaves, there is still uncertainty about these assignments.

A phylogenetic analysis of the tribe Myrteae (Vasconcelos et al., 2017) suggested that the most recent common ancestor of the genera *Psidium*, *Myrrhinium* Schott and *Mosiera* Small [named *Psidium* group in Vasconcelos et al., 2017 and part of the

current Pimentinae subtribe in Lucas et al., 2019] dates to 25.62 mya, during the Oligocene. Today, we know that a progressive decline in temperature since the Oligocene and the subsequent climatic fluctuations during the middle-Miocene, caused the megathermal angiosperms (e.g., Myrtaceae) to migrate toward lower latitudes (Panti, 2016) from their former range at the southern tip of Patagonia.

Systematics and Taxonomy of Guava

Psidium guajava (**Figure 1**) is a member of the Myrtaceae family that contains the pantropical tribe Myrteae (*sensu* Wilson et al., 2005), which comprises 51 genera and ~2,500 species, including the large genus *Psidium* (Lucas et al., 2007; Govaerts et al., 2019). *Psidium*, as presently circumscribed, contains at least 60 species and perhaps as many as 100 (McVaugh, 1968; Landrum, 2017), and it is undoubtedly a Neotropical genus (Lucas et al., 2005; Landrum, 2017). The Atlantic Coastal Forest, Cerrado, and Caatinga biomes in Brazil, with ca. 50 species are recognized as the center of diversity for *Psidium* (Landrum, 2017, 2021).

Phylogenetic studies based on nuclear and plastid DNA regions have confirmed the monophyly of *Psidium* with *Myrrhinium* as its sister group (Lucas et al., 2005; Vasconcelos et al., 2017; Nadra et al., 2018; Flickinger et al., 2020). However, evolutionary relationships within the genus remain unresolved due to its high rates of diversification (Vasconcelos et al., 2017), its wide Neotropical distribution (Richardson and Rejmánek, 2011; Landrum, 2017), variable ploidy levels (Costa and Forni-Martins, 2007), and hybridization among species (Landrum et al., 1995). In this context, Landrum (2003; 2005; 2021) proposed three species complexes based on morphological, ecological, and geographical characteristics: the *Psidium salutare* complex, the *Psidium grandifolium* complex, and the *Psidium guajava* complex. The *Psidium guajava* complex is also proposed as a working hypothesis of the putative relatives of guava (Landrum, 2021).

The Closest Relatives of Guava

The *Psidium guajava* complex (**Figures 2**, **3**) is conceived as a morphological cluster of species that have similarities in their characteristics (states) that make them distinct from other species of the genus. There are no unique characteritics that define the group, but rather they share a combination of characteristics, such as venation pattern, seed number and size, indumentum pattern, and calyx structure. The complex as presently conceived includes *Psidium guajava* L., *P. guineense* Sw., *P. guyanense* Pers., *P. nutans* O. Berg., *P. rostratum* McVaugh, and *P. rutidocarpum* Ruiz & Pav. (**Figures 2**, **3** and **Supplementary Data 1**). The geographic distribution of the complex extends from Mexico to northern Argentina and Uruguay, including the Antilles (**Figure 4**). A molecular phylogenetic analysis shows *P. guajava* and *P. guineense* as sister taxa (Salywon, 2003), but the relationships with the other species are unclear due to lack of representation (Rivero et al., 2012; Nadra et al., 2018).

Within the complex, *Psidium guajava* and *P. guineense* (**Figures 2A,B**, respectively) are "weedy" species that frequently grow together and are sometimes confused with each other.

27

FIGURE 2 | *Psidium guajava* and three other members of the *P. guajava* complex. **(A)** *P. guajava*, fruit from Lee Lee International Market, Mesa, AZ, United States, the twig and leaves grown from seed from the same variety of guava sold in same market. **(B)** *P. guineense,* cultivated plant at Tempe, Arizona, United States; grown from seeds collected in Chiapas, Mexico. **(C)** *P. rutidocarpum*, wild growing plant from Estación Biológica Huampal, Parque Nacional Yanachaga-Chemillen, Oxapama, Peru. **(D)** *P. rostratum,* twig with leaves and fruit, wild growing plant from Bosque Protector Cerro Blanco, Guayas, Ecuador. Bars all = 3 cm. Photos by **(A,C)**, L. R. Landrum **(B)**, R. Rojas **(D)**, and X. Cornejo.

Psidium guajava shares with *P. rutidocarpum* (**Figure 2C**) leaves with numerous lateral veins (not found in other species of *Psidium*). While *P. guajava* is a widespread weedy species, *P. rutidocarpum* has a restricted range in apparently natural habitats in the upper Amazon of Peru. *Psidium guineense*, *P. guyanense* (**Figures 3A–C**), and *P. nutans* (**Figures 3D–F**) are remarkably similar among themselves and could be merged into one large variable species, but they can be distinguished using the characters described in the key proposed by Landrum (2021). *Psidium rostratum* (**Figure 2D**) is geographically separated from other members of the *P. guajava* complex and only shares the characteristics of flower size (long style, numerous stamens), closed calyx, and placenta, but differs from other species in having fewer ovules and a few large seeds; its venation is similar to *P. guineense*, which justifies its tentative assignment to this complex.

Habitat, Distribution, Dispersers, and Ecology of Guava

Current habitats of *P. guajava* and *P. guineense* are mainly disturbed areas, such as roadsides, grasslands, and orchards, from

near sea level to 1600 m (Landrum, 2017). Their distributions extend from Mexico and the Antilles to northwestern Argentina (**Figure 4**). *Psidium nutans* is found primarily in humid grasslands or riparian habitats at elevations from 150 to 750 m in Argentina, Bolivia, Brazil, Paraguay, and Venezuela (Landrum, 2017). *Psidium rutidocarpum* is an endemic species of the Peruvian Yunga (500 to 2300 m) and open Amazonian forest (80 to 500 m) (Kawasaki and Holst, 2006), while the natural habitat of *P. rostratum* is the Equatorial dry forest located in northwestern Peru and southwestern Ecuador (Linares-Palomino et al., 2010). *Psidium guyanense* has been reported in moist forests in the Amazon biome (Guayana region and Brazil) and the Brazilian Cerrado (Conceição and Aragão, 2010; **Figure 4**). An exploratory richness analysis (**Figure 5**) identified savannas and yungas as high species richness areas for the *P. guajava* complex. Nevertheless, we ruled Yungas vegetation out as suitable for *P. guajava* because this region is rich in dense, layered vegetation, typical of montane forests, which prevents the direct light necessary for guava development.

Most *Psidium* fruits are fleshy berries that have a strong sweet scent and are externally yellow or green. Today, guava

28

FIGURE 3 | Members of the *P. guajava* complex. *Psidium guyanense*. **(A)** Herbarium sheet, showing elliptic leaves with acute to acuminate apexes. **(B)** Open flowers. **(C)** Old flower and closed flower buds showing apical pore in calyx. Outer surfaces of flowers covered with very short, erect hairs. Specimen *S. Mori 17404* from Mun. de Macapá, Amapá, Brazil. *Psidium nutans*. **(D)** Herbarium sheet, showing elliptic leaves with few lateral veins. Specimen from *Landrum & Basualdo 8841*, Canineyú, Paraguay. **(E)** Closed flower buds, the minute calyx lobes barely visible in upper left bud. Specimen from *Guillén & Choré 2426*, Santa Cruz, Bolivia. **(F)** Longitudinal section of fruit showing many small seeds. Specimen from from *Zardini & Guerrero 38496*, Misiones, Paraguay. All photos by L. R. Landrum of specimens at ASU.

29

FIGURE 4 | Map with the current distribution of the six species of the *Psidium guajava* complex based on geo-referenced records available in the Global Biodiversity Information Facility (GBIF), the Botanical Information and Ecology Network (BIEN), the Integrated Digitized Biocollections (IDigBio), Species Link, and The Cooperative Taxonomic Resource for American Myrtaceae (CoTRAM). The distribution of *P. guajava* is certainly more extensive than shown, because botanists and ethnobotanists don't always collect common cultivated species.

FIGURE 5 | Map with species richness of the *Psidium guajava* complex by one-degree grids and selected South American ecoregions. The geo-referenced records were obtained from the following databases: the Global Biodiversity Information Facility (GBIF), the Botanical Information and Ecology Network (BIEN), the Integrated Digitized Biocollections (IDigBio), Species Link, and The Cooperative Taxonomic Resource for American Myrtaceae (CoTRAM).

30

seeds may be dispersed effectively by birds and mammals (e.g., bats, monkeys, ungulates, marsupials) (Gressler et al., 2006; Torres and Gutiérrez, 2018). Some studies have shown that domestic ungulates can disperse between 18,000 and 49,000 guava seeds per day during the fruiting peak (Somarriba, 1985, 1986; Herrera, 2013). Although Perissodactyla (horses and tapirs) are far more likely to disperse and consume sugary fruits than true ruminants (Spengler, 2019), cows are enthusiastic guava consumers (Somarriba, 1985). Double digestion of Artiodactyla presumably does not affect guava seed germination, but seed losses occur due to chewing (Somarriba, 1986). The morphological traits of guava and its abundant fruit production suggest that this tree probably had more efficient dispersers, such as Pleistocenic megafauna, capable of dispersing large quantities of seeds over long distances. Megafaunal fruits are characterized by having seeds that can escape dental grinding by large mammals (one or two large seeds or numerous small seeds) and present dull coloration (brown, yellow, or green). Janzen and Martin (1982) defined this as the Megafaunal dispersal syndrome and guava appears to fit the description.

Guava is a heliophyllic species. Its seeds need a long photoperiod (>10 h) with temperatures ranging between 20 and 30°C to attain > 90% germination (Sugahara and Takaki, 2004). Exposure of mature trees to direct sunlight guarantees abundant fruiting with increased branch growth (Paull and Duarte, 2012). The highest fruit yields are at mean air temperatures of 23 to 28°C (Menzel, 1985). These temperatures are also adequate for the germination of *P. guineense* seeds (Mugnol et al., 2014). Guavas are drought-resistant and well-adapted to different rainfall conditions, although optimal growth and fruit yield occur with 1,000 to 2,000 mm of rainfall well-distributed throughout the year (Menzel, 1985). They are sensitive to cold and are killed or severely injured by prolonged temperatures below freezing, but they can withstand light frosts. Guava adapts to various soil types, although it prefers deep, fertile, well-drained loamy, or sandy-clay soils (Menzel, 1985). Trees survive floods and can grow in seasonally waterlogged soils.

Reproductive Biology and Pollination

Research on breeding systems in *Psidium* focuses primarily on *P. guajava*, which is an allogamous species characterized by an open pollination reproductive cycle (Torres and Gutiérrez, 2018). Nevertheless, self-pollination can occur. Bee-pollination, in which pollen is the sole reward, is the dominant pollination system in *Psidium* (**Figure 1C**; Proença and Gibbs, 1994; Lughadha and Proença, 1996; Gressler et al., 2006). Species of the superfamily Apoidea are important pollinators of both *P. guajava* and *P. guineense* (Lughadha and Proença, 1996; Boti, 2001; Alves and Freitas, 2007; de Siqueira, 2012; Ramos et al., 2019). Both species are also pollinated by the European honeybee, *Apis mellifera* (Soubihe Sobrinho and Gurgel, 1962; Boti, 2001; Gressler et al., 2006; Pommer and Murakami, 2009). *Psidium guyanense*, *P. nutans*, *P. rostratum*, and *P. rutidocarpum* lack comprehensive studies of their breeding systems, pollination, and dispersal.

Cytogenetics and Hybridization

The basic chromosome number for Myrtaceae is x = 11 (Atchison, 1947). Usually, this family shows little variation in chromosome number, with 2n = 22 in most genera, but some studies have reported polyploidy in *Psidium* (Costa et al., 2008; Tuler et al., 2019). Guava karyotype analyses show a predominance of 2n = 2x = 22 chromosomes (**Supplementary Table 1**), with five metacentric (pairs 3, 4, 8, 9, and 10) and six submetacentric chromosomes (pairs 1, 2, 5, 6, 7, and 11) (Coser et al., 2012; Marques et al., 2016). The genome size is small (2C = 0.95 pg; **Supplementary Table 1**), following the classification suggested by Soltis et al. (2004). Variable levels of ploidy have been reported, mainly associated with polyploidy in cultivated plants (**Supplementary Table 1**).

The karyotype of *P. guineense* has 2n = 4x = 44 chromosomes (Chakraborti et al., 2010; Souza et al., 2015; Marques et al., 2016; Tuler et al., 2019), with mainly metacentric (11, 12) and submetacentric (1–10, 13–22) chromosomes (Marques et al., 2016).

Hybridization events between *P. guajava* and *P. guineense* were reported by Landrum et al. (1995). Using macro-, micro-morphological and chemical characteristics, they found intermediate individuals, which exhibit a combination of traits from both species. Molecular analyses are necessary to corroborate introgression between these species, backcrossing, and probable F1 hybrids.

Genetic Insights

Guava genetic studies have used different molecular markers to characterize genetic diversity in feral individuals, germplasm collections, and cultivars, inside and outside its native distribution (**Supplementary Table 2**). These characterizations aimed to identify genotypes of interest for developing breeding programs, which can be useful to understand the domestication of guava, although the characteristics of the different markers used are not easily comparable. For this reason, and because of its overall performance (Selkoe and Toonen, 2006), we only considered the data obtained using microsatellites (simple sequence repeats – SSR).

More than 300 specific microsatellite markers have been developed for guava (Risterucci et al., 2005; Guavamap, 2008), which have been used in studies for guava genetic characterization in several countries (**Supplementary Table 3**). Within the range of its natural distribution, the Mexican germplasm bank, which includes both feral and cultivated samples, showed slightly higher values of diversity (H$_e$ = 0.75; Sánchez-Teyer et al., 2010), than those obtained from Venezuelan wild samples (H$_e$ = 0.73; Aranguren et al., 2008a). Outside of its native Neotropical distribution (Asia and the United States), guava germplasm banks (Viji et al., 2010; Mehmood et al., 2016; Urquía et al., 2019) showed similar levels of genetic diversity (**Supplementary Table 3**), while Asian crops and collections presented both high and low diversity values (Kanupriya et al., 2011; Chaithanya et al., 2014; Sitther et al., 2014; Kherwar et al., 2018; Kumari et al., 2018).

Most of the studies reviewed here showed low observed heterozygosity values compared to expected heterozygosity (**Supplementary Table 3**). Inbreeding or founder effects could explain this pattern (Motamayor et al., 2002; Erfani et al., 2012). Evolutionary genetic studies that explore the geographical origin and genetic diversity of the guava throughout the range of its natural distribution are necessary to understand the guava domestication process and dispersal routes.

CULTURAL HISTORY OF *PSIDIUM GUAJAVA*

Archeological Evidence

Fruit trees were important resources for the pre-Columbian inhabitants of the Neotropics (Piperno and Pearsall, 1998; Clement et al., 2021). Along the Peruvian coast, archeological sites with guava macro-remains date back to 7.000 calibrated years before the present (cal. BP) (**Supplementary Table 4** and **Figure 6**) (Cárdenas, 1999). According to Hastorf (2006), guava arrived on the Peruvian coast from northwestern Amazonia via the low gap in the Andes of Ecuador, while for Pearsall (2008) crops like guava, peanut, manioc, and gourd were introduced from the eastern lowlands by way of the Andes. Later, by ca. 5400 cal. BP guava occurs in archeological sites located on the northern, central, and southern Peruvian coast (**Supplementary Table 4**).

The earliest evidence of guava in northern South America (Colombia), the Antilles (Monserrat), and Mesoamerica (Belize) date to ca. 2600 cal. BP, from the sites at Finca Limoncillos, Trants, and Cuello, respectively (**Supplementary Table 4** and **Figure 6**). This chronology may be coincidental, due to the archeological sample, or a product of rapid human dispersal of the plant. However, it is suggestive of dispersal occurring only millennia after the widespread establishment of guava on the Pacific coast. Most of the sites in these regions date to the last 1500 years, suggesting later widespread utilization of the fruit.

The dry tropical forests of Mesoamerica provide a different scenario. The only remains recovered thus far are four fruit fragments dating to ca. 670 cal. BP from Purron cave (Palo Blanco phase) in the Tehuacán Valley (**Figure 6**) and may coincide with the period in which the area was under intensive cultivation with irrigated fields (Smith, 1967). Smith (1967) interprets the late introduction of guava as it being a luxury food, one of a group of fruits that were not grown before the establishment of an irrigation system that permitted the cultivation of enough staples in this dry environment.

Recent research in Brazilian Amazonia could push this date back from the earliest coastal Peruvian remains, with a *Psidium* seed from the Teotonio archeological site (**Figure 6**), near Porto Velho (state of Rondonia), in a layer dated to between 9490 and 6505 cal. BP (**Supplementary Table 4**). In this context Watling et al. (2018) also found carbonized seeds of other perennial food species, such as Brazil nut (*Bertholletia excelsa* Humb. & Bonpl.), pequiá (*Caryocar* sp.), and micro-remains (phytoliths) of léren (*Calathea* sp.). These findings give insights about food production in southwestern Amazonia, and are also considered evidence of disturbed areas around the human occupation. Guava is a hardy plant and highly adaptable to disturbed habitats, which explains its presence near areas of dense Amazonian vegetation.

Even older evidence comes from pollen remains recovered in the Pay Paso-1 site, near the Cuareim River in Uruguay (**Figure 6**), dated to ca. 11,735 cal. BP (Suárez, 2018). This suggests that *Psidium* trees were part of riverside vegetation and were available for human groups that occupied the Cuareim River region during the Pleistocene-Holocene transition (Suárez, 2018). In Mesoamerica, specifically in the Santa Marta cave (Chiapas, Mexico), Acosta et al. (2013, 2018) also recovered pollen grains, some of them tentatively attributed to *Psidium* sp., dated to the same millennia as the Uruguayan remains (**Figure 6** and **Supplementary Table 4**).

We consider that the pollen remains attributed to *Psidium*, both in the Pay-Paso 1 site (Uruguay) and Santa Marta cave (Mexico), should be interpreted with caution. Usually, *Psidium* pollen grains are tricolpates and occasionally tetracolpates (Thornhill et al., 2012). Those reported from Santa Marta Cave are tetracolpate, apparently the least common form. For the Pay Paso site, there is no information on the shape of *Psidium* pollen grains. The determination of the Myrtaceae genus by pollen recovered in archeological sites using only light microscopy (LM) is notably troublesome and especially difficult at the species level (Tuler et al., 2017). Therefore, more evidence is needed to corroborate the early presence of *Psidium* in Mesoamerican and southern South American archeological sites.

Post-Columbian Historical Evidence

In the Americas, fruits have historically represented a substantial component of indigenous people's diets (Patiño, 1963). Curiously, guava was despised by some of the first Europeans who arrived in the New World. For instance, a Dominican Friar wrote "It [guava] stinks like a bug, and it was an abomination to eat it" (Rodríguez Demorizi, 1942). But later, its medicinal and nutritional values were recognized and then encouraged beyond the Americas. Guava historical records from México and the Caribbean date back to the 16th century. The official historian of the West Indies, Gonzalo Fernández de Oviedo y Valdés, reported the presence of guava fruits during the Yucatán expedition in 1517 (Patiño, 2002). Later, Francisco Cervantes de Salazar, who chronicled the Conquest of Mexico, first mentioned guava for its medicinal use to treat diarrhea (Cervantes de Salazar, 1971; Patiño, 2002). Differences between continental and Caribbean guava fruits were also recorded early by Bartolomé de las Casas: the fruits found in the Dominican Republic were smaller than those grown on the mainland (Patiño, 2002). The Licentiate Echagoian, judge for the Audience of Santo Domingo, considered guava to be a plague (Rodríguez Demorizi, 1942; Patiño, 2002). In Puerto and Santo Domingo, some chroniclers recorded guava as a potent weed in cattle pastures (Patiño, 2002).

In the expeditions which explored western El Nuevo Reino de Granada in 1553, Pedro Cieza de León (conqueror, chronicler, and historian) described the abundance of good quality guavas used for stomach illnesses in Colombia and the Darién Gap (Patiño, 2002). In 1526, Oviedo y Valdés reported that indigenous people from Darien, current Panama, distinguished between

32
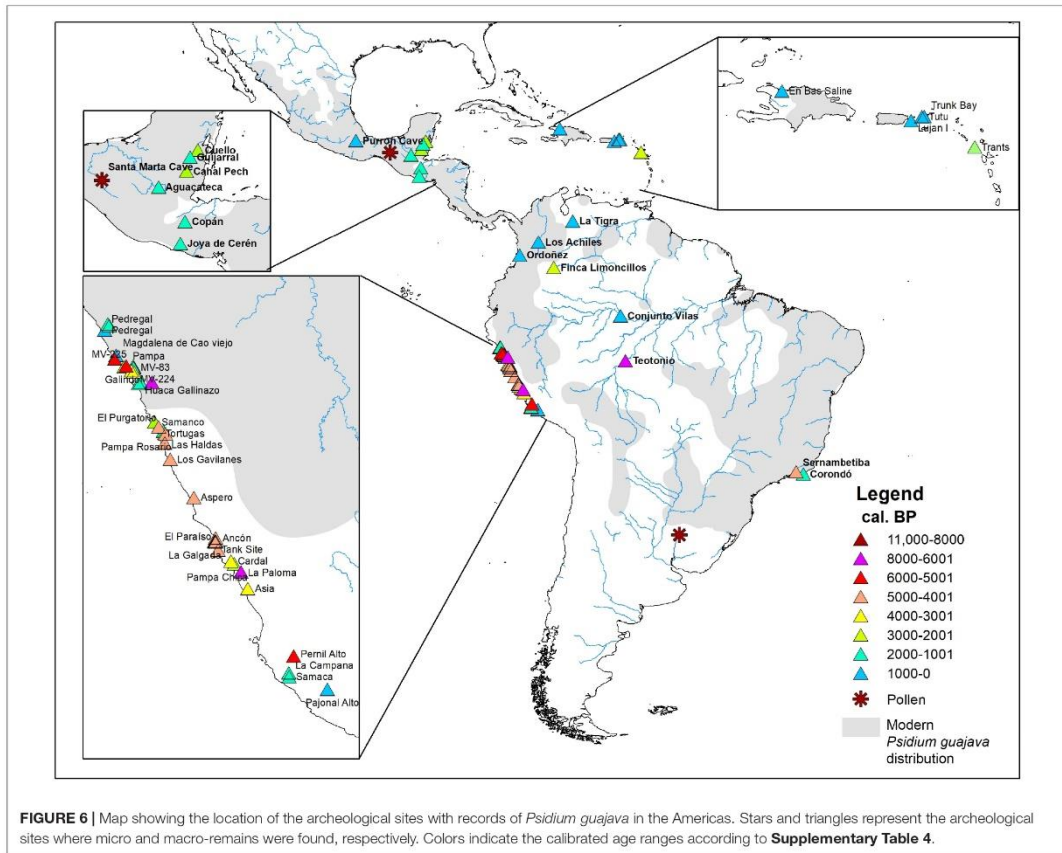
FIGURE 6 | Map showing the location of the archeological sites with records of *Psidium guajava* in the Americas. Stars and triangles represent the archeological sites where micro and macro-remains were found, respectively. Colors indicate the calibrated age ranges according to **Supplementary Table 4**.

domestic and spontaneous guava trees. He also highlighted the quality of the wood and a guava variety with red-white flesh (de Vedia, 1852). Friar Pedro Aguado recorded, in the central region of the El Nuevo Reino de Granada, the consumption of pink guavas and registered the guava wood use in domestic activities (Aguado, 1917; Patiño, 2002). Reports of guava throughout the El Nuevo Reino de Granada (current Colombia, mainly) were included by several chroniclers in their descriptions of the trees and fruits used by indigenous people between 1571 and 1743 (Patiño, 2002). In Venezuela, the Spaniards found guavas in the municipalities of Lagunillas and Mérida, as recorded by Friar Pedro (Aguado, 1917). Guava trees also were reported by several chroniclers between 1535 and 1743 along the Orinoco River (Patiño, 2002).

Diego de Trujillo, Spanish chronicler and conqueror, in 1530-1533, observed guavas at several places along the Ecuadorian coast during the expedition that ended with the conquest of the Inca Empire (Patiño, 2002). Around the same time, guava crops were reported near the city of Cuzco in 1534 and north coast of Peru (1535 and 1630) (Patiño, 2002). Different kinds of

guava with various colors, shapes, and sizes were found during expeditions through Peru and Ecuador in the mid and end of the 16th century (Patiño, 2002).

The chronicler Fray Gaspar de Carvajal observed guava trees in the Amazon Valley near the Andes in 1541–1542. Likewise, guavas were reported along the Manu River expedition in 1567–1569. The Oviedo y Valdés chronicles from the 16th century reported the use of guava by the Guarani in southeastern and southern Brazil and Paraguay (Patiño, 2002). Guava was also recorded in Bahia, Pernambuco, and Paraíba states (northeastern Brazil) by Portuguese chroniclers between 1582 and 1618 (Patiño, 2002).

FIVE PROPOSALS TO GUIDE FUTURE RESEARCH

This review of the natural and cultural histories of guava in the Neotropics provides information that can inform future research. We propose that: (1) guava originated in the savannas

33

and semi-deciduous forests of South America; (2) megafauna were important dispersers of guava; (3) guava domestication started in southwestern Amazonia; (4) subsequent expansion of guava's geographic range is the outcome of human migrations during the pre-Columbian times; and (5) European conquest and the introduction of domestic animals (exotic megafauna) in post-Columbian times promoted secondary contact among guava populations.

Proposal 1: Guava (*Psidium guajava*) Originated in the Savannas and Semi-Deciduous Forests of South America

A common characteristic among species of the *Psidium guajava* complex is their affinity for open forests habitats. Guava plants, in particular, flourish in full sunlight, warm temperatures, rainfall well-distributed throughout the year, and well-drained soils. These ecological characteristics are present, with their respective variations, in the tropical savanna and tropical deciduous forest biomes, where several of the species of the complex are currently distributed.

Due to a drop in temperature since the Eocene, some Myrtaceae lineages' distributions shifted from Patagonia toward tropical zones during subsequent epochs. Thus, the lineage from which the current genus *Psidium* derived possibly required warmer climatic conditions and open forests, which were only present from the Middle or Late Miocene. Reconstructions of the distribution of biomes in the later Miocene, using palaeoecological data and model-predicted vegetation, show evidence for a warmer world than at present and the emergence of tropical savanna and tropical deciduous forest in the center and east of South America (Pound et al., 2011). Tropical savannah covered a large part of the current areas of the humid Chaco and the Cerrado (see Figure 6 in Pound et al., 2011). Palaeobotanical studies showed fossil evidence of woody species related to the current plant taxa of the Chaco and Cerrado, which date from the Late Miocene. According to Anzótegui et al. (2019), the palaeoflora that gave rise to the current Chaco and Cerrado vegetation could have diversified earlier, in the Middle Miocene, and the fossil ancestors of Cerrado vegetation could have had a broader and more southern geographical distribution.

With this in mind, and the analysis of *P. guajava* complex species (**Figure 5**), we propose that the guava ancestor originated during the Middle or Late Miocene in savanna-like vegetation. Later, the Humid Chaco and the Cerrado ecoregions formed during the Pleistocene, would have presented the most suitable ecological conditions for the diversification of the species of *Psidium* that prefer open forests. We suggest the Humid Chaco and/or Cerrado are likely the areas of origin for *P. guajava*. Phylogenetic and biogeographic analyses will be useful to determine the relationships among *Psidium* species, especially within the *P. guajava* complex, to unravel its diversification and dispersal processes. They will also be important to support or refute our proposal for the center of origin of guava (*P. guajava*).

Proposal 2: Megafauna Disperse Guava Into Other Suitable South American Ecosystems

The large fleshy fruits (4 to >10 cm diameter, details in Guimarães et al., 2008) of many tree species, including guava, share a set of traits inconsistent with current seed dispersers (Janzen and Martin, 1982). Species with large fleshy fruits, numerous seeds and abundant sugars could be the outcome of mutualistic relationships of plants with now-extinct animals, primarily with Pleistocene megafauna (Janzen and Martin, 1982; Jordano, 1995; Guimarães et al., 2008).

Reconstructions of ancient diets and habitat preferences of Pleistocene South American mega-omnivores through stable isotope analyses revealed that these animals' subsistence consisted of C3 plants, or a mixture of C3 and C4 plants (MacFadden, 2005; Dantas et al., 2013). Since, from a functional and evolutionary perspective, fleshy fruits are usually accepted as adaptations for animal dispersal (Spengler, 2019), in South America fruit-trees could have been a C3 food source relevant to the megafauna's diets (Janzen and Martin, 1982; Donatti et al., 2007; Guimarães et al., 2008). The evolution of large fruits in *Psidium guajava* with high sugar content was likely accompanied by strong selection pressures for seed dispersal, which led to the recruitment of large animals (Spengler, 2020).

Megafauna are effective long-distance dispersers as they depend on the vast extensions of open spaces and grassland to survive (De Vivo and Carmignotto, 2004). South American savannas were the habitat of *Toxodon*, *Glyptodon*, *Holmesina*, *Pampatherium*, *Megatherium*, *Stegomastodon*, *Tapirus*, and *Equus* (Carlini et al., 2004; Ríos et al., 2014). Paleogeographic evidence has demonstrated similarities in the taxonomic composition of the megafauna in the savannas of Bolivia, Paraguay, Argentina, and the faunal associations found in the tropical areas of Brazil. This suggests migratory phenomena and faunal exchanges between savanna ecosystems and tropical open forests, which consequently would have allowed the dispersal of plants (Carlini et al., 2004; Ríos et al., 2014). Pires et al. (2018) proved that the long-distance dispersal of seeds by gravigrade species could be up to ten times greater than that of current smaller-sized mammals, favoring gene flow and the ability of these species to adapt (Spengler, 2020).

In the case of guava, it has been demonstrated that ungulates, especially domestic ones, can consume and disperse many fruits and seeds of guava. In this perspective, we propose that guava's distribution was expanded by large omnivores within the Humid Chaco or the Cerrado, and colonized open areas in the Atlantic Forest, the Beni savannas (Llanos de Mojos), or during humid intervals, parts of Northeastern Brazil.

Proposal 3: Guava's Neotropical Distribution Is the Outcome of Human Migrations in Pre-Columbian Times

Dispersal of guava by megafauna (Proposal 2), does not account for the modern geography of this fruit due to ecological and environmental restrictions, both of megafauna and guava, from

34

thriving in tropical rainforests. Guava is widely dispersed in South America and Mesoamerica today. The archeological and historical records demonstrate that at least part of this range existed before European arrival.

Following megafauna extinction, indigenous people's food use became the principal dispersal agent for a subset of large-fruited species. van Zonneveld et al. (2018) demonstrated that megafaunal fruits, such as the sapodilla tree (*Manilkara zapota* (L.) P. Royen), soursop (*Annona muricata* L.), and hog plum (*Spondias mombin* L.), increased or maintained their distribution ranges in the Americas through their incorporation into human diets. This suggests active human consumption and management of Neotropical fruit species and, moreover, pre-Columbian exchange, including between subcontinents (van Zonneveld et al., 2018). Dispersal and adoption of edible plants benefited from routes of movement, connection and trade between widespread human groups, both overland and by waterways. Maritime routes have facilitated the transfer of edible plants over long-distance dispersal, allowing humans to exchange and trade products between regions separated by hundreds of kilometers.

When examining the archeological occurrence of guava, we observe remains in numerous archeological sites that fall outside of its current distribution, specifically along the Pacific coastline and the interior of Amazonian forests. In both instances, the area is not optimal for its growth so the creation of additional spaces with ecological conditions favorable would have been the key to maintaining guava in new places; human settlements had these ecological conditions (Clement et al., 2021).

The archeological chronology of remains in the Antilles, with older sites in the south, suggests that the maritime introduction of guava was by way of northern South America, a hypothesis that is testable using genetic methods. The dispersal of guava into Mesoamerica could be either maritime or terrestrial or both, from the Antilles or overland from Colombia's tropical lowlands through Central America, a route demonstrated for the spread of numerous Amazonian plants (Piperno et al., 2000; Powis et al., 2011). Geographic range expansion of *P. guajava* is likely a consequence of human migrations, exchange, and trade in pre-Columbian times, alongside the creation of additional spaces with favorable ecological conditions for guava.

Proposal 4: Guava Domestication Began in the Lowlands of Southwestern Amazonia

Southwestern Amazonia is considered one of the most important centers of crop genetic diversity in the New World, where several native species were domesticated to some degree (Clement, 1999; Clement et al., 2015, 2016; Watling et al., 2018). The compilation of archeological evidence for guava is surprising because the earliest confirmed macrobotanical remains come from locations outside of the plant's current range (**Figure 6**): the Pacific coast of Peru. For guava to occur there, necessary conditions were that (1) the plant was dispersed far beyond the expected range of Pleistocene megafaunal dispersal and (2) an appropriate ecological niche existed or was created. With respect to condition 1, humans are effective Holocene

dispersers and the archeological dates support this, with the possible earliest date in SW Amazonia. With respect to condition 2, guava is an invasive plant that grows vigorously in areas disturbed by humans (Hastorf, 2006) and we can assume that it was as popular in the early Holocene and after European arrival.

The data from the Teotônio site can be used to suggest the early domestication of guava in southwestern Amazonia (Watling et al., 2018) and subsequent introduction to the Peruvian coast. However, we are approaching this hypothesis with caution as at Teotônio the evidence is sparse and there are only two dates that bracket the stratum with the "oldest" archeological remains (we have chosen to use the midpoint between the radiocarbon assays, 9490 to 6505 cal BP, for the **Figure 6**). Support for SW Amazonian cultivation of guava comes predominantly from secondary sources: the aforementioned importance of the region as a center of crop domestication (Clement et al., 2016) from which plants like manioc and peanut were introduced to the Pacific coast (Piperno, 2011), and the probable engagement of humans at Teotônio with cultivation of the root crop léren (Watling et al., 2018) and those in the Llanos de Mojos with léren, manioc, and squash on constructed forest islands (Lombardo et al., 2020).

Proposal 5: The Introduction of Domestic Animals and Transfer of Guava Fruits in Post-Columbian Times Promoted Secondary Contact

Historical records indicate Europeans did not initially value the guava fruits for cultural reasons (for instance, their strong aroma). Nevertheless, guava started to gain acceptance in European communities due to the diverse sizes and colors of fruits and its medicinal benefits, which probably led to selection and later transfer of plant material.

Beyond the intense indigenous interactions during millennia described above, the European conquerors established new maritime routes for exchange between different communities and geographic areas. The contact between plant populations could have produced tree fruits with characteristics from different pre-conquest plant populations, as well as introducing new situations that could change selections desired by both indigenous peoples and new European arrivals.

European colonization also introduced domestic livestock to many areas of the Americas, especially cattle to regions where the ecological conditions were conducive to ranching. According to Janzen and Martin (1982), domestic megafauna may replicate some of the interactions between fruit trees and extinct megafauna. The introduction of such domestic animals from Europe to America provided suitable dispersers for megafaunal fruits and seeds (Janzen and Martin, 1982; Barlow, 2000). Livestock may have increased guava dispersal by favoring the establishment of the guava populations in areas where the species previously did not occur and enhancing gene flow. Given the weedy nature of guava, it may be expected that if it is dispersed into areas with suitable rainfall, light, and temperature, it would have easily established itself. European livestock, while

35

plentiful on ranches, were also frequently maintained on small farms where some pigs and cattle were raised for domestic meat and milk products. Tracing records and the paths of livestock introduction in different regions of the Americas may contribute to interpreting the genetic history of guava, especially outside of areas where it is well represented in the archeological record.

FINAL REMARKS

This review of the literature about the natural and cultural history of guava offers guidance to better understand guava's domestication and dispersal in the Neotropics. The third and fourth proposals are especially amenable to testing with modern molecular methods (genetics/genomics), although they will require extensive collaboration by a Neotropical network of researchers to collect samples and analyze them with uniform methods. A preliminary analysis is underway and will be published soon.

AUTHOR CONTRIBUTIONS

EA-M, AC, and CC conceived the idea. AC and CC conceptualized and supervised the project. EA-M analyzed the data and drafted the manuscript. LL conducted the taxonomic analysis. MS assisted in the collection and analysis of archeological data. HA-S and ER-S assisted in manuscript conceptualization. All authors discussed the results, contributed, and revised the final manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.714763/full#supplementary-material

REFERENCES

Acosta, G., Beramendi, L. E., González, G., Rivera, I., Eudave, I., Hernández, E., et al. (2018). Climate change and peopling of the Neotropics during the Pleistocene-Holocene transition. *Boletín la Soc. Geol. Mex.* 70, 1–19. doi: 10.18268/BSGM2018v70n1a1

Acosta, G. O., Pérez, P. M., and Rivera, I. I. G. (2013). Metodología para el estudio del procesamiento de plantas en sociedades cazadoras-recolectoras: un estudio de caso. *Bol. do Mus. Para. Emílio Goeldi. Ciênc. Hum.* 8, 535–550. doi: 10.1590/S1981-81222013000300004

Adhiambo, R., Muyekho, F., Creed, I. F., Enanga, E., Shivoga, W., Trick, C. G., et al. (2019). Managing the invasion of guava trees to enhance carbon storage in tropical forests. *For. Ecol. Manage.* 432, 623–630. doi: 10.1016/j.foreco.2018.10.010

Aguado, F. P. D. (1917). *Historia*. Available online at: http://babel.banrepcultural.org/cdm/ref/collection/p17054coll10/id/2475 (accessed May 8, 2019).

Altendorf, S. (2018). *Minor Tropical Fruits: Mainstreaming a niche market in Food Outlook Biannual Report on Global Food Markets*. Rome: Food and Agriculture Organization, 67–74.

Altendorf, S. (2019). *Major Tropical Fruits Market Review 2018*. Rome: Food and Agriculture Organization.

Alves, J. E., and Freitas, B. M. (2007). Requerimentos de polinização da goiabeira. *Ciênc. Rural* 37, 1281–1286. doi: 10.1590/S0103-84782007000500010

Anzótegui, L., Garalla, S., Mautino, L., and Prado, D. (2019). Evolución de Bosques Estacionales Secos Neotropicales, Chaco y Cerrado en el Neógeno del Centro-Norte de Argentina. *Publicación Electrónica la Asoc. Paleontol. Argentina* 19, 7–17. doi: 10.5710/PEAPA.12.04.2019.279

Aranguren, Y., Valecillos, C., and Fermin, G. (2008b). "Variability of Venezuelan guava geographic landraces employing phenotypic markers," in *II International Symposium on Guava and Other Myrtaceae*, eds W. Rohde and G. Fermin (Leuven: International Society for Horticultural Science), 87–94. doi: 10.17660/ActaHortic.2010.849.9

Aranguren, Y., Briceño, A., and Fermin, G. (2008a). "Assessment of the variability of Venezuelan guava landraces by microsatellites," in *II International Symposium on Guava and Other Myrtaceae*, eds W. Rohde and G. Fermin (Leuven: International Society for Horticultural Science), 147–154. doi: 10.17660/ActaHortic.2010.849.16

Atchison, E. (1947). Chromosome Numbers in the Myrtaceae. *Am. J. Bot.* 34, 159–164. doi: 10.1002/j.1537-2197.1947.tb12970.x

Barlow, C. (2000). *The Ghosts Of Evolution: Nonsensical Fruit, Missing Partners, And Other Ecological Anachronisms*. New York: Basic Books.

Boti, J. B. (2001). *Polinização entomófila da goiabeira (Psidium guajava L., Myrtaceae): Influência da distância de fragmentos florestais em Santa Teresa, Espírito Santo*. Ph.D. thesis. Viçosa: Universidade Federal de Viçosa.

Brand, D. D. (1939). The Origin and Early Distribution of New World Cultivated Plants. *Agric. Hist.* 13, 109–117.

Cárdenas, M. (1999). *El Periodo Precerámico en el valle de Chao.* Peru: Boletín de Arqueología PUCP, 141–169.

Carlini, A. A., Zurita, A. E., Gasparini, G., and Noriega, J. I. (2004). *Los mamíferos del Pleistoceno de la mesopotamia Argentina y su relación con los del centro norte de la Argentina, Paraguay y sur de Bolivia, y los del sur de Brasil y Oeste de Uruguay: Paleobiogeografía y Paleoambientes, Vol. 12.* Miscelánea: INSUGEO, 83–90.

Cervantes de Salazar, F. (1971). *Crónica de la Nueva España.* Available online at: http://www.cervantesvirtual.com/nd/ark:/59851/bmcrr1s3 (accessed March 8, 2019).

Chaithanya, M. V. N., Dinesh, M. R., Vasugi, C., Reddy, D. C. L., Sailaja, D., and Aswath, C. (2014). Assessment of genetic diversity in guava (*Psidium guajava*) germplasm using microsatellites. *J. Hortic. Sci.* 9, 117–125.

Chakraborti, S., Sinha, S., and Sinha, R. K. (2010). Chromosome number and karyotype analysis of wild guava *Psidium guineense* Sw. – a new report from Tripura, India. *Indian J. Sci. Technol.* 3, 925–927. doi: 10.17485/ijst/2010/v3i8.9

Clement, C. R. (1999). 1492 and the loss of Amazonian crop genetic resources. I. The relation between domestication and human population decline. *Econ. Bot.* 53, 188–202. doi: 10.1007/BF02866498

Clement, C. R., Casas, A., Parra-Rondinel, F. A., Levis, C., Peroni, N., Hanazaki, N., et al. (2021). Disentangling domestication from food production systems in the Neotropics. *Quaternary* 4:4. doi: 10.3390/quat4010004

Clement, C. R., Denevan, W. M., Heckenberger, M. J., Junqueira, A. B., Neves, E. G., Teixeira, W. G., et al. (2015). The domestication of Amazonia before European conquest. *Proc. R. Soc. B Biol. Sci.* 282:20150813. doi: 10.1098/rspb.2015.0813

Clement, C. R., Rodrigues, D. P., Alves-Pereira, A., Mühlen, G. S., De Cristo-Araújo, M., Moreira, P. A., et al. (2016). Crop domestication in the upper Madeira River basin. *Bol. do Mus. Para. Emilio Goeldi Cienc. Hum.* 11, 193–205. doi: 10.1590/1981.81222016000100010

Conceição, G. M. D., and Aragão, J. G. (2010). Diversidade e importância econômica das Myrtaceae do Cerrado, Parque Estadual do Mirador, Maranhão. *Sci. Plena* 6, 1–8.

Coser, S. M., Ferreira, M. F., da, S., Ferreira, A., Mitre, L. K., Carvalho, C. R., et al. (2012). Assessment of genetic diversity in *Psidium guajava* L. using different approaches. *Sci. Hortic.* 148, 223–229. doi: 10.1016/j.scienta.2012.09.030

Costa, I. R., Dornelas, M. C., and Forni-Martins, E. R. (2008). Nuclear genome size variation in fleshy-fruited Neotropical Myrtaceae. *Plant Syst. Evol.* 276, 209–217. doi: 10.1007/s00606-008-0088-x

Costa, I. R., and Forni-Martins, E. R. (2007). Karyotype analysis in South American species of Myrtaceae. *Bot. J. Linn. Soc.* 155, 571–580. doi: 10.1111/j.1095-8339.2007.00704.x

Dantas, M. A. T., Dutra, R. P., Cherkinsky, A., Fortier, D. C., Kamino, L. H. Y., Cozzuol, M. A., et al. (2013). Paleoecology and radiocarbon dating of the Pleistocene megafauna of the Brazilian Intertropical Region. *Quat. Res.* 79, 61–65. doi: 10.1016/j.yqres.2012.09.006

de Siqueira, K. M. M. (2012). "Polinização da Goiabeira (*Psidium guajava* L.)," in *III Semana dos Polinizadores: palestras e resumos*, ed. Márcia de Fátima Ribeiro (Petrolina: Embrapa Semiárido), 139–156. Available online at: http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/936996

de Vedia, E. (1852). *Historiadores primitivos de Indias. Rivadeneyra.* Available online at: http://bdh-rd.bne.es/viewer.vm?id=0000010223&page=1 (accessed March 7, 2019).

De Vivo, M., and Carmignotto, A. P. (2004). Holocene vegetation change and the mammal faunas of South America and Africa. *J. Biogeogr.* 31, 943–957. doi: 10.1111/j.1365-2699.2004.01068.x

Donatti, C. I., Galetti, M., Pizo, M. A., Guimaraes, P. R. J., and Jordano, P. (2007). "Living in the land of ghosts: fruit traits and the importance of large mammals as seed dispersers in the Pantanal, Brazil," in *Seed dispersal: theory and its application in a changing world*, eds A. J. Dennis, E. W. Schupp, R. J. Green, and D. A. Westcott (Wallingford: CAB International), 104–123. doi: 10.1079/9781845931650.0104

Erfani, J., Ebadi, A., Abdollahi, H., and Fatahi, R. (2012). Genetic diversity of some pear cultivars and genotypes using simple sequence repeat (SSR) markers. *Plant Mol. Biol. Rep.* 30, 1065–1072. doi: 10.1007/s11105-012-0421-y

Flickinger, J. A., Jestrow, B., Oviedo Prieto, R., Santiago-Valentín, E., Sustache-Sustache, J., Jiménez-Rodríguez, F., et al. (2020). A phylogenetic survey of Myrtaceae in the Greater Antilles with nomenclatural changes for some endemic species. *Taxon* 69, 448–480. doi: 10.1002/tax.12263

Gaut, B. S., Díez, C. M., and Morrell, P. L. (2015). Genomics and the contrasting dynamics of annual and perennial domestication. *Trends Genet.* 31, 709–719. doi: 10.1016/j.tig.2015.10.002

Govaerts, R., Sobral, M., Ashton, P., Barrie, F., Holst, B. K., Landrum, L. L., et al. (2019). *World Checklist of Selected Plant Families. World Checklist of Myrtaceae, Facil. by R. Bot. Gard. Kew.* Available online at: http://wcsp.science.kew.org/ (Accessed May 7, 2019).

Gressler, E., Pizo, M. A., and Morellato, L. P. C. (2006). Polinização e dispersão de sementes em Myrtaceae do Brasil. *Rev. Bras. Bot.* 29, 509–530. doi: 10.1590/S0100-84042006000400002

Guavamap (2008). *Improvement of Guava: Linkage mapping and QTL analysis as a basisfor marker-assisted selection.* Available online at: https://cordis.europa.eu/project/id/15111/reporting/es (accessed August 14, 2020).

Guimarães, P. R., Galetti, M., and Jordano, P. (2008). Seed dispersal anachronisms: rethinking the fruits extinct megafauna ate. *PLoS One* 3:e1745. doi: 10.1371/journal.pone.0001745

Gutiérrez, R. M., Mitchell, S., and Solis, R. V. (2008). *Psidium guajava*: a review of its traditional uses, phytochemistry and pharmacology. *J. Ethnopharmacol.* 117, 1–27. doi: 10.1016/j.jep.2008.01.025

Hastorf, C. A. (2006). "Domesticated food and society in early coastal Peru," in *Time and complexity in historical ecology: Studies in the Neotropical lowlands*, eds W. Baleé and C. L. Erickson (New York: Columbia University Press), 87–126. doi: 10.7312/bale13562-006

Herrera, M. X. A. (2013). *Posibles dispersores de Psidium guajava en la Isla San Cristóbal, Galápagos - Ecuador.* Ph.D. thesis. Quito: Universidad San Francisco de Quito.

Janzen, D. H., and Martin, P. S. (1982). Neotropical anachronisms: the fruits the gomphotheres ate. *Science* 215, 19–27. doi: 10.1126/science.215.4528.19

Jordano, P. (1995). Angiosperm fleshy fruits and seed dispersers: a comparative analysis of adaptation and constraints in plant-animal interactions. *Am. Nat.* 145, 163–191. doi: 10.1086/285735

Kanupriya, Latha, P. M., Aswath, C., Reddy, L., Padmakar, B., Vasugi, C., et al. (2011). Cultivar identification and genetic fingerprinting of guava (*Psidium guajava*) using microsatellite markers. *Int. J. fruit Sci.* 11, 184–196. doi: 10.1080/15538362.2011.578521

Kawasaki, L., and Holst, B. K. (2006). Myrtaceae endémicas del Perú. *Rev. Peru. Biol* 13, 463s–468s. doi: 10.15381/rpb.v13i2.1885

Kherwar, D., Usha, K., Mithra, S. V. A., and Singh, B. (2018). Microsatellite (SSR) marker assisted assessment of population structure and genetic diversity for morpho-physiological traits in guava (*Psidium guajava* L.). *J. Plant Biochem. Biotechnol.* 27, 284–292. doi: 10.1007/s13562-017-0438-2

Kumari, S., Nagaraja Arumugam, R. S., Srivastav, M., Banoth, S., Mithra, A. C., GoswamiI, A. K., et al. (2018). Diversity analysis of guava (*Psidium guajava*) germplasm collection. *Indian J. Agric. Sci.* 88, 489–497.

Landrum, L. R. (2003). A revision of the *Psidium salutare* complex (Myrtaceae). *SIDA* 20, 1335–1354.

Landrum, L. R. (2005). A revision of the *Psidium grandifolium* complex (Myrtaceae). *SIDA* 21, 1335–1354.

Landrum, L. R. (2017). The genus *Psidium* (Myrtaceae) in the state of Bahia, Brazil. *Canotia* 13, 1–101.

Landrum, L. R. (2021). "*Psidium guajava* L.: taxonomy, relatives and possible origin," in *Guava: botany, production and uses*, ed. S. Mitra (Wallingford: Cab International), 1–21. doi: 10.1079/9781789247022.0001

Landrum, L. R., Clark, W. D., Sharp, W. P., and Brendecke, J. (1995). Hybridization between *Psidium guajava* and *P. guineense* (Myrtaceae). *Econ. Bot.* 49, 153–161. doi: 10.1007/BF02862919

Linares-Palomino, R., Kvist, L. P., Aguirre-Mendoza, Z., and Gonzales-Inca, C. (2010). Diversity and endemism of woody plant species in the Equatorial Pacific seasonally dry forests. *Biodivers. Conserv.* 19, 169–185. doi: 10.1007/s10531-009-9713-4

Lombardo, U., Iriarte, J., Hilbert, L., Ruiz-Pérez, J., Capriles, J. M., and Veit, H. (2020). Early Holocene crop cultivation and landscape modification in Amazonia. *Nature* 581, 190–193. doi: 10.1038/s41586-020-2162-7

Lucas, E. J., Belsham, S. R., Nic Lughadha, E. M., Orlovich, D. A., Sakuragui, C. M., Chase, M. W., et al. (2005). Phylogenetic patterns in the fleshy-fruited

37

Myrtaceae - Preliminary molecular evidence. *Plant Syst. Evol.* 251, 35–51. doi: 10.1007/s00606-004-0164-9

Lucas, E. J., Harris, S. A., Mazine, F. F., Belsham, S. R., Nic Lughadha, E. M., Telford, A., et al. (2007). Suprageneric phylogenetics of Myrteae, the generically richest tribe in Myrtaceae. *Taxon* 56, 1105–1128. doi: 10.2307/25065906

Lucas, E. J., Holst, B., Sobral, M., Mazine, F. F., Nic Lughadha, E. M., Barnes Proença, C. E., et al. (2019). A new subtribal classification of tribe Myrteae (Myrtaceae). *Syst. Bot.* 44, 560–569. doi: 10.1600/036364419X15620113920608

Lughadha, E. N., and Proença, C. (1996). A Survey of the reproductive biology of the Myrtoideae (Myrtaceae). *Ann. Mo. Bot. Gard.* 83, 480–503. doi: 10.2307/2399990

MacFadden, B. J. (2005). Diet and habitat of toxodont megaherbivores (Mammalia. Notoungulata) from the late Quaternary of South and Central America. *Quat. Res.* 64, 113–124. doi: 10.1016/j.yqres.2005.05.003

Marques, A. M., Tuler, A. C., Carvalho, C. R., Carrijo, T. T., Da Silva, Ferreira, M. F., et al. (2016). Refinement of the karyological aspects of *Psidium guineense* (Swartz, 1788): a comparison with *Psidium guajava* (Linnaeus, 1753). *Comp. Cytogenet.* 10, 117–128. doi: 10.3897/CompCytogen.v10i1.6462

McVaugh, R. (1968). The Genera of American Myrtaceae: an Interim. *Taxon* 17, 354–418. doi: 10.2307/1217393

Mehmood, A., Luo, S., Ahmad, N. M., Dong, C., Mahmood, T., Sajjad, Y., et al. (2016). Molecular variability and phylogenetic relationships of guava (*Psidium guajava* L.) cultivars using inter-primer binding site (iPBS) and microsatellite (SSR) markers. *Genet. Resour. Crop Evol.* 63, 1345–1361. doi: 10.1007/s10722-015-0322-7

Menzel, C. M. (1985). Guava: an exotic fruit with potential in Queensland. *Queensl. Agric. J.* 111, 93–98.

Miller, A. J., and Gross, B. L. (2011). From forest to field: perennial fruit crop domestication. *Am. J. Bot.* 98, 1389–1414. doi: 10.3732/ajb.1000522

Motamayor, J. C., Risterucci, A. M., Lopez, P. A., Ortiz, C. F., Moreno, A., and Lanaud, C. (2002). Cacao domestication I: the origin of the cacao cultivated by the Mayas. *Heredity* 89, 380–386. doi: 10.1038/sj.hdy.6800156

Mugnol, D. D., Silvana, S., Quintão, D. P., Eliane, N., Elisa, M. T., and Mara, M. R. (2014). Effect of pre-treatments on seed germination and seedling growth in *Psidium guineense* Swartz. *Agrociencia Uruguay* 18, 33–39. doi: 10.31285/AGRO.18.460

Nadra, M. G., Giannini, N. P., Acosta, J. M., and Aagesen, L. (2018). Evolution of pollination by frugivorous birds in Neotropical Myrtaceae. *PeerJ.* 6:e5426. doi: 10.7717/peerj.5426

Nakasone, H. Y., and Paull, R. E. (1998). *Tropical Fruits*. New York: Cab International.

Panti, C. (2016). Myrtaceae fossil leaves from the Río Turbio Formation (Middle Eocene), Santa Cruz Province, Argentina. *Hist. Biol.* 28, 459–469. doi: 10.1080/08912963.2014.976635

Patiño, V. M. (1963). *Plantas cultivadas y animales domésticos en América Equinoccial I: Frutales*. Cali: Imprensa Departamental.

Patiño, V. M. *Historia y dispersión de los frutales nativos del neotrópico*. Colombia: Centro Internacional de Agricultura Tropical.

Paull, R. E., and Duarte, O. (2012). "Guava," in *Tropical fruits, Volume 2*, eds O. Paull and R. E. Duarte (Wallingford: CABI International), 91–122. doi: 10.1079/9781845937898.0091

Pearsall, D. M. (2008). "Plant domestication and the shift to agriculture in the Andes," in *The handbook of South American archaeology*, eds H. Silverman and W. H. Isbell (New York: Springer Science & Business Media), 105–120. doi: 10.1007/978-0-387-74907-5_7

Piperno, D. R. (2011). The Origins of Plant Cultivation and Domestication in the New World Tropics. *Curr. Anthropol.* 52, S453–S470. doi: 10.1086/659998

Piperno, D. R., and Pearsall, D. M. (1998). *The Origin of Agriculture in the Lowland Neotropics*. San Diego: Academic Press, 243–320. doi: 10.1016/B978-012557180-7/50006-0

Piperno, D. R., Ranere, A. J., Holst, I., and Hansell, P. (2000). Starch grains reveal early root crop horticulture in the Panamanian tropical forest. *Nature* 407, 894–897. doi: 10.1038/35038055

Pires, M. M., Guimarães, P. R., Galetti, M., and Jordano, P. (2018). Pleistocene megafaunal extinctions and the functional loss of long-distance seed-dispersal services. *Ecography* 41, 153–163. doi: 10.1111/ecog.03163

Pommer, C., and Murakami, K. (2009). "Breeding guava (*Psidium guajava* L.)," in *Breeding Plantation Tree Crops: Tropical Species*, eds S. M. Jain and P. M. Priyadarshan (Berlin: Springer Science+Business Media), 83–120. doi: 10.1007/978-0-387-71201-7_3

Pound, M. J., Haywood, A. M., Salzmann, U., Riding, J. B., Lunt, D. J., and Hunter, S. J. (2011). A Tortonian (Late Miocene, 11.61-7.25Ma) global vegetation reconstruction. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 300, 29–45. doi: 10.1016/j.palaeo.2010.11.029

Powis, T. G., Cyphers, A., Gaikwad, N. W., Grivetti, L., and Cheong, K. (2011). Cacao use and the San Lorenzo Olmec. *Proc. Natl. Acad. Sci. U. S. A.* 108, 8595–8600. doi: 10.1073/pnas.1100620108

Proença, C. E. B., and Gibbs, P. E. (1994). Reproductive biology of eight sympatric Myrtaceae from Central Brazil. *New Phytol.* 126, 343–354. doi: 10.1111/j.1469-8137.1994.tb03954.x

Rajan, S., and Hudedamani, U. (2019). "Genetic Resources of Guava: importance, Uses and Prospects," in *Conservation and Utilization of Horticultural Genetic Resources*, eds P. E. Rajasekharan, V. R. Rao (Singapore: Springer) 363–383. doi: 10.1007/978-981-13-3669-0_11

Ramos, E., Souza, R., Ribeiro, M., and de, F. (2019). *Visitantes florais do araçazeiro (Psidium guineense Sw.)(Myrtaceae), in Anais da XIV Jornada de Iniciação Científica da Embrapa Semiárido. (Embrapa Smeiárido)*. Petrolina: JORNADA DE INICIAÇÃO CIENTÍFICA DA EMBRAPA SEMIÁRIDO.

Richardson, D. M., and Rejmánek, M. (2011). Trees and shrubs as invasive alien species a global review. *Divers. Distrib.* 17, 788–809. doi: 10.1111/j.1472-4642.2011.00782.x

Rios, S. D., Luna, C. A., Souberlich, R., Aguilera, P., Cervera, A. M. G., de, C., et al. (2014). Mamíferos del Cuaternario de Puerto Santa Rosa, departamento de San Pedro. Paraguay. *Bol. Mus. Nac. Hist. Nat. Parag.* 18, 67–76.

Risterucci, A. M., Duval, M. F., Rohde, W., and Billotte, N. (2005). Isolation and characterization of microsatellite loci from *Psidium guajava* L. *Mol. Ecol. Notes* 5, 745–748. doi: 10.1111/j.1471-8286.2005.01050.x

Rivero, G., Salazar, G., Pacheco, D., Sánchez, A., Quirós, M., and Sthormes, G. (2012). Relaciones filogeneticas entre especies de *Psidium* (Myrtaceae) presentes en el occidente de Venezuela a partir de secuencias de ADN nuclear (ITS) y plastidial (trnH-psbA). *Interciencia* 37, 838–844.

Rodríguez Demorizi, E. (1942). *Relaciones históricas de Santo Domingo*. Available online at: https://books.google.com.mx/books?id=tHIKAQAAIAAJ (accessed March 7, 2019).

Salywon, A. M. (2003). *A monograph of Mosiera (Myrtaceae)*. Ph.D. thesis. Tempe: University of State Arizona.

Sánchez-Teyer, L. F., Barraza-Morales, A., Keb, L., Barredo, F., Quiroz-Moreno, A., O'Connor-Sánchez, A., et al. (2010). Assessment of genetic diversity of Mexican guava germplasm using DNA molecular markers. *Acta Hortic.* 849, 133–138. doi: 10.17660/ActaHortic.2010.849.14

Selkoe, K. A., and Toonen, R. J. (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecol. Lett.* 9, 615–629. doi: 10.1111/j.1461-0248.2006.00889.x

Sitther, V., Zhang, D., Harris, D. L., Yadav, A. K., Zee, F. T., Meinhardt, L. W., et al. (2014). Genetic characterization of guava (*Psidium guajava* L.) germplasm in the United States using microsatellite markers. *Genet. Resour. Crop Evol.* 61, 829–839. doi: 10.1007/s10722-014-0078-5

Smith, C. E. (1967). "Plant remains," in *The Prehistory of Tehuacan Valley. Volume One: Environment and Subsistence*, ed. D. S. Byers (Austin: University of Texas Press), 220–255.

Soltis, D. E., Soltis, P. S., and Tate, J. A. (2004). Advances in the study of polyploidy since plant speciation. *New Phytol.* 161, 173–191. doi: 10.1046/j.1469-8137.2003.00948.x

Somarriba, E. (1985). Árboles de Guayaba (*Psidium guajava* L.) en Pastizales II: consumo de Fruta y Dispersión de Semillas. *Rev. Interam. Ciencias Agríc.* 35, 329–332.

Somarriba, E. (1986). Effects of livestock on seed germination of guava (*Psidium guajava* L.). *Agrofor. Syst.* 4, 233–238. doi: 10.1007/BF02028357

Soubihe Sobrinho, J., and Gurgel, J. T. A. (1962). Taxa de panmixia na goiabeira (*Psidium guajava* L.). *Bragantia* 21, 15–20. doi: 10.1590/S0006-87051962000100002

Souza, A. G., Resende, L. V., Lima, I. P., Martins, L. S. S., and Techio, V. H. (2015). Chromosome number and nuclear DNA amount in *Psidium* spp. resistant and susceptible to *Meloidogyne enterolobii* and its relation with compatibility

38

between rootstocks and commercial varieties of guava tree. *Plant Syst. Evol.* 301, 231–237. doi: 10.1007/s00606-014-1068-y

Spengler, R. N. (2019). Origins of the apple: the role of megafaunal mutualism in the domestication of *Malus* and rosaceous trees. *Front. Plant Sci.* 10:617. doi: 10.3389/fpls.2019.00617

Spengler, R. N. (2020). Anthropogenic Seed Dispersal: rethinking the Origins of Plant Domestication. *Trends Plant Sci.* 25, 340–348. doi: 10.1016/j.tplants.2020.01.005

Suárez, R. (2018). "The Peopling of Southeastern South America: Cultural Diversity, Paleoenvironmental Conditions, and Lithic Technological Organization During the Late Pleistocene and Early Holocene," in *Lithic Technological Organization and Paleoenvironmental Change.*, eds E. Robinson and F. Sellet, (Cham: Springer) 281–300. doi: 10.1007/978-3-319-64 407-3

Sugahara, V. Y., and Takaki, M. (2004). Effect of light and temperature on seed germination in guava (*Psidium guajava* L. - Myrtaceae). *Seed Sci. Technol.* 32, 759–764. doi: 10.15258/sst.2008.36.3.30

Thornhill, A. H., Hope, G. S., Craven, L. A., and Crisp, M. D. (2012). Pollen morphology of the Myrtaceae. Part 4: tribes Kanieae, Myrteae and Tristanieae. *Aust. J. Bot.* 60, 260–289. doi: 10.1071/BT11177

Torres, M. D. L., and Gutiérrez, B. (2018). "A preliminary assessment of the genetic diversity and population structure of guava, *Psidium guajava*, in San Cristobal," in *Understanding Invasive Species in the Galapagos Islands*, eds M. L. Torres and C. F. Mena (Berlin: Springer), 3–17. doi: 10.1007/978-3-319-67177-2_1

Tuler, A. C., Carrijo, T. T., Peixoto, A. L., Garbin, M. L., da Silva, Ferreira, M. F., et al. (2019). Diversification and geographical distribution of *Psidium* (Myrtaceae) species with distinct ploidy levels. *Trees* 33, 1101–1110. doi: 10.1007/s00468-019-01845-2

Tuler, A. C., da Silva, T., Carrijo, T. T., Garbin, M. L., Mendonça, C. B. F., Peixoto, A. L., et al. (2017). Taxonomic significance of pollen morphology for species delimitation in *Psidium* (Myrtaceae). *Plant Syst. Evol.* 303, 317–327. doi: 10.1007/s00606-016-1373-8

Urquía, D., Gutierrez, B., Pozo, G., Pozo, M. J., Espín, A., Torres, M., et al. (2019). *Psidium guajava* in the Galapagos Islands: population genetics and history of an invasive species. *PLoS One* 14:e0203737. doi: 10.1371/journal.pone.0203737

van Zonneveld, M., Larranaga, N., Blonder, B., Coradin, L., Hormaza, J. I., and Hunter, D. (2018). Human diets drive range expansion of megafauna-dispersed fruit species. *Proc. Natl. Acad. Sci. U. S. A.* 115, 3326–3331. doi: 10.1073/pnas.1718045115

Vasconcelos, T. N., Proenca, C. E., Ahmad, B., Aguilar, D. S., Aguilar, R., Amorim, B. S., et al. (2017). Myrteae phylogeny, calibration, biogeography and diversification patterns: increased understanding in the most species rich tribe of Myrtaceae. *Mol. Phylogenet. Evol.* 109, 113–137. doi: 10.1016/j.ympev.2017.01.002

Viji, G., Harris, D. L., Yadav, A. K., and Zee, F. T. (2010). Use of microsatellite markers to characterize genetic diversity of selected accessions of guava (*Psidium guajava*) in the United States. *Acta Hortic.* 859, 169–176. doi: 10.17660/ActaHortic.2010.859.20

Watling, J., Shock, M. P., Mongeló, G. Z., Almeida, F. O., Kater, T., De Oliveira, P. E., et al. (2018). Direct archaeological evidence for Southwestern Amazonia as an early plant domestication and food production centre. *PLoS One* 13:e0199868. doi: 10.1371/journal.pone.0199868

Wilson, P. G., O'Brien, M. M., Heslewood, M. M., and Quinn, C. J. (2005). Relationships within Myrtaceae *sensu lato* based on a *matK* phylogeny. *Plant Syst. Evol.* 251, 3–19. doi: 10.1007/s00606-004-0162-y

39

SUPPORTING INFORMATION

The Taming of *Psidium guajava*: Natural and Cultural History of a Neotropical Fruit

Cita completa:

Arévalo-Marín E, Casas A, Landrum L, Shock MP, Alvarado-Sizzo H, Ruiz-Sanchez E and Clement CR (2021) The Taming of *Psidium guajava:* Natural and Cultural History of a Neotropical Fruit. Front. Plant Sci. 12:714763. doi: 10.3389/fpls.2021.714763

Supplementary Material

Supplementary Data 1. Taxonomic description of the *P. guajava* complex

Psidium guajava L.

      Shrub or tree up to ca. 12 m high, subglabrous to densely appressed pubescent on young growth and lower leaf surfaces, the trunk smooth, light brown to light gray-green, with large flaky scales; hairs whitish, yellowish, or silvery, up to ca. 0.7 mm long, erect or appressed; young twigs quadrangular, slightly to strongly winged, often sulcate (at least when dry), densely to moderately appressed-pubescent, the older twigs at first scaly with longitudinal striations or fibers, eventually smooth with irregular scales falling as patches. LEAVES elliptic, oblong, elliptic-oblanceolate, elliptic- obovate, or lanceolate, 4.5–14 cm long, 2.4–7.5 cm wide, 1.6–4 times as long as wide, densely to sparsely appressed pubescent below, subglabrous except for puberulent midvein above; apex acute, acuminate, to rounded; base rounded to slightly cordate; petiole 2–5 mm long, 1–2 mm thick, channeled, densely pubescent to subglabrous; venation brochidodromous distally to eucamptodromous proximally, the midvein impressed above, prominent below, the lateral veins 9–22 prominent pairs, ascending at angle of ca. 45°, nearly straight, curving toward apex near the margin and connecting with the next lateral, the marginal vein not clearly present or arching between the laterals, the tertiary veins connecting the laterals in a ladder-like to reticulate pattern; blades coriaceous to submembranous, drying yellow-green, gray-green, to dark reddish brown. FLOWER BUDS subfusiform to pyriform, 9–14 mm long, sometimes strongly constricted near the midpoint, the hypanthium narrowly campanulate, barrel shaped or fusiform 4–6 mm long, the distal portion of bud more or less ovoid, sometimes strongly so with a conical apex, 4.5–9.5 mm long; indumentum pattern of buds with peduncles, hypanthium, and bracteoles sparsely to moderately appressed pubescent, the calyx without glabrous to sparsely pubescent (usually less densely covered than that hypanthium), the calyx within glabrous or densely pubescent, the petals, disk, and style glabrous; peduncles 1–3-flowered, 1–3.5 cm long, 1–1.5 mm thick, terete; bracteoles linear to narrowly triangular, 2–5 mm long. CALYX closed, tearing irregularly as the bud opens, persisting or falling in ca. 3 parts; petals obovate to elliptic, 13–22 mm long; disk 4–6 mm across; stamens 280–720, 7–15 mm long; anthers 0.7–1 mm long, with 1–7(–10) glands; style 10–15 mm long; ovary 3–6-locular; ovules 90–180 per locule, multiseriate. FRUIT globose to pyriform, 2–6(–8) cm long, green to yellow without, with pink, yellow, or white flesh, aromatic; seeds numerous, subreniform, 3–4 mm long, more or less smooth, the seed coat ca. 0.25 mm thick. 2n = 22.

P. guineense Sw.

Shrub or small tree up to about 6 m high, typically densely covered with reddish brown to yellowish gray velvety indumentum on the inflorescence and young growth but sometimes glabrous or nearly so, the trunk smooth to scaly; hairs simple, usually spreading, grayish to reddish brown, ca. 0.3–0.5 mm long; young twigs densely to moderately velutinous, or less often glabrous, compressed to terete in section, losing indumentum in about 1 year, usually not angled but sometimes grooved when young, vigorous shoots sometimes weakly angled, the older bark usually remaining more or less smooth, less often somewhat flaky or stringy. LEAVES elliptic, elliptic-oblong,

obovate, 4–11.5 cm long, 2–8 cm wide, 1.3–2.4 times as long as wide, usually densely to moderately velutinous below, glabrous to covered with hairs along the midvein above, the margin entire; apex obtuse, rounded, or acute; base rounded to acute; petiole 4–12 mm long, 1.5–2 mm thick, channeled, densely to sparsely pubescent, rarely glabrous; venation brochidodromous to eucamptodromous distally, the midvein impressed or nearly flat above, prominent below, the lateral veins 5–10 pairs, ascending at an angle of ca. 45°, diminishing and looping near the margin to connect with the next lateral, a clear marginal vein not formed, the tertiary veins, connecting the laterals in a ladder-like to reticulate pattern; blades coriaceous, drying yellowish brown to reddish brown, concolorous to somewhat darker above, when dry often mottled and/or lustrous above. FLOWER BUDS pyriform, 8–15(–17) mm long, the hypanthium ellipsoid to obconic, 3.5–7 mm long, the distal portion of bud ellipsoid, subglobose, or ovoid, 4.5– 10 mm long; indumentum pattern of buds with all external surfaces moderately to densely pubescent (rarely subglabrous), the calyx pubescent without, but less densely so than hypanthium, distally pubescent within, the petals pubescent without, the disk sparsely pubescent (less often glabrous), the style glabrous; peduncles terete to compressed, 9–25(–30) mm long, 1–2 mm wide, uniflorous or triflorous, the branches of the dichasium when present 2–12 mm long; bracteoles narrowly triangular, ca. 2–3 mm long, caducous at about anthesis. CALYX closed completely, or with a terminal, pore-like opening at the apex, tearing longitudinally to the staminal ring, usually in 5 parts, these sometimes persisting until the fruit matures, the margin of calyx pore if present sinuate or with 5 small lobes; petals elliptic to obovate, 7–11 mm long; disk 4– 5 mm across; stamens 180–300, 7–10 mm long; anthers 1–3 mm long, more or less introrsely dehiscent, the glands in the connective 1 to over 50; style 8–10 mm long; ovary 3–5-locular; ovules 50–100 per locule, ca. 8-seriate. FRUIT subglobose to ellipsoidal, 1–3 cm long; seeds (22–)27–250 per fruit, 3–4 mm long. 2n = 44.

P. guyanense Pers.
Tree or shrub to 1-5 m high, minutely hirtellous on young growth; *hairs* mainly less than 0.1 mm long, erect, whitish to yellowish brown; *young twigs* minutely and sparsely hirtellous, sometimes appearing glabrous, drying dark reddish brown to light brown, the bark of older twigs smooth or flaky, often gland-dotted, the hairs persisting on first bark until it falls. LEAVES ovate, lanceolate, or elliptic, 6-13.5 cm long, 3-6.5 cm wide, (1.3-)1.8-2.6(-3) times as long as wide; *apex* acute to acuminate; *base* rounded, acute or cuneate; *petiole* 3-9 mm long, 1.5-2 mm thick, slightly to deeply channeled, minutely hirtellous; *venation* brochidodromous distally, eucamptodromous proximally, the midvein impressed or flat above, prominent below, the lateral veins 4-7, the marginal arching broadly distally, 0.5-3(-5) mm from margin, the tertiary venation dendritic; *blades* sometimes submembranous at anthesis, coriaceous at maturity, drying dark reddish brown. FLOWER BUDS 8-12 mm long, pyriform, the hypanthium obconic to campanulate, ca. 3.5-5 mm long, the distal portion of bud subglobose, 4-7 mm long, often wider than long, sparsely hirsutilose; *indumentum pattern of buds* with external surfaces sparsely to densely hirsutilose, the calyx often less densely so than adjacent hypanthium, the calyx within densely hirsutilose, the disk hirsutilose to pubescent, the style glabrous or with scattered hairs; *peduncles* 1-3-flowered, 4-20 mm long, 1-1.5 mm wide, compressed; *bracteoles* caducous before anthesis, perhaps small, narrowly triangular, those seen narrowly elliptic, 5-7 mm long. CALYX nearly closed (with a small

terminal pore), or bowl-like with a large terminal pore, with a sinuate margin, tearing into 5-lobes at anthesis, these ca. 5 mm long, 3-5 mm wide; *petals* obovate to suborbicular, 6-9 mm long, glandular; *disk* including staminal ring 5-6 mm wide after anthesis; *stamens* 200-400, ca. 10 mm long, sparsely pubescent to glabrous; *anthers* oblong, 1-1.5 mm long, with 3-5 glands in the connective; *style* ca. 10 mm long, the stigma peltate, ca. ca. 1 mm wide; *ovary* 3-5 locular; *ovules* 50-75 per locule, the placenta protruding, sometime reflexed with ovules pointing towards center of flower. FRUIT oblong to suborbicular, to 5 cm long, the fruit wall 4-6 mm thick; seeds (2.5)3-5 mm long, 44-188, with rounded and flat sides.

P. nutans O. Berg.
Tree or shrub 1–5 m high, glabrous or sparsely appressed antrorsely pubescent on distal inner surface of calyx (rarely puberulent on young growth); hairs if present colorless, to ca. 0.5 mm long; young twigs reddish brown to blackish brown when dry, smooth with numerous darker glands, the bark of older twigs lighter brown or tan, falling as flakes. LEAVES elliptic to obovate, 5–17 cm long, 3–12 cm wide, 1.4–2.6 times as long as wide; apex acute, acuminate, obtuse to rounded, often with a cuspidate tip; base cuneate, obtuse, rounded, oblique, or subcordate; petiole shallowly channeled, 3–10 mm long, 1–2.5 mm thick; venation eucamptodromous proximally, brochidodromous distally, the midvein impressed proximally to nearly flat above, prominent below, the lateral veins 4–8(–10) pairs, leaving the midvein at an angle of 30–60°, nearly straight near midvein, arching distally towards apex, the marginal vein usually evident distally, arching mostly between 1–3 mm from the margin, the tertiary veins dendritic to ladder-like in pattern; blades coriaceous, often lustrous above and below, drying reddish to blackish brown, often mottled above with lighter spots when dry, the margin sometime crenulate in part. FLOWER BUDS pyriform, moderately to strongly constricted at ovary summit, 7–12(–14) mm long, the hypanthium ellipsoid to campanulate, 3–6 mm long, the distal portion of bud ovoid to subglobose, 4–6.5(–9.5) mm long; indumentum pattern of buds with all surfaces glabrous or essentially so; peduncles 1–3-flowered, 0.4–2 cm long, flattened, 1–2 mm wide, the branches of the dichasia 0.4–1 cm long; bracteoles narrowly triangular, ca. 2 mm long, caduceus at or before anthesis. CALYX closed, or with a small apical pore-like opening, with 5 minute lobes on the margin of the opening, tearing irregularly at anthesis, usually in 4–5 persistent pieces 4–8 mm long, the tears not cutting the staminal ring, glabrous to sparsely pubescent distally within; petals elliptic to obovate, 0.8–1.4 cm long; disk ca. 3.5 mm across; stamens 110–240, 6–10 mm long; anthers 1.2–2(–3) mm long, with a few to several glands in the connective; style 9–13 mm long, the stigma peltate, 0.5–0.8 mm across; ovary 3–5- locular; ovules 50–105 per locule, the placenta protruding, sometimes somewhat peltate. FRUIT globose, to subpyriform, 0.7–2 cm long; seeds numerous (50 in one fruit), subtriangular with rounded edges, 3–4 mm long.

P. rostratum McVaugh
Tree or shrub 1.8-15 m high, densely to moderately appressed pubescent to strigose on young growth and some floral structures, glabrescent with age, the trunk smooth, tannish or reddish; *hairs* whitish to tawny yellow, ca. 0.5 mm long; *young twigs* subterete, compressed or obscurely quadrangular, without clear wings, densely appressed pubescent to glabrous, drying dark reddish black to tan, dotted with small

glands, the bark of older twigs scaly to smooth, generally light reddish brown. LEAVES elliptic, ovate, or oblong, 2.6-13.5 cm long, 2-4.8 cm wide, 1.3-2.6 times as long as wide, moderately strigose-pubescent to glabrous below, sparsely pubescent to glabrous above; *apex* rounded, acute, or acuminate, the tip sometimes apiculate; *base* rounded, acute, or acuminate; *petiole* channeled or not, 2-5 mm long, 1-1.5 mm thick, sparsely strigose to glabrous; *venation* brochidodromous (sometimes eucamptodromous proximally), the midvein impressed or sulcate proximally above, prominent below, the lateral veins 7-13 pairs, usually leaving the midvein at an angle greater than 45 degrees, arcing slightly upwards, up to ca. 0.25 mm wide, the marginal vein arching between the laterals, 1-10 mm from the margin, the tertiary veins forming a dendritic to ladder-like pattern between the laterals; *blades* submembranous to subcoriaceous, drying gray-green to blackish green or reddish brown, often covered with numerous amber colored protruding glands, these largest along the midvein. FLOWER BUDS often with a rostrate tip (rostrum), discounting the rostrum, broadly pyriform, 10-12 mm long, the rostrate tip 1-5 mm long when present, the hypanthium obconic, 3-4 mm long, the distal portion of bud globose, 7-8 mm long; *indumentum pattern of buds* with external surfaces sparsely pubescent to glabrous, or calyx and hypanthium densely puberulent without, the calyx sparsely to densely pubescent within, the petals glabrous or ciliate, sometimes densely glandular, the disc entirely glabrous or puberulent on staminal ring, the style glabrous; *peduncles* uniflorous, 0.2-4.5 cm long, ca. 1 mm wide, subterete; *bracteoles* caducous before anthesis, unknown. CALYX closed in bud, often prolonged as a rostrum 1-5 mm long, or open only as a small pore at the apex, tearing irregularly at anthesis, the tears sometimes cutting the staminal ring; *petals* obovate to suborbicular, ca. 1 cm long; *disk* within staminal ring 2-10 mm across; *stamens* ca. 600 or perhaps more, ca. 1 cm long; *anthers* ca. 1 mm long, with 0 or 1 gland in the connective, the gland reddish; *style* ca. 1 cm long; *ovary* 3-4-locular in flower seen; *ovules* 15-26, radiating from an elongate, slightly peltate placenta. FRUIT subpyriform to subglobose, 1.5-4 cm long, the wall 2-6 mm thick; seeds 4-12 (or perhaps more), 8-12 mm long, blackish or brown, smooth, with rounded with flat sides.

P. rutidocarpum G. Don.
Tree 5-11 m high, pubescent on young growth and lower leaf surfaces, the trunk "mottled reddish brown, peeling" (ex Smith); *hairs* reddish brown to whitish, usually curled, ca. 0.5 (-1) mm long; *young twigs* quadrangular, winged, moderately to densely pubescent, the bark falling in about 1 year, the older twigs more or less smooth or scaly, round, the scales gray, the inner bark dark reddish brown. LEAVES lanceolate, often narrowly so, 4-11 cm long, 1.4-3.4 cm wide, (2-)2.9-4.4 times as long as wide, the margin slightly revolute; *apex* attentuate-acuminate; *base* rounded to obtuse; *petiole* channeled, pubescent, 3-5 mm long, ca. 1.2 mm thick; *venation* mainly eucamptodromous, sometimes brochidodromous distally, the midvein impressed above, prominent below, the lateral veins 11-25 prominent pairs, usually impressed above, ascending at an angle of ca. 45 degrees, nearly straight but curving upwards near margin, the tertiary veins numerous (more than 10 per cm of lateral), connecting the laterals in more or less ladder-like pattern; *blades* coriaceous, drying dark reddish brown to dark gray-green. FLOWER BUDS unknown but probably similar to *P. guajava* or *P. guineense*; *indumentum pattern of buds* probably with external surfaces glabrous to moderately pubescent, the inner suface of the calyx densely covered with reddish brown

hairs, the disc subglabrous; *peduncles* 5-23 mm long, 1-1.2 mm thick; *bracteoles* unknown. CALYX closed in bud, tearing irregularly, usually in 3 to 5 parts, tearing to staminal ring, but usually not cutting into it; *petals* unknown; *hypanthium* with ca. 10 poorly defined longitudinal ridges (in young fruits at least); *disk* ca. 5 mm across; *stamens* ca. 300; *anthers* unknown; *style* unknown; ovary 3-locular; *ovules* 25–40 ovules per locule. FRUIT subglobose, 1.5-2.5 cm long; seeds 65-90, ca. 2.5 mm long.

Supplementary Table 2. Summary of the published genetic studies using dominant markers for guava genetic characterization

Molecular marker	Objetive	Reference
RAPD	Determine the effectiveness of RAPDs in identifying cultivars.	Dahiya, Archak, and Karihaloo (2002)
	Characterize guava cultivars in Mexico.	Padilla-Ramírez et al. (2002)
	Analyze molecular diversity of Indian *Psidium* genotypes.	Prakash, Narayanaswamy, and Sondur (2002)
	Determinate genetic variability of the 82 guava genotypes in Mexico.	Domínguez et al. (2005)
	Examine the diversity among *Psidium guajava* cultivars and two aditional species *P. catleyanum* and *P. friedrichthalianum*.	Sharma et al. (2005)
	Characterize genetic diversity of guava germplasm in Colombia.	Rueda et al. (2006)
	Evaluate genetic variability of 17 guava genotypes from Mexico.	Tapia Pérez andLegaria Solano (2007)
	Assess the genetic variability present in cultivated commercial varieties, breeding lines and unimproved cultivars.	Bajpai et al. (2008)
	Evaluate the genetic divergence among six cultivars and 19 accessions of guava.	Gomes Filho et al. (2010)
	Characterize genetic diversity of guava germplasm in Bangladesh.	Ahmed, Mannan, & Hossain (2011)
	Analyze molecular variation among guava native genotypes from Aguascalientes, Mexico.	Valera-Montero et al. (2016)
AFLP	Examine genetic diversity from 52 mexican guava germplasm.	Hernández-Delgado et al. (2007)
	Evaluate the genetic diversity of Mexican guava accessions to identify potential parental materila for genetic improvment programmes.	Sánchez-Teyer et al. (2010)
	Determine genetic similarity among Brasilian guava accessions.	Corrêa et al. (2011)
	Characterize genetic diversity of guava germplasm in Thailand.	Thaipong et al. (2017)
RAM (random microsatellites)	Characterize genetic diversity of guava trees from Valle del Cauca, Colombia.	Sanabria (2006)

References

Ahmed, B., Mannan, M. A., and Hossain, S. A. (2011). Molecular characterization of guava (*Psidium guajava* L.) germplasm by RAPD analysis. Int. J. Nat. Sci. 1, 62–67.

Bajpai, A., Chandra, R., Rajan, S., and Srivastava, N. (2008). RAPD and minisatellite markers for genetic diversity and relationship in guava varieties. Indian J. Genet. Plant Breed. 68, 441.

Corrêa, L. C., Santos, C. A. F., Lima, G. P. P., Rodrigues, M. A., and Costa, T. P. P. (2011). Similaridade genética entre acessos de goiabeiras e araçazeiros baseada em marcadores moleculares AFLP. Embrapa Semiárido-Artigo em periódico indexado.

Dahiya, K. K., Archak, S., and Karihaloo, J. L. (2002). DNA fingerprinting of guava (*Psidium guajava* L.) cultivars using RAPD markers. Indian J. Plant Genet. Resour. 15, 112–115.

Domínguez, J. L. Á., Legaria-Solano, J. P., Nieto-Ángel, R., Barrientos-Priego, A. F., and Pineda-Pineda, J. (2005). Variabilidad genética según RAPD de árboles de guayabo "Media China" procedentes de cuatro plantaciones y su respuesta morfológica a baja disponibilidad de nutrimentos. Rev. Chapingo Ser. Hortic. 11, 329–343.

Gomes Filho, A., Oliveira, J. G. de, Viana, A. P., Siqueira, A. P. de O., Oliveira, M. G., and Pereira, M. G. (2010). Marcadores moleculares RAPD e descritores morfológicos na avaliação da diversidade genética de goiabeiras (*Psidium guajava* L.). Acta Sci. Agron. 32, 627–633.

Hernández-Delgado, S., Padilla-Ramírez, J. S., Nava-Cedillo, A., and Mayek-Pérez, N. (2007). Morphological and genetic diversity of Mexican guava germplasm. Plant Genet. Resour. Charact. Util. 5, 131–141. doi:10.1017/s1479262107827055.

Padilla-Ramírez, J. S., González-Gaona, E., Esquivel-Villagrana, F., Mercado-Silva, E., Hernández-Delgado, S., and Mayek-Pérez, N. (2002). Caracterización de germoplasma sobresaliente de guayabo de la región Calvillo-Cañones, México. Rev. Fitotec. Mex. 25, 393–399.

Prakash, D. P., Narayanaswamy, P., and Sondur, S. N. (2002). Analysis of molecular diversity in guava using RAPD markers. J. Hortic. Sci. Biotechnol. 77, 287–293.

Rueda, A., Palacio-Mejía, J. D., Muñoz, J. E., Saavedra, R., Bravo, E., Palacio, J. D., et al. (2006). Caracterización molecular del banco de germoplasma de guayaba *Psidium* spp. Centro de Investigación Corpoica-Palmira. Fitotec. Colomb. 6, 26–32.

Sanabria, H. L. (2006). Caracterización molecular con marcadores RAM de árboles nativos *de Psidium guajava* (guayaba) en el Valle del Cauca. Acta Agronómica 55, 23–30.

Sánchez-Teyer, L. F., Barraza-Morales, A., Keb, L., Barredo, F., Quiroz-Moreno, A., O'Connor-Sánchez, A., et al. (2010). Assessment of Genetic Diversity of Mexican Guava Germplasm Using DNA Molecular Markers. Acta Hortic., 133–138. doi:10.17660/ActaHortic.2010.849.14.

Sharma, A. S., Sehrawat, S. K., Singhrot, R. S., and Boora, K. S. (2005). Assessment of genetic diversity and relationship among Psidium spp. through RAPD analysis. in I International Guava Symposium 735, 71–78.

Tapia Pérez, D., and Legaria Solano, J. P. (2007). Variabilidad genética en cultivares de guayabo (*Psidium guajava* l.). Rev. Fitotec. Mex. 30, 391–401.

Thaipong, K., Promchot, S., Auvuchanon, A., and Boonprakob, U. (2017). Genetic analysis of guava germplasm using AFLP markers. Int. J. Agric. Technol. 13, 741–752.

Valera-Montero, L. L., Muñoz-Rodríguez, P. J., Silos-Espino, H., and Flores-Benítez, S. (2016). Genetic diversity of guava (*Psidium guajava* L.) from Central Mexico revealed by morphological and RAPD markers. Phyton (B. Aires). 85, 176–183.

Supplementary Table 3. Summary of published guava genetic diversity values using microsatellite (SSR) markers.

Sample provenance	Objetive	Status	N	Number loci	N_T	N_a	Ho	He	PIC	Reference
Natural area distribution										
Cuba	Characterize genetic diversity of guava germplasm in Cuba.	Germplasm bank	34	7	34	4.5	0.38	-		Valdés-Infante et al. (2007)
Mexico	Evaluate the genetic diversity of Mexican guava accessions to identify potential parental material for genetic improvement programmes.	Germplasm bank	57	6	79	13	-	0.75	-	Sánchez-Teyer et al. (2010)
Venezuela	Characterize genetic diversity of natural populations of *P. guajava* in Venezuela.	Feral	31	14	91	6.9	-	0.73	-	Aranguren, Briceno, and Fermin (2008)
Brazil	Characterize the genetic diversity among *P. guajava* genotypes	Cultivars and Feral	28	24	70	2.7	-	-	-	Coser el al. (2012)
Ecuador	Assess the genetic diversity and population structure of guava in San Cristobal (Galapagos), Ecuador	Feral	94 - 18*	17	-	17.8 - 15.2*	0.153 - 0.252*	-	-	Torres and Gutiérrez (2018)
Outside and inside natural area distribution										
Ecuador	Analyze the genetic diversity and population structure of *P. guajava* populations on the Isabela, Santa Cruz and San Cristobal islands in the Galapagos (Ecuador)	Feral	269	11	52	4.7	0.16	0.35	-	Urquía et al. (2019)
USA, El Salvador, Brazil, South Africa, Malaysia, Thailand	Characterize diversity of guava accesions in the United States.	Germplasm bank	13	14	91	4.7	0.20	0.61	-	Viji et al. (2010)
India	Evaluate genetic diversity of nine guava cultivars and generated barcodes.	Cultivars (hybrids)	9	23	147	6.3	-	0.82	0.74	Kanupriya et al. (2011)

Location	Objective	Material	N		N_T	N_a	Ho	He	PIC	Reference
USA, Puerto Rico, El Salvador, Brazil, South Africa, Malaysia, Singapore, Thailand, Israel, Taiwan, Australia	Characterize diversity of guava germplasm in the United States.	Germplasm bank	38	20	178	8.9	0.20	0.70	-	Sitther et al. (2014)
India and other countries	Measure genetic diversity in guava germplasm at Indian Institute of Horticultural Research, Bengaluru, India	Germplasm bank	72	48	249	-	0.171	0.616	0.56	Chaithanya et al. (2014)
Pakistan, India, Hawai, Mexico	Assess the molecular variation and genetic structure of 51 promising Pakistani guava compared with 19 others from different geographical regions.	Cultivars	70	18	172	9.5	-	0.091	0.61	Mehmood et al. (2016)
India	Determine the genetic structure and evaluate marker transfer to closely related wild species of *P. guajava*.	Germplasm bank	36	24	81	3.6	-	-		Kherwar et al. (2018)
India	Characterize diversity of guava accesions in New Delhi, India.	Germplasm bank	33	26	64	2.3	0.22	0.22	0.23	Kumari et al. (2018)

* Data from island and continent respectively

N: number of individuals analyzed; N$_T$: total number of alleles; N$_a$: average number of alleles for locus; Ho: observed heterozygosity: He: expected heterozygosity; PIC: polymorphism information content.

References

Aranguren, Y., Briceño, A., and Fermin, G. (2008a). "Assessment of the variability of Venezuelan guava landraces by microsatellites," in II International Symposium on Guava and Other Myrtaceae, eds. W. Rohde and G. Fermin, 147–154.

Chaithanya, M. V. N., Dinesh, M. R., Vasugi, C., Reddy, D. C. L., Sailaja, D., and Aswath, C. (2014). Assessment of genetic diversity in guava (*Psidium guajava*) germplasm using microsatellites. J. Hortic. Sci. 9, 117–125.

Coser, S. M., Ferreira, M. F. da S., Ferreira, A., Mitre, L. K., Carvalho, C. R., and Clarindo, W. R. (2012). Assessment of genetic diversity in Psidium guajava L. using different approaches. Sci. Hortic. (Amsterdam). 148, 223–229. doi:10.1016/j.scienta.2012.09.030

Kanupriya, Latha, P. M., Aswath, C., Reddy, L., Padmakar, B., Vasugi, C., et al. (2011). Cultivar identification and genetic fingerprinting of guava (*Psidium guajava*) using microsatellite markers. Int. J. fruit Sci. 11, 184–196. doi:10.1080/15538362.2011.578521

Kherwar, D., Usha, K., Mithra, S. V. A., and Singh, B. (2018). Microsatellite (SSR) marker assisted assessment of population structure and genetic diversity for morpho-physiological traits in guava (*Psidium guajava* L.). J. Plant Biochem. Biotechnol. 27, 284–292. doi:10.1007/s13562-017-0438-2

Kumari, S., Nagaraja Arumugam, R. S., Srivastav, M., Banoth, S., Mithra, A. C., GoswamiI, A. K., et al. (2018). Diversity analysis of guava (*Psidium guajava*) germplasm collection. Indian J. Agric. Sci. 88, 489–497.

Mehmood, A., Luo, S., Ahmad, N. M., Dong, C., Mahmood, T., Sajjad, Y., et al. (2016). Molecular variability and phylogenetic relationships of guava (*Psidium guajava* L.) cultivars using inter-primer binding site (iPBS) and microsatellite (SSR) markers. Genet. Resour. Crop Evol. 63, 1345–1361. doi:10.1007/s10722-015-0322-7

Sánchez-Teyer, L. F., Barraza-Morales, A., Keb, L., Barredo, F., Quiroz-Moreno, A., O'Connor-Sánchez, A., et al. (2010). Assessment of genetic diversity of Mexican guava germplasm using DNA molecular markers. Acta Hortic., 133–138. doi:10.17660/ActaHortic.2010.849.14

Sitther, V., Zhang, D., Harris, D. L., Yadav, A. K., Zee, F. T., Meinhardt, L. W., et al. (2014). Genetic characterization of guava (*Psidium guajava* L.) germplasm in the United States using microsatellite markers. Genet. Resour. Crop Evol. 61, 829–839. doi:10.1007/s10722-014-0078-5

Torres, M. de L., and Gutiérrez, B. (2018). "A preliminary assessment of the genetic diversity and population structure of guava, *Psidium guajava,* in San Cristobal," in Understanding Invasive Species in the Galapagos Islands, 3–17. doi:10.1007/978-3-319-67177-2_1

Urquía, D., Gutierrez, B., Pozo, G., Pozo, M. J., Espín, A., and Torres, M. de L. (2019). *Psidium guajava* in the Galapagos Islands: population genetics and history of an invasive species. PLoS One 14, e0203737. doi:10.1371/journal.pone.0203737

Valdés-Infante, J., Rodríguez, N. N., Becker, D., Velázquez, B., Sourd, D., Espinosa, G., et al. (2007). Microsatellite characterization of guava (*Psidium guajava* L.) germplasm collection in Cuba. Cultiv. Trop. 28, 5–11.

Viji, G., Harris, D. L., Yadav, A. K., and Zee, F. T. (2010). Use of microsatellite markers to characterize genetic diversity of selected accessions of guava (*Psidium guajava*) in the United States. Acta Hortic., 169–176.

Supplementary Table 4. Chronological data for guava remains (macrobotanical and pollen) from various archaeological sites in South America, Mesoamerica, and the Antilles based on published radiocarbon dates. The dates were calibrated, for calendar age ranges (2σ) and median calibrated age, using OxCal 4.4 (Bronk, 2009; Bronk and Lee, 2013) with the Northern Hemisphere Intcal20 (Reimer et al. 2020) or the Southern Hemisphere ShCal20 (Hogg et al. 2020) calibration curve, as appropriate.

Country	Area	Archaeological site	Published C^{14} age years BP	Calibrated age BP (2σ)	Median calibrated age BP	References
Uruguay	Cuareim River	Pay Paso 1 Cave	10,115 ± 25	11,835 – 11,605 11,530 – 11,500 11,425 – 11,405	11,735	Suárez (2011, 2018)
Brazil	Upper Madeira River	Teotonio	8460 ± 30 to 5720 ± 30*	9535 - 9455 to 6600 - 6585 6565 – 6440 6420 – 6410	9490 to 6505	Watling et al. (2018)
	Guanabara Bay	Sernambetiba	4260 ± 65 to 3010 ± 80*	4875 – 4810 4755 – 4710 4670 – 4650 to 3345 – 2380 3275 – 3070	4840 to 3205	Scheel-Ybert and Boyadjian (2020)
	Cabo Frio	Corondó	1920 ± 70*	1930 – 1735	1835	Scheel-Ybert and Boyadjian (2020)
	Solimões River	Conjunto Vilas			1500 to 900[†]	Shock *in press.*
Peru	Chao Valley	Piedras Negras A	6100 ± 90	7245 – 7210 7170 – 6740	6975	Cárdenas (1999)
	Chilca River Valley	La Paloma	6030 ± 180 to 4900 ± 130	7315 – 6485 6465 – 6455 to 6180 – 6145 6120 – 6070 6065 – 6045 6020 – 5275 5170 – 5135 5105 – 5070	6890 to 5645	Dering and Weir (1979, 1981) *In*: Piperno and Pearsall (1998); Benfer (1999)

Río Grande River Valley	Pernil Alto	4677 ± 31[†]	5475 – 53185	5395	Gorbahn (2013)
Chicama Valley	Huaca Prieta	4507 ± 270	5880 – 5825 5755 – 4500 4495 – 4435	5150	Bird et al. (1985)
Ancón-Chillon Valley	Pampa	4450 ± 110	5445 – 5400 5330 – 4830	5095	Patterson and Moseley (1968)
Ancón-Chillon Valley	Yacth Club	4440 ± 110 to 3885 ± 95[†]	5445 – 5405, 5330 – 4830 to 4575 – 4540 4535 – 4070 4045 – 3985	5085 to 4305	Patterson and Moseley (1968)
Ancón-Chillon Valley	Ancón	4200 ± 80 to 3760 ± 95[†]	4960 – 4930 4885 – 4515 4475 – 4445 to 4415 – 3885	4715 – 4135	Patterson and Moseley (1968)
Huarmey Valley	Los Gavilanes	4140 ± 160	5265 – 5185 5120 – 5115 5055 – 4225 4205 – 4155	4660	Bonavia and Grobman (1979); Bonavia (1996)
Ancón-Chillon Valley	Banco Verde	4125 ± 105	4875 – 4400 4370 – 4355 4330 – 4300	4645	Patterson and Moseley (1968)
Supe Valley	Aspero	4060 ± 150	4960 – 4925 4915 – 4895 4885 – 4140 4125 – 4090	4560	Feldman (1983)
Casma River Valley	Las Haldas	3960 ± 80	4795 – 4760 4695 – 4675 4625 – 4150	4410	Pozorski (1987)

Tablachaca River Valley	La Galgada	3820 ± 100 to 3440 ± 80	4520 – 4475 4450 – 3960 3950 – 3925 to 3895 – 3480	4220 to 3695	Pearsall (2003); Washburn et al. (2020)
Ancón-Chillon Valley	Tank Site	3810 ± 150	4795 – 4760, 4625 – 3820, 3795 – 3770, 3745 – 3725	4205	Patterson and Moseley (1968)
Ancón-Chillon Valley	El Paraíso, PV46-35 PV46-35 o Chuquitanta	3810 ± 150 to 3455 ± 105[†]	4795 – 4760, 4625 – 3820, 3795 – 3770, 3745 – 3725 to 3980 – 3455	4205 to 3720	Patterson and Moseley (1968)
Ancón-Chillon Valley	Punta Grande	3760 ± 95	4415 – 3885	4135	Patterson and Moseley (1968)
Casma Valley	Tortugas	3750 ± 65	4355 – 4325 4300 – 3905	4115	Ugent et al. (1986); Pozorski (1987)
Casma Valley	Pampa de las Llamas - Moxeke	3735 ± 75	4400 – 4370 4355 – 4325 4300 – 3880	4095	Pozorski (1987); Pearsall (2003)
Moche Valley	Alto Salaverry	3430 ± 110	3980 – 3445 3420 – 3410	3690	Pozorski and Pozorski (1990)
Chilca Valley	Asia	3270 ± 100	3825 – 3790 3765 – 3750 3725 – 3320 3305 – 3250	3505	Engel (1963)
Moche Valley	Gramalote	3140 ± 30	3450 – 3325 3295 – 3250	3365	Prieto (2015)
Lurín Valley	Cardal	3120 ± 90	3560 – 3530 3495 – 3070	3320	Burger and Salazar-Burger, (1991)

Casma Valley	Pampa Rosario	2760 ± 75	3060 – 2745	2870	Ugent et al. (1984); Ugent et al. (1986); Pozorski (1987)
Lurín Valley	Pampa Chica	2460 ± 70	2720 – 2355	2540	Chevalier (1999); Dulanto (2002)
Moche Valley	La Poza	2240 ± 30	2340 – 2290 2270 – 2150	2225	Bardolph (2017)
Nepeña Valley	Samanco	2166 ± 29	2310 – 2225 2210 – 2190 2185 – 2050 2015 – 2005	2170	Helmer (2014)
Virú Valley	Huaca Santa Clara	2010 ± 50*	2000 – 1870	1950	Millaire (2010); Masur et al. (2018)
Moche Valley	MV-225	1890 ± 30	1875 – 1715	1790	Bardolph (2017)
Moche Valley	MV-224	1790 ± 15	1725 – 1690 1670 – 1620	1655	Bardolph (2017)
Virú Valley	Huaca Gallinazo	1730 ± 40*	1700 – 1655 1640 – 1545	1615	Millaire (2010); Masur et al. (2018)
Late Nasca	La Campana			1450 ^	Beresford-Jones et al. (2009, 2011)
Moche Valley	MV-83	1510 ± 15	1405 – 1350	1380	Bardolph (2017)
Moche Valley	Galindo	1373 ± 41	1355 – 1245 1210 – 1175	1295	Pozorski (1979), Koons and Alex (2014)
Moche Valley	Cerro Chepén	1373 ± 41 to 1292 ± 35	1355 – 1245 1210 – 1175 to 1295 – 1170 1160 – 1130	1295 to 1225	Koons and Alex (2014); Szpak and Chiou (2020)
Casma Valley	El Purgatorio			1250 to 550 ^	Vogel et al. (2016)
Moche Valley	San José Moro	1292 ± 35	1295 – 1170 1160 – 1130	1225	Koons and Alex (2014); Szpak and Chiou (2020)

	Middle Horizon	Samaca H13 F-23			1150[†]	Beresford-Jones et al. (2009, 2011)
	Jetquetepeque Valley	Pedregal			1050 to 480[†]	Cutright (2014)
	Nasca	Pajonal Alto			900[†]	Conlee (2000)
	Jetquetepeque Valley	Pedregal	766 ± 34	735 – 660	695	Cutright (2013)
		Magdalena de Cao viejo			450[†]	Williams (2013)
Colombia	Guaviare	Finca Limoncillos	2541 ± 38	2750 – 2670 2655 – 2610 2600 – 2490	2620	Kosztura (2020)
	La Miel	Los Achiles	565 ± 115	725 – 425 395 – 385 380 – 315	570	Morcote and Cavelier (1999); Morcote-Ríos (2006)
	Calima	Ordóñez	525 ± 35	630 – 595 560 – 540	540	Romero-Picón (1996); Morcote-Ríos (2006)
Venezuela	Barinas	La Tigra			1400 to 950^	Spencer et al. (1994); Iriarte and Dickau (2012)
Honduras		Copán			ca. 1100[†]	Turner and Miksicek (1984)
Guatemala	Department El Petén	Aguateca			ca. 1350[†]	Lentz et al. (2013) *In* Cavallaro (2013)
El Salvador		Joya de Cerén			1300[†]	Slotten and Lentz (*in press*)
Belize		Cuello	2540 ± 70	2760 – 2407 2390 – 2367	2596	Miksicek et al. (1991); Hammond et al. (1995)
		Tolok group, Cahal Pech	2220 ± 150	2705 – 2629 2620 – 2555 2545 – 1870 1852 – 1841	2219	Powis et al. (1999)

	NW Belize	Bronco		1350 to 1050[†]	Hageman and Goldstein (2009)
	NW Belize	Guijarral		1350 to 1050[†]	Hageman and Goldstein (2009)
	NW Belize	Chispas		1350 to 1050[†]	Hageman and Goldstein (2009)
Monserrat		Trants		2750 to 2350[†]	Newsom and Wing (2004); Fitzpatrick (2006)
St. John		Trunk Bay		1150 to 750[†]	Newsom and Wing (2004); Wild (2013)
Vieques		Lujan I		950 to 650[†]	Newsom and Wing (2004); Kimura et al. (2016)
St. Thomas		Tutu	720 ± 120	910 – 840 835 – 520	675	Pearsall (2002); Lundberg (2002)
Hispaniola		En Bas Saline	340 ± 70	515 – 280 170 – 150	395	Deagan (2004); Newsom and Wing (2004)
Mexico	Ocozocoautla Valley	Santa Marta Cave	9850 ± 40	11,395 – 11,375 11,335 – 11,195	11,250	Acosta et al. (2013, 2018)
Mexico	Tehuacán Valley	Purron Cave	730 ± 30	725 – 645 585 – 570	670	Smith (1967); Kaplan and Lynch (1999)

* Dates in the original publication in 2σ. † Data obtained from regional cultural chronology. ^ Date range given by author for unpublished radiocarbon assays.

References

Acosta, G., Beramendi, L. E., González, G., Rivera, I., Eudave, I., Hernández, E., et al. (2018). Climate change and peopling of the Neotropics during the Pleistocene-Holocene transition. Boletín la Soc. Geológica Mex. 70, 1–19. doi:10.18268/bsgm2018v70n1a1

Acosta, G. O., Pérez, P. M., and Rivera, I. I. G. (2013). Metodología para el estudio del procesamiento de plantas en sociedades cazadoras-recolectoras: un estudio de caso. Bol. do Mus. Para. Emílio Goeldi. Ciências Humanas 8, 535–550. doi:10.1590/S1981-81222013000300004

Bardolph, D. N. (2017). Reimagining ancient agricultural strategies and gendered labor in the Prehispanic Moche Valley of North Coastal Peru. [dissertation/doctor's thesis]. [Santa Barbara (CA)]: University of California.

Benfer, R. A. (1999). Proyecto de excavaciones en Paloma, valle de Chilca, Perú. Boletín Arqueol. PUCP, 213–237.

Beresford-Jones, D. G., Whaley, O., Ledesma, C. A., and Cadwallader, L. (2011). Two millennia of changes in human ecology: Archaeobotanical and invertebrate records from the lower Ica Valley, South Coast Peru. Veg. Hist. Archaeobot. 20, 273–292. doi:10.1007/s00334-011-0292-4

Beresford-Jones, D., Lewis, H., and Boreham, S. (2009). Linking cultural and environmental change in Peruvian prehistory: Geomorphological survey of the Samaca Basin, Lower Ica Valley, Peru. Catena 78, 234–249. doi:10.1016/j.catena.2008.12.010

Bird, J., Hyslop, J., and Skinner, M. D. (1985). The Preceramic Excavations at the Huaca Prieta Chicama Valley, Peru. Anthropol. Pap. Am. Mus. Nat. Hist. 62, 1–294.

Bonavia, D. (1996). De la caza-recolección a la agricultura: una perspectiva local. Arqueol. y vida Duccio Bonavia, 97–118. doi:10.4000/books.ifea.5666

Bonavia, D., and Grobman, A. (1979). Sistema de depósitos y almacenamiento durante el período precerámico en la Costa del Perú. J. Soc. des Am. 66, 21–43. doi:10.3406/jsa.1979.2169

Burger, R., and Salazar-Burger, L. (1991). The second season of investigations at the initial period center of Cardal, Peru. J. F. Archaeol. 18, 275–296. doi:10.1179/009346991791548654

Cárdenas, M. (1999). El Periodo Precerámico en el valle de Chao. Boletín Arqueol. PUCP, 141–169

Cavallaro, D. A. (2013). Reconstructing the past: Paleoethnobotanical evidence for ancient Maya plant use practices at the Dos Pilas site, Guatemala. [dissertation/master's thesis]. [Cincinnati (OH)]: University of Cincinnati.

Chevalier, A. (1999). Archaeology+ Ethnobotany= Palaeoethnobotany? Bull. suisse des américanistes, 149–162.

Conlee, C. A. (2000). Late prehispanic occupation of Pajonal Alto, Nasca, Peru: Implications for imperial collapse and societal reformation. University of California, Santa Barbara.

Cutright, R. E. (2013). Household ofrendas and community feasts: ritual at a Late Intermediate Period village in the Jequetepeque Valley, Peru. Ñawpa Pacha 33, 1–21.

doi:10.1179/0077629713z.0000000001

Deagan, K. (2004). Reconsidering Taíno Social Dynamics after Spanish Conquest: Gender and Class in Culture Contact Studies. Am. Antiq. 69, 597–626. doi:10.2307/4128440

Engel, F. (1963). A Preceramic settlement on the Central Coast of Peru: Asia, Unit 1. Trans. Am. Philos. Soc. 53, 1–139.

Feldman, R. A. (1983). "From maritime chiefdom to agricultural state in formative coastal Peru," in Civilization in the Ancient Americas, eds. R. Leventhal and A. Kolata (Univ. New Mexico Press, Albuquerque), 289–310.

Fitzpatrick, S. M. (2006). A Critical Approach to 14C Dating in the Caribbean: Using Chronometric Hygiene to Evaluate Chronological Control and Prehistoric Settlement. Am. Archaeol. 17, 389–418. doi:

Gorbahn, H. (2013). The middle archaic site of Pernil Alto, Southern Peru: the beginnings of horticulture and sedentariness in Mid-Holocene conditions. Diálogo Andin., 61–82. doi:10.4067/s0719-26812013000100005

Hageman, J. B., and Goldstein, D. J. (2009). An integrated assessment of archaeobotanical recovery methods in the Neotropical rainforest of northern Belize: flotation and dry screening. J. Archaeol. Sci. 36, 2841–2852. doi:10.1016/j.jas.2009.09.013

Hammond, N., Clarke, A., and Donaghey, S. (1995). The long goodbye: Middle preclassic Maya archaeology at Cuello, Belize. Lat. Am. Antiq. 6, 120–128.

Helmer, M. (2014). The Archaeology of an ancient seaside town: Performance and community at Samanco, Nepeña Valley Peru. [dissertation/doctor's thesis]. [Norwich (NR)]: University of East Anglia.
Hogg, A. G., Heaton, T. J., Hua, Q., Palmer, J. G., Turney, C. S., Southon, J., et al. (2020). SHCal20 Southern Hemisphere calibration, 0–55,000 years cal BP. Radiocarbon, 62, 759-778. doi: 10.1017/RDC.2020.59

Iriarte, J., and Dickau, R. (2012). ¿Las culturas del maíz?: Arqueobotánica de las sociedades hidráulicas de las tierras bajas sudamericanas. Amazônica 4, 30–58.

Kaplan, L., and Lynch, T. F. (1999). Phaseolus (Fabaceae) in archaeology: AMS radiocarbon dates and their significance for Pre-colombian agriculture. Econ. Bot. 53, 261–272. doi:10.1007/BF02866636

Kimura, B. K., LeFebvre, M. J., deFrance, S. D., Knodel, H. I., Turner, M. S., Fitzsimmons, N. S., et al. (2016). Origin of pre-Columbian guinea pigs from Caribbean archeological sites revealed through genetic analysis. J. Archaeol. Sci. Reports 5, 442–452. doi:10.1016/j.jasrep.2015.12.012

Koons, M. L., and Alex, B. A. (2014). Revised Moche chronology based on Bayesian models of reliable radiocarbon dates. Radiocarbon 56, 1039–1055. doi:10.2458/56.16919

Kosztura, N. J. M. (2020). Cultivares y plantas silvestres en las Terras Pretas de la Amazonia colombiana (Guaviare-Colombia). [dissertation/marster's thesis]. [Bogotá (Bogotá)]: Universidad

Nacional de Colombia.

Lundberg, E. R. (2002). "Tutu pottery and ceramic chronology," in The Tutu archaeological village site. A multi-disciplinary case study in human adaptation, ed. E. Righter (Routledge), 166–198.
Masur, L. J., Millaire, J. F., and Blake, M. (2018). Peanuts and power in the andes: The social archaeology of plant remains from the Virú Valley, Peru. J. Ethnobiol. 38, 589–609. doi:10.2993/0278-0771-38.4.589

Miksicek, C. H. (1991). "The ecology and economy of Cuello. The natural and cultural landscape Preclassic Cuello," in Cuello and early Maya community in Belize, ed. N. Hammond (Cambridge University Press).
Millaire, J. F. (2010). Primary state formation in the Virú Valley, north coast of Peru. Proc. Natl. Acad. Sci. U. S. A. 107, 6186–6191. doi:10.1073/pnas.0911226107

Morcote-Ríos, G. (2006). Tumbas y plantas antiguas del suroccidente colombiano. Boletín Mus. del Oro 54, 46–71.
Morcote, G., and Cavelier, I. (1999). Estrategias adaptativas y subsistencia en grupos humanos precolombinos del Medio Magdalena, Colombia. Rev. Acad. Colomb. Ciencias 23, 41–48.

Newsom, L. A., and Wing, E. S. (2004). On land and sea: Native American uses of biological resources in the West Indies. University of Alabama Press.

Patterson, T. C., and Moseley, M. E. (1968). Late Preceramic and early ceramic cultures of the Central Coast of Peru. J. Andean Archaeol. 6, 115–133. doi:10.1179/naw.1968.6.1.006

Pearsall, D. M. (2003). "Plant food resources of the Ecuadorian Formative: an overview and comparison to the Central Andes," in Archaeology of Formative Ecuador, eds. J. Quilter, J. S. Raymond, and R. L. Burger (Dumbarton Oaks Research Library and Collection Washington, DC.), 213-257.
Pearsall, D. M. (2002). "Analysis of charred botanical remains from the Tutu site," in The Tutu Archaeological Village Site. Routledge, 143-147.

Piperno, D. R., and Pearsall, D. M. (1998). "From Small Scale Horticulture to the Formative Period: The Development of Agriculture," in The origins of agriculture in the Lowland Neotropics (San Diego, CA (USA): Academic Press), 243–320.

Powis, T. G., Stanchly, N., White, C. D., Healy, P. F., Awe, J. J., and Longstaffe, F. (1999). A reconstruction of Middle Preclassic Maya subsistence economy at Cahal Pech, Belize. Antiquity 73, 364–376. doi:10.1017/S0003598X00088311

Pozorski, S. G. (1979). Prehistoric Diet and Subsistence of the Moche Valley, Peru. World Archaeol. 11, 163–184.

Pozorski, S. G. (1987). Theocracy vs. militarism: the significance of the Casma Valley in understanding early state formation. Orig. Dev. Andean state, 15–30.

Pozorski, S., and Pozorski, T. (1990). Reexamining the critical Preceramic/Ceramic Period transition: New data from coastal Peru. Am. Anthropol. 92, 481–491.

Prieto, G. (2015). Gramalote: Domestic life, economy and ritual practices of a Prehispanic maritime community. [dissertation/doctor's thesis]. [New Haven (CT)]: Yale University.

Ramsey, C. B., and Lee, S. (2013). Recent and planned developments of the program OxCal. Radiocarbon 55, 720-730. doi:10.1017/S0033822200057878

Ramsey, C. B. (2009). Bayesian analysis of radiocarbon dates. Radiocarbon 51, 337-360. doi:10.1017/S0033822200033865

Reimer, P. J., Austin, W. E., Bard, E., Bayliss, A., Blackwell, P. G., Ramsey, C. B. et al. (2020) The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 cal kBP). Radiocarbon 62, 725-757. doi: 10.1017/RDC.2020.59

Romero-Picón, Y. (1996). Estudio paleoetnobotánico en la costa Pacifica vallecaucana. Caldasia 18, 155–163.

Scheel-Ybert, R., and Boyadjian, C. (2020). Gardens on the coast: Considerations on food production by Brazilian shellmound builders. J. Anthropol. Archaeol. 60, 101211. doi:10.1016/j.jaa.2020.101211

Shock, M. P. (in press). "As Seen Through the Trees, a Lens into Amazonian Mobility and its Lasting Landscape," in South American Contributions to World Archaeology, eds. M. Bonomo and S. Archila (Springer).

Slotten, V., and Lentz, D. L. (in press) Trees, shrubs, and forests at Joya de Cerén, a Late Classic Mesoamerican village. Quat. Int. doi:10.1016/j.quaint.2021.02.026

Smith, C. E. (1967). "Plant remains," in The Prehistory of Tehuacan Valley. Volume One: Environment and Subsistence, ed. D. S. Byers (Austin: University of Texas Press), 220–255.

Spencer, C. S., Redmond, E. M., and Rinaldi, M. (1994). Drained fields at La Tigra, Venezuelan Llanos: A Regional Perspective. Lat. Am. Antiq., 119–143. doi:10.2307/971559.

Suárez, R. (2011). Arqueología durante la Transición Pleistoceno-Holoceno en Uruguay. BAR International Series 2220.

Suárez, R. (2018). "The Peopling of Southeastern South America: Cultural Diversity, Paleoenvironmental Conditions, and Lithic Technological Organization During the Late Pleistocene and Early Holocene," in Lithic Technological Organization and Paleoenvironmental Change. Global and Diachronic Perspectives, eds. E. Robinson and F. Sellet, 281–300. doi:10.1007/978-3-319-64407-3

Szpak, P., and Chiou, K. L. (2020). A comparison of nitrogen isotope compositions of charred and desiccated botanical remains from northern Peru. Veg. Hist. Archaeobot. 29, 527–538. doi:10.1007/s00334-019-00761-2

Turner, B. L., and Miksicek, C. H. (1984). Economic plant species associated with prehistoric agriculture in the Maya lowlands. Econ. Bot. 38, 179–193. doi:10.1007/BF02858831

Ugent, D., Pozorski, S., and Pozorski, T. (1984). New Evidence for Ancient Cultivation of Canna edulis in Peru. Econ. Bot. 38, 417–432.

Ugent, D., Pozorski, S., and Pozorski, T. (1986). Archaeological Manioc (Manihot) from Coastal Peru. Econ. Bot. 40, 78–102.

Washburn, E., Nesbitt, J., Burger, R., Tomasto-Cagigao, E., Oelze, V. M., and Fehren-Schmitz, L. (2020). Maize and dietary change in early Peruvian civilization: Isotopic evidence from the Late Preceramic Period/Initial Period site of La Galgada, Peru. J. Archaeol. Sci. Reports 31, 102309. doi:10.1016/j.jasrep.2020.102309

Vogel, M., Garren, A., Pacifico, D., and Turner, B. L. (2016). Urban political ecology in late prehistory: New evidence from El Purgatorio, Peru. J. F. Archaeol. 41, 448–466. doi:10.1080/00934690.2016.1195231

Watling, J., Shock, M. P., Mongeló, G. Z., Almeida, F. O., Kater, T., De Oliveira, P. E., et al. (2018). Direct archaeological evidence for Southwestern Amazonia as an early plant domestication and food production centre. PLoS One 13, e0199868. doi:10.1371/journal.pone.0199868

Wild, K. S. (2013). A Timeline of Taíno Development in the Virgin Islands. in International Association for Caribbean Archaeology 25th Congress held in Puerto Rico.

Williams, S. (2013). Palynological investigations during the Colonial Reducciones at Magdalena de Cao Viejo, Peru. [dissertation/master's thesis]. [Pullman (DC)]: Washington State University.

CAPÍTULO II

Evolutionary history of *Psidium guajava* (Myrtaceae) in The Americas inferred by chloroplast and nuclear markers

Autores:

E. Arévalo-Marín, A. Casas, H. Alvarado-Sizzo, E. Ruiz-Sanchez, A. Cosacov, I. Manolopoulou, G. Fermin, J. S. Padilla-Ramírez, C. R. Clement

Evolutionary history of *Psidium guajava* (Myrtaceae) in the Americas inferred by chloroplast and nuclear markers

E. Arévalo-Marín[1,2]*, A. Casas[1]*, H. Alvarado-Sizzo[3], E. Ruiz-Sanchez[4], A. Cosacov[5], I. Manolopoulou[6], G. Fermin[7], J. S. Padilla-Ramírez[8], C. R. Clement[9]

[1]Instituto de Investigaciones en Ecosistemas y Sustentabilidad - IIES, Universidad Nacional Autónoma de México, Morelia, Michoacán, México

[2]Posgrado en Ciencias Biológicas, Unidad de Posgrado, Edificio D, 1º Piso, Circuito de Posgrado, Ciudad Universitaria, Coyoacán, CDMX, México

[3]Laboratorio de Biogeografía y Sistemática, Departamento de Biología Evolutiva, Facultad de Ciencias, Universidad Nacional Autónoma de México, Ciudad de México, México

[4]Departamento de Botánica y Zoología, Centro Universitario de Ciencias Biológicas y Agropecuarias, Universidad de Guadalajara, Zapopan, Jalisco, México

[5]Laboratorio de Ecología Evolutiva-Biología Floral, Instituto Multidisciplinario de Biología Vegetal (IMBIV), CONICET - Universidad Nacional de Córdoba, Córdoba, Argentina

[6]Department of Statistical Science, University College London, London, UK

[7]Instituto Jardín Botánico de Mérida, Facultad de Ciencias, Universidad de Los Andes. Mérida, Mérida, Venezuela

[8]Campo Experimental Pabellón, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP), Pabellón de Arteaga, Aguascalientes, México

[9]Coordenação de Tecnologia e Inovação, Instituto Nacional de Pesquisas da Amazônia, Manaus, Amazonas, Brasil

***Corresponding Authors**
Edna Arévalo-Marín earevalo@cieco.unam.mx
Alejandro Casas acasas@cieco.unam.mx

Abstract

Psidium guajava L. (guava) is a Neotropical semi-domesticated tree widely cultivated because of its fleshy edible fruits. Despite its cultural and economic importance, questions about where and when its evolutionary history originated remain unanswered. Previous hypotheses suggested that guava originated somewhere in South America or in Mesoamerica regions, but little genetic evidence has been published to assess its geographical origin and dispersion process. We used three plastid (*ndhf-rpl32*, *petA-psbj,* and *ycf4-cemA)* and two nuclear (ITS and single-copy gene MeNu39) sequences data from 241 accessions covering almost the entire distribution of guava in the Neotropics. We employed phylogeographic and demographic approaches to determine processes that shaped the guava distribution in the Americas. Our data recovered 24 (chloroplast), 42 (ITS), and 15 (Single-copy MeNu39) haplotypes and rybotypes. Several of the most frequent haplotypes and ribotypes were broadly distributed. However, South American and Mesoamerican lineages were recovered. Bayesian Phylogeographic and Ecological Clustering (BPEC) analysis also supported these findings. Our results indicate that guava could have originated in the Miocene (~13-8 Myr), rapidly diverging into two main lineages: the South American lineage, which diversified ~11 - ~8 Myr, and the Mesoamerican lineage with a diversification of ~5 - ~1 Myr. Our historical demography analyses revealed demographic expansions mainly for the South American lineage. This evidence suggests a continuous historical distribution of the species across the Americas, with independent evolutionary histories occurring in each region. The present sampling of loci did not conclusively resolve the geographic origin of guava. It is plausible that the area of origin was obscured by gene flow, a result of fruit movement by dispersers, or that it is a widely distributed species. This current study provides the first phylogeographical characterization of *Psidium guajava* in its range distribution and provides baseline data for further studies on the evolutionary history of this species.

Keywords: crop evolution, Neotropical crops, origin and diffusion of crops, phylogeography

1. Introduction

The evolutionary history of long-lived fruit trees is particularly complex, with their areas of origin being especially difficult to disentangle (Kantar et al., 2017). The genetic diversity preserved within a species is the result of both historical events and recent evolutionary processes. In this context, the genetic configuration of natural populations reflects the interplay of several population parameters, such as mating systems, effective population sizes, inbreeding, gene exchange, natural selection, and genetic drift (Lewontin, 1974; Slatkin, 1987; Wright, 1969), together with biogeographic processes of dispersion and vicariance (Avise, 2009; Avise et al., 1987). In addition, human activities represent a factor with significant effect on the genetic structure and diversity of economically and culturally relevant fruiting plant populations (Casas & Blancas, 2023; Clement et al., 2021). Over the last three decades, archaeological, genetic, and genomic data have shown the complexity of the evolutionary processes of cultivated plants, some of them characterized by multiple independent domestication events occurring throughout large or diffuse geographic regions (Kantar et al., 2017). Therefore, documenting the evolutionary history of these plants can be even more complex task since wild, feral, and cultivated populations may have coexisted and interacted in the same region for long time (Arévalo-Marín et al., 2021; Clement & Cassino, 2020; Kraft et al., 2014).

The Neotropical areas of Mesoamerica, the Central Andes, and the South American lowlands are home to numerous useful plant species, including some of the most important crops in the world and minor crops with regional significance (Clement et al., 2021; Meyer et al., 2012). Genetic studies employing different types of molecular markers, such as nuclear and plastid sequences and microsatellites, have unraveled the evolutionary history of trees like cocoa (*Theobroma cacao* L.; Zarrillo et al., 2018), the peach palm (*Bactris gasipaes* Kunth; Clement et al., 2017), of papaya (*Carica papaya* L.; Chávez-Pesqueira & Núñez-Farfán, 2016), caimito (*Chrysophyllum cainito* L.; Petersen et al., 2012), chirimoya (*Annona cherimola* Mill.; Larranaga et al., 2017), and jocote (*Spondias purpurea* L.; Miller & Schaal, 2006).

Psidium guajava L., the guava, is one of the Neotropical semi-domesticated (Arévalo-Marín et al., 2021) minor trees most widely cultivated because of its fleshy

fruits (Altendorf, 2018, 2019). It is a hermaphrodite allogamous species (Torres & Gutiérrez, 2018) belonging to the Myrtaceae family (Landrum, 2017). The Neotropical genus *Psidium* includes 90 accepted species, but this number could increase to 122, according to Proença et al. (2022). Five of these species have been proposed as possible wild relatives of guava: *P. guineense* Sw., *P. guyanense* Pers., *P. nutans* O. Berg., *P. rostratum* McVaugh, and *P. rutidocarpum* Ruiz & Pav. (Landrum, 2021). Although the biogeographical history of the species of the genus is unknown, the Brazilian biomes of the Atlantic Forest, the Cerrado, and the Caatinga are recognized as the center of *Psidium* diversity, hosting 50 species (Landrum, 2017).

Currently, *P. guajava* occurs from Mexico and the Antilles to northwestern Argentina (Landrum, 2017). Despite being a genetic resource of great economic and cultural importance, its evolutionary history and dispersal throughout the Americas are unclear (Arévalo-Marín et al., 2021). Therefore, recognizing the ancestral genetic inheritance in guava populations is necessary since, currently, there are no known wild populations for the species, and the conservation of its genetic diversity is not guaranteed.

In a recent review, based on natural history and archaeological and historical information about guava, we introduced a series of hypotheses considered useful to guide research on the evolutionary history of this species (Arévalo-Marín et al., 2021). We proposed South America as the geographic area of origin of the species, and species of the megafauna as the main dispersers towards other suitable ecosystems of South America and a subsequent dispersion towards Mesoamerica by humans (Arévalo-Marín et al., 2021). Through this study, we provide an overview of the evolutionary context of *P. guajava* in the Americas, looking for testing the South American origin and the dispersion of guava from South to North. We hypothesized that (1) If South America were the origin region of guava, we would find there the highest genetic diversity and the highest number of unique genetic variants, despite the current wide distribution of guava, (2) we expected that Mesoamerica would harbor South American lineages and could have unique genetic variants derived from South American lineages. But in addition, we considered possible to expect (3) deepest/oldest phylogenetic divergence events driven by biogeographic processes through evolutionary time. This study aim: 1)

provide detailed insights about the phylogeographical diversity pattern of guava in the Americas, 2) identify possible genetic divergence between guava samples from South America and Mesoamerica regions, 3) elucidate the phylogenetic relationships among guava lineages, and 4) recognize biogeographic and/or anthropogenic, and demographic processes that could explain the current guava distribution.

Using nuclear (nDNA) and plastid (cpDNA) genetic markers and covering the almost entire range of guava distribution, we analyzed genealogical relationships, the genetic diversity and structure, divergence times, and demographic history. Ultimately, we seek to elucidate the evolutionary history and a likely area of origin of this important tree species.

2. Materials and Methods

2.1 Sampling and study areas

We studied samples of 243 guava trees. Leaf tissue samples were obtained from 84 individuals in the guava Germplasm banks of the Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP) in Aguas Calientes, Mexico; 22 from the Tropical Agricultural Research and Higher Education Center (CATIE) in Turrialba, Costa Rica; and 34 from the Corporación Colombiana de Investigación Agropecuaria (Agrosavia) in Palmira, Colombia. Colombian and Mexican germplasm collections included guava landraces and modern breeding lines from each country. Mexican, Costa Rican, and Colombian germplasm banks also provided some accessions from Puerto Rico, Colombia, and Venezuela; Guatemala, Honduras, El Salvador, and Cuba, respectively (Table S1). Another set of 103 South American accessions from Venezuela, Peru (one population from the Andean region and three populations from the Amazonas), and Brazil (western Amazon) (Table S1).

2.2 DNA extraction and sequencing

Total DNA was isolated using a modified CTAB protocol (Doyle & Doyle, 1987). We chose three plastid intergenic spacers: *ndhf-rpl32*, *petA-psbj,* and *ycf4-cemA* considering Shaw et al. (2014). Guava oligos were designed for this study using the Primer 3 software (Table S2). Additionally, two bi-parentally inherited nuclear regions

(hereafter nDNA) were used, the nuclear ribosomal internal transcribed spacer (ITS), and a single copy nuclear gene (MeNu39) developed for the genus *Metrosideros* (Pillon et al., 2014). Sequences of *P. guineense*, *P. cattleyanum,* and *P. friedrichsthalianum*, used as outgroups, were generated in this study. Other species of *Psidium* also used as an outgroup for the ITS dating analysis were obtained from (Vasconcelos et al., 2017).

We performed amplification by polymerase chain reaction (PCR) in an 11 µL final reaction volume using 3 µL Platinum Multiplex PCR Master Mix (Applied Biosystems, Foster City, CA, USA), 6 µL PCR grade H_2O, 1.0 µL DNA template (50 ng/µL), and 1.0 µL G/C enhancer volume to assure amplification success. PCRs were carried out in MultiGene OptiMax (Labnet International, Inc., Edison, NJ, USA) or a 2700 thermal cycler (Applied Biosystems, Foster City, CA, USA). For *ycf4-cemA*, *ndhf-rpl32*, and *petA-psbj*, we used a primer annealing temperature at 62°C for 1 min, while we used 60ºC for ITS and MeNu39. Additional cycling conditions were implemented following manufacturer directions. We sequenced amplicons with a forward primer at Macrogen USA Maryland headquarters (https://lims.psomagen.com). Plastid and nuclear sequences were trimmed and aligned using the Geneious Prime v. 2019.2 software. Because the chloroplast is typically inherited uniparentally without recombination in angiosperms, the three plastid regions (*ndhf-rpl32*, *petA-psbj*, and *ycf4-cemA*) were combined for the following analyses. We coded indels as single binary characters using the simple indel coding method (Simmons & Ochoterena, 2000). All Sanger sequences showed polymorphism.

2.3 Haplotype Relationships

We obtained chloroplast and nuclear haplotypes using DnaSP v. 5.10.01 (Rozas & Rozas, 1999). In the case of nuclear sequences containing ambiguities, haplotypes were reconstructed using the coalescent-based Bayesian method PHASE v2.1 (Stephens et al., 2001; Stephens & Scheet, 2005). We estimated the genealogical relationships among haplotypes and ribotypes for each dataset using the TCS Network distance method implemented in PopArt 1.7 software (Leigh & Bryant, 2015), and geographical distributions of haplotypes were mapped using QGIS v. 3.20 Odense (QGIS.org, 2021).

We resolved ambiguous connections (loops) in networks of haplotypes using the frequency and topological criteria (Crandall & Templeton, 1993).

2.4 Bayesian Phylogeographic and Ecological Clustering (BPEC)

We used the Bayesian phylogeographic and ecological clustering (BPEC), implemented in the R package '*BPEC*' (Manolopoulou et al., 2020) to determine geographical population clusters for *P. guajava*. The method explores different haplotype trees and migration events and provides an entire posterior distribution over phylogeographic clustering through the Markov chain Monte Carlo (MCMC) sampler. Additionally, under a coalescent-based migration-mutation model, BPEC accounts for haplotype connection ambiguities caused by loops, identifies missing haplotypes, and estimates posterior probabilities (Manolopoulou et al., 2020). This analysis assumes that population substructure is the result of migration events into new sites, which can be explained by geographical and ecological restrictions to gene flow (Manolopoulou et al., 2020). The clustering and migration model do not explicitly consider the geographical distance between clusters. It only separates observations in distinct geographical clusters. Therefore, it is possible for migration to result in two distant clusters (Manolopoulou et al., 2020). We conducted the BPEC analysis on a two-dimensional dataset (longitude, latitude) for all cpDNA and nDNA haplotypes. Since BPEC was designed primarily for non-recombinant DNA sequences, we assumed non-recombination for nuclear markers. For concatenated cpDNA, we set the maximum number of migrations in 3, (Max-Mig = 3) and the parsimony relaxation parameter in 4 (ds = 4). For ITS and the single-copy gene MeNu39 the settings were Max-Mig = 3 and ds = 2. We ran two MCMS chains for 5×10^7, 1×10^7, and 2×10^5 iterations for cpDNA ITS and MeNu 39, with parameters sampled every 1000 steps.

2.5 Molecular diversity and genetic structure

For concatenated plastid (cpDNA) and two nuclear markers (nDNA), we calculated standard diversity indices in the "*pegas*" and "*ape*" R package (Paradis, 2010; Paradis et al., 2004), including gene diversity (h), and nucleotide diversity (π), and considering the genetic groups found in BPEC analysis. To assess the level of geographic structure

among regions: South America (Brazilian and Peruvian Amazon [AMA-BR and AMA-PER], Peruvian Andes [PER-AND], and northern South America [NSA] areas), Mesoamerica (Mexico and Central America), and the Antilles, we carried out analyses of molecular variance (AMOVA; Excoffier et al., 1992), using Arlequin 3.1.1 (Excoffier et al., 2005). We conducted the AMOVA using 10,000 permutations and ΦST (pairwise differences) to determine genetic differentiation within and among regions.

2.6 Divergence time estimates

We reconstructed phylogenetic relationships among haplotypes and ribotypes using Bayesian inference (BI) with BEAST v2.6.7 (Bouckaert et al., 2019). We used the nucleotide substitution models HKY for cpDNA and MeNu39 dataset, and TrNef + G for ITS under the Bayesian Information Criteria (BIC) suggested by jModelTest v.2.2.10 (Darriba & Posada, 2012; Guindon & Gascuel, 2003). We ran independent analyses two times for 1×10^7 generations each time for all markers, sampling parameters every 1000 steps.

Based on time estimation for the *Psidium* group previously elucidated using pollen fossils (Vasconcelos et al., 2017), we used the timing of *Psidium* group diversification as a calibration point in our dating analyses. The *Psidium* group dating 25.62 Myr (95%CI: 24.14-31.07 Myr; Vasconcelos et al., 2017) was used to calibrate the root of our tree, using a normal distribution prior. We used an uncorrelated lognormal relaxed clock to allow for rate variation among lineages and the Yule model as a prior for branching rates. We used Tracer v1.7.2 (Rambaut et al., 2018) to verify high effective sample sizes (ESS > 200) of each parameter; log and tree files were combined using LogCombiner and summarized as a maximum clade credibility tree using TreeAnnotator (Bouckaert et al., 2019) with a burn-in of 25%. We visualized the topologies using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

2.7 Demographic History

We conducted demographic analyses based on the phylogenetic results. We estimated changes in population size over time in the main lineages identified (South America and Mesoamerica lineages) using Bayesian Skyline Plots analysis (BSPs; Drummond et al.,

2005) in BEAST v.2.6.7. We performed the BSPs in the piecewise-constant coalescent Bayesian skyline tree prior to 5 starting groups. Two MCMCs starting with a random tree were run for 4 and 1 x 107 generations with parameters sampled every 4000 and 1000 steps for cpDNA lineages A and B, respectively. For ITS and MeNu39 markers, we used 3 and 5 x 107 generations, sampling every 3000 and 5000 steps, respectively. The priors for these analyses were the same as those used in the BI haplotype tree reconstruction, as previously described.

3. Results

3.1 Phylogeographic patterns

We identified 24 haplotypes for the cpDNA-concatenated data (Figure 1a-b, Table S1). Network haplotypes revealed three common haplotypes, H1, H3, and H15 (Figure 1b). Haplotype H1 was the most frequent, occurring in all sampled countries (108 samples, 48.8% of individuals; Figure 1a). Haplotype H15, the second most frequent (37 samples, 16.7% of individuals), was located exclusively in the South America region, specifically in northern South America (NSA), the Peruvian Andes (PER-AND), and the Amazon (BR-AMA and PER-AMA) (Figure 1a-b). Given its high number of connections and abundance, H1 haplotype is probably the most ancestral haplotype in our network. Haplotype 3 was found in both Mesoamerica (9 individuals) and South America (15 individuals; NSA + BR-AMA and PER-AMA) and was the third most frequent haplotype, being present in 10% of our samples (Fig. 1a-b). Haplotypes H9 and H11 were found in Mesoamerica and South America, although in low frequency (1.8% and 0.9%, respectively; Figure 1a-b). We recovered six private haplotypes from Mesoamerican samples (H2, H4, H5, H6, H8, and H10; Figure 1b). Thirteen private South American haplotypes of low frequency were also recovered: H7, H12, H15, and H20–22 from the Amazon, H13–H14, and H18–H19 from NSA, and H23–H24 from the PER-AND (Figure 1a-b), and a unique haplotype from the Antilles (H17, Figure 1b). The most polymorphic localities were identified in South America (Loreto in the PER-AMA and Huánuco in the PER-AND with six and four haplotypes, respectively (Table S1).

Figure 1. (a) Geographic distribution of 24 haplotypes of guava depicted in the network (H1-H24). (b) Haplotype network based on combined cpDNA haplotypes. The small black rhombus represents missing haplotypes. The size of each circle is proportional to the number of samples (n) harboring a certain haplotype, with the smallest circles representing n=1 and the largest representing n=108.

The ITS-aligned data set for 227 samples of *P. guajava* (454 phases) yielded 44 ribotypes (Figure 2a-b). R1 was found in 64 (105 phases) of the 227 samples, being the most frequent and widespread ribotype (Figure 2a). We also found that ribotypes R2, R5, and R8 were at high frequencies (Figure 2a-b). R2 ribotype, with 88 phases (53 individuals), is present mainly in the Mesoamerican region and includes two samples from NSA (Figure 2a-b). The ribotype R5 (83 phases) is present mainly in the Amazon, along with some samples of NSA and Mesoamerica. R8 ribotype (73 phases) is present in the BR-AMA, PER-AMA, PER-AND, NSA, and some samples from Mesoamerica (Table S1; Figure 2a-b). Other ribotypes occurring in the two regions were R7, R10, R12, R16, R18, and R20. We recovered 12 private ribotypes from Mesoamerica and none from the Antilles. We found 22 private ribotypes from South America: ten from NSA (H23-H25, H27, H29, and H30-34), ten from the BR-AMA and PER-AMA (H35-H44), and two shared for both areas (H26 and H28; Figure 2b, Table S1).
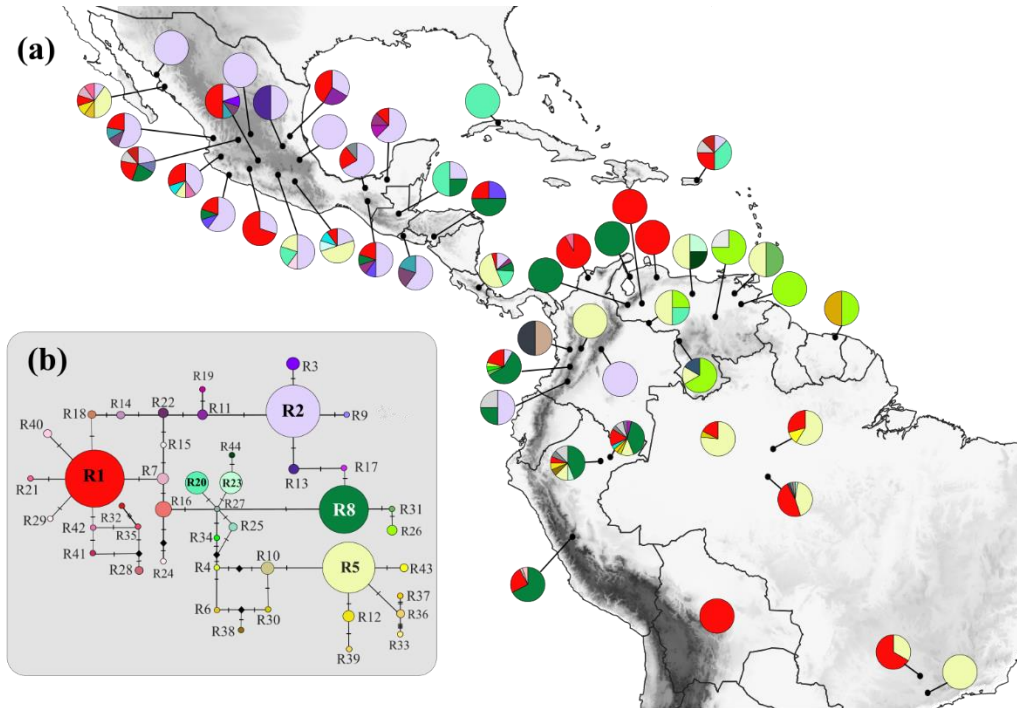
Figure 2. (a) Geographic distribution of 44 ribotypes of guava depicted in the network (R1-R44). (b) Haplotype network based on ITS ribotypes. The small black circles represent missing ribotypes. The size of each circle is proportional to the number of samples (n) harboring a certain ribotype, with the smallest circles representing n = 1 and the largest representing n = 105.

For the single-copy nuclear marker MeNu39, we reconstructed 15 haplotypes (denominated here with the letter M; Figure 3a-b; Table S1). Haplotype M1 was shared between Mesoamerica and South America (Figure 3a-b) and was the most frequent haplotype found in 90% of samples, being harbored in almost all provenances except El Salvador (Figure 3a). Haplotypes M2–M4 and M7–M9 were private to Mesoamerica (Figure 3b), M5–M6 (from Brazilian germplasm collection), M10–M11 (from NSA), and M12 and M15 (from the PER-AND) were the unique from South America (Figure 3b). The locality of Huánuco in the Peruvian Andes was the highest polymorphic site, with five private haplotypes (Figure 3a).

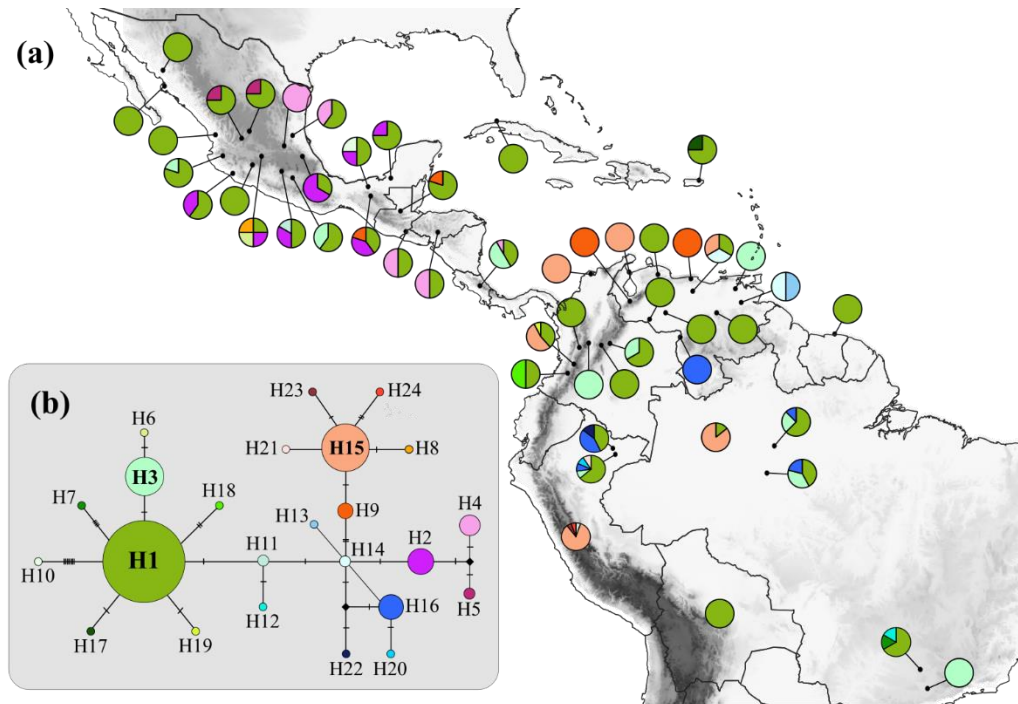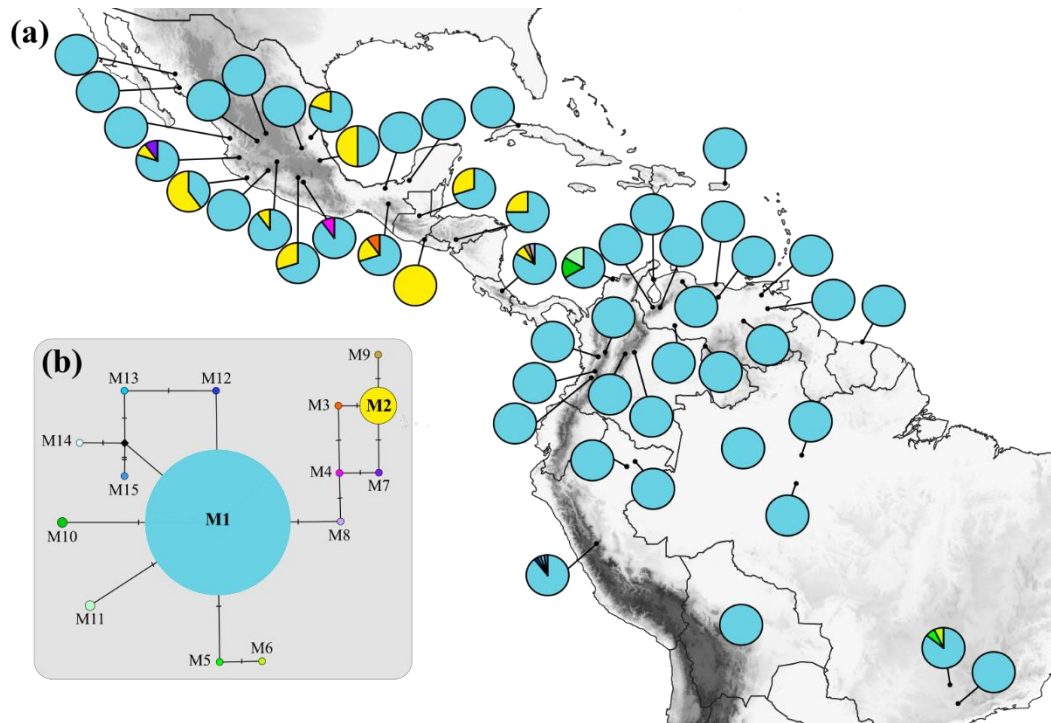**Figure 3.** (a) Geographic distribution of 15 haplotypes of guava depicted in the network (M1-M15). (b) Haplotype network based on single-copy nuclear gene MeNu39 haplotypes. The small black circles represent missing haplotypes. The size of each circle is approximately proportional to the number of samples (n) harboring a certain haplotype, with the smallest circles representing n = 1 and the largest representing n = 429.

3.2 Spatial Structure – BPEC

Based on cpDNA, the BPEC parsimony network identified 30 haplotypes: the 24 haplotypes detected in this study and previously mentioned (Figure 1), and six missed additional haplotypes (Figure 4a). The haplotype groupings in the parsimony tree (Figure 4a) correspond to the main lineages detected by BEAST (Figure 4a) and the two main subclades within lineage A and the unique divergent clade within lineage B. BPEC clustering assigned haplotypes to one of two phylogeographic clusters with high posterior probabilities (PP = 0.81 – 0.99), except the common H1 haplotype, which was divided between Clusters 1 and 3 (PP = 0.52 and 0.47, respectively). Cluster 1 (Figure 4a) encloses 13 private haplotypes from South America, samples of H1 haplotypes, the haplotype from Antilles, and two private haplotypes from Mesoamerica (H6 and H8). The South American haplotypes H15, H23, and H24 had low posterior assignment

probabilities for the Cluster 2 (PP = 0.07 and 0.16, respectively; Figure 4a). However, given the higher posterior probabilities of these three haplotypes belonging to Cluster 1 (PP = 0.92 and 0.82, respectively, Figure 4a). Clusters 1 and 2 were considered a single group called the South American cluster since 86% of samples were from this region. Cluster 3 includes the remaining private Mesoamerican haplotypes (H2, H4, H5, H10) and some samples from H1, henceforth the Mesoamerican cluster. The estimated number of migration events was two (PP = 0.82).

BPEC analysis showed a strong association between genetic groups and geographic areas for nuclear markers. For the ITS, the BPEC analysis identified the 44 ribotypes shown in Figure 2 and 10 undetected ribotypes (Figure 4b). Most ribotypes showed PP > 0.86, belonging to one of the three main clusters retrieved. We discarded an additional cluster that had very low assignment probabilities. Cluster 1 includes 22 South American private ribotypes, six shared ribotypes (R1, R5, R7, R10, R12, R16 and R20) and Mesoamerican exclusive R4 and R6 (Figure 4b). The R8 ribotype, which brings together mainly South American individuals, was assigned to Cluster 2. However, because its posterior density largely overlaps with Cluster 1, we consider Cluster 1 and 2 a single genetic cluster, henceforth the South American cluster (90% of samples from South America). Cluster 3 is predominantly Mesoamerican (Figure 4b), including shared ribotypes R2 and R18. This analysis also identified two probable migration events (PP = 1).

Regarding single-copy gene MeNu39, BPEC found 16 haplotypes, of which only one was undetected. The common and shared haplotype M1, which includes 90% of samples, showed membership divided among Clusters 1 and 3 with PP = 0.44, and 0.55, respectively (Figure 4c). BPEC assigned six (M10 - M15) of the eight private South American haplotypes to Cluster 1 with PP = 1 (Figure 4c), and the other two haplotypes (M5 and M6; PP = 1) were nested in Cluster 2 (Figure 4c). All private haplotypes from Mesoamerica, with PP = 1, were assigned to Cluster 3 (Figure 4c). The three clusters will be called from now on as: South American-1 (Cluster 1), South American-2 (Cluster 2) and Mesoamerican cluster (Cluster 3). The number of migrations estimated with this marker was two (PP = 1).

There were no loops that led to ambiguous connections between haplotypes and ribotypes for no markers.
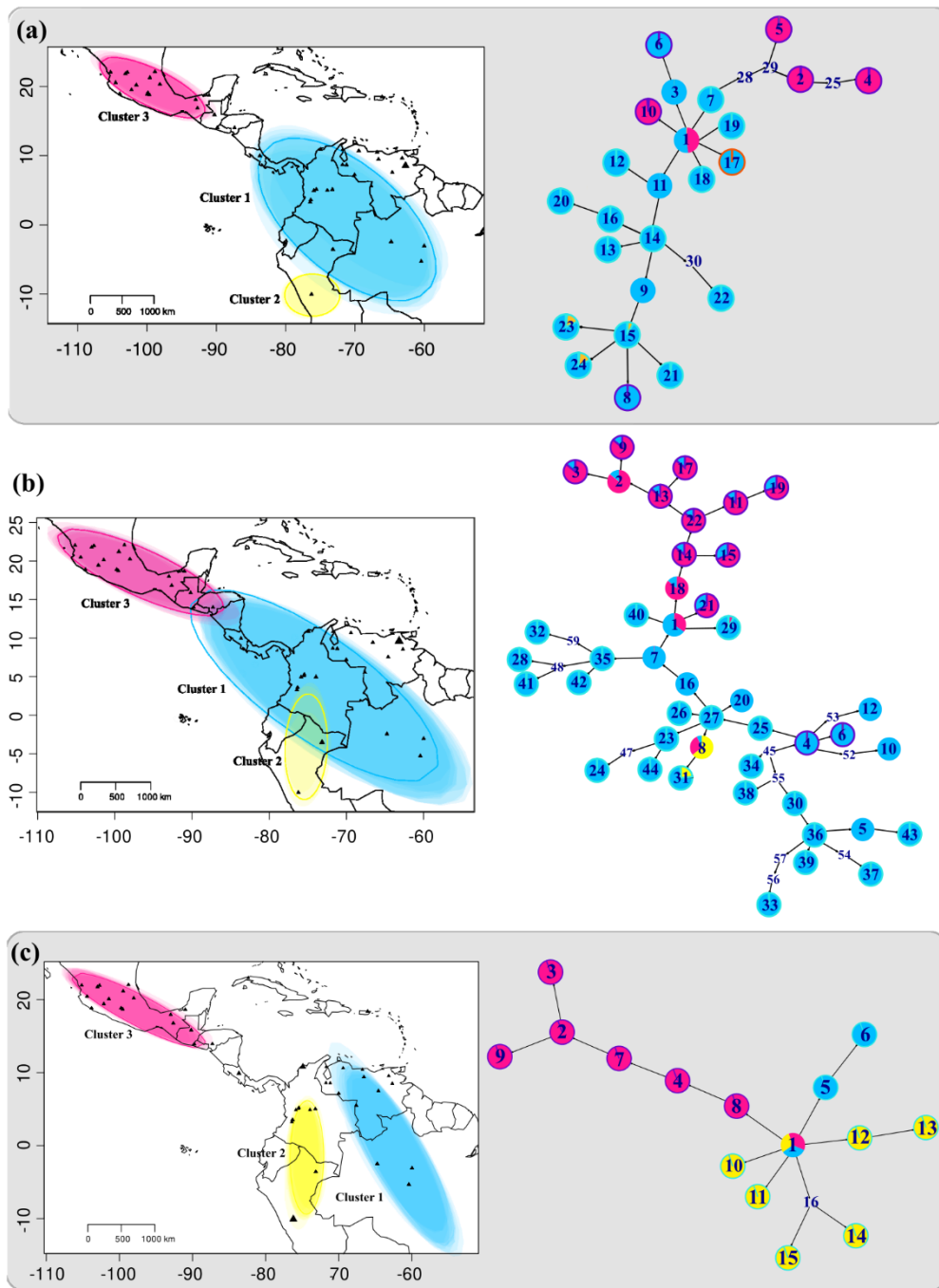

Figure 4. Clustering maps of populations based on (a) cpDNA combined, (b) ITS, and (c) single-copy MeNu39 sequences from a Bayesian phylogeographic clustering analysis as implemented in BPEC. The identified clusters are shown in different colors, and their spatial coverage is drawn with a transparent ellipse representing the 50% highest posterior density in each run,

whereas the solid lines represent posterior means. Each map is accompanied by the corresponding haplotype network showing the coalescent relationships among haplotypes and K = 2 or K=3 clusters. The haplotype number ID is the same as in Figure 1B, 2b, and 3b. The color of each haplotype denotes its most likely cluster membership to the three different lineages, and the green, orange, and blue color on the edges of each circle represents if they are private haplotypes for Mesoamerica, the Antilles, or South America, respectively.

3.3 Molecular diversity patterns

Of the 243 samples of guavas collected, we compiled sequence data for 221 individuals for plastid *ndhf-rpl32* (701 bp), *petA-psbj* (856 bp), and *ycf4-cemA* (744 bp) spacers, 227 individuals for ITS, and 236 for MeNu39. The combined plastid matrix comprised 2301 bp with 34 polymorphic sites and 10 parsimony informative sites. The aligned sequences of ITS and MeNu39 were 770 and 452 bp long, with 22 and 14 polymorphic sites and 19 and 9 parsimony informative sites, respectively.

Both cpDNA and ITS markers had higher haplotype diversity in the South America cluster compared with the Mesoamerican cluster (Table 1). However, nucleotide diversity was low in all groups analyzed. Contrary to the cpDNA and ITS, the single-copy gene MeNu39 showed higher haplotype diversity values for Mesoamerican compared to South American-1 cluster. The AMOVA analyses showed highly significant genetic structure among regions for all markers (Table S3). Within regions, there was also significant structure among populations.

Table 1. Estimate of genetic variability of cpDNA, ITS, and MeNu 39 markers for *P. guajava*

Genetic clusters	N	N_H	S	*h*	π
cp					
South American	141	20	26	0.780 ± 0.0004	0.00130 ± 5.7e-07
Mesoamerican	80	5	11	0.434 ± 0.0039	0.00071 ± 2.1e-07
ITS					
South American	340	33	21	0.806 ± 0.00015	0.00586 ± 1.0e-05
Mesoamerican	122	14	9	0.496 ± 0.00304	0.00161± 1.2e-06
MeNu39					
South American-1	266	7	7	0.059 ± 0.00041	0.00018 ± 1.4e-07
South American-2	2	2	1	1.000 ± 0.125	0.00210 ± 8.8e-06
Mesoamerican	196	7	5	0.289 ± 0.00145	0.00228 ± 2.7e-06

N = number of samples sequenced, N_H = number of unique haplotypes, S = number of segregating sites, h = haplotype diversity, π = nucleotide diversity, D = Tajima's D, F = Fu's F). Bold numbers indicate summary values.

3.4 Divergence time estimation

Based on cpDNA data, the most recent common ancestor (MRCA) of *P. guajava* lineages was dated 12.22 Mya (95% highest posterior density [HPD] = 5.57 − 21.79 Mya, Figure 5). Two main lineages, A and B (Figure 5a) were retrieved within *P. guajava* and showed, agreement with the genetic groups primarily found in the BPEC analysis. The lineage A, dating back to 10.06 Mya (HPD: 4.07 − 19.72 Mya), splits into two subclades. The first subclade (A1) dates from 7.03 Mya (HPD: 2.17 − 15.04 Mya), and the second subclade (A2) is about 8.28 Mya, although it presents a low support value (Figure 5a-b). The lineage B, dating back to 5.57 Mya (HPD: 0.5 − 14.04 Mya), is a Mesoamerican lineage (Figure 5a-b).

Figure 5. (a) Chronogram of *Psidium guajava* haplotypes and other *Psidium* species based on the maximum credibility tree from the Bayesian dating analysis using the *ndhf-rpl32/petA-psbj/ycf4-cemA* combined dataset. The numbers above the branches are time scale (Mya with 95% HPD intervals in square brackets). Ages in geological time are shown at the base of the tree. (b) The haplotype networks. Color coding in a and b indicates the assignment of each ribotype to a genetic group related with the figure 4a.
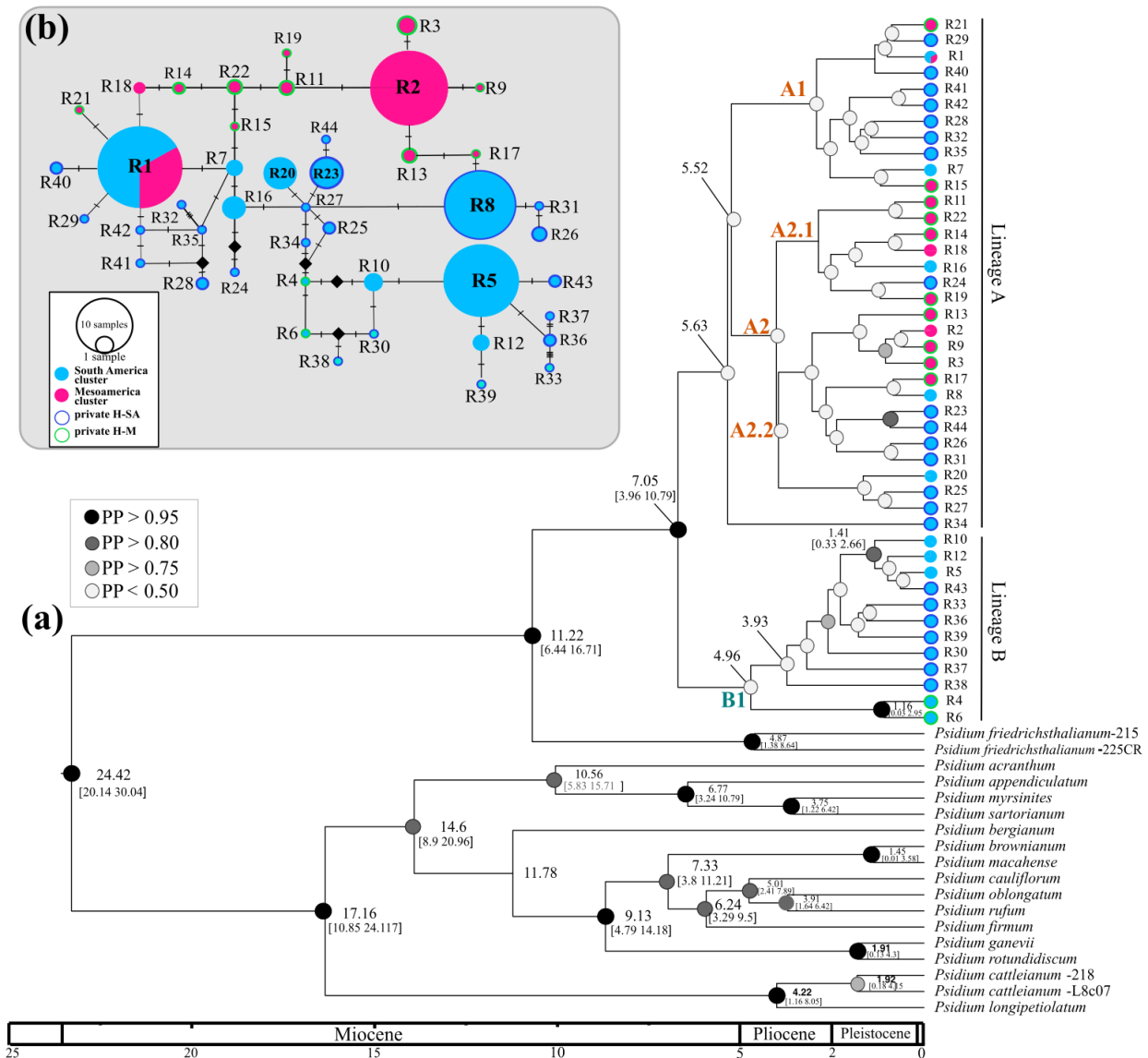
Figure 6. (a) Chronogram of *Psidium guajava* haplotypes and other *Psidium* based on the maximum credibility tree from the Bayesian dating analysis using the ITS-aligned. The numbers above the branches are time scale (Mya with 95% HPD intervals in square brackets). Ages in

geological time are shown at the base of the tree. (b) The ribotype network. Color coding in a and b indicates the assignment of each ribotype to a genetic group related with the figure 4b.

The phylogenetic tree using the ITS dataset suggests a guava age between 11 (95% HPD = 6.44 – 16,71 Mya) and 7.05 Mya (95% HPD = 3.96 – 10,79 Mya; Figure 6a), followed by rapid diversification. Two main lineages were recovered, nesting most of the retrieved ribotypes. The low-supported lineages A and B, with median ages of diversification around 5.63 and 4.96 Mya, respectively (Figure 6a-b). Lineage A is formed by two main sub-clades. In the subclade A1, we find ribotypes mainly from the genetic group South American. Subclade A2 presents two main nodes. Node A2.1 is a predominantly Mesoamerican clade. Meanwhile, subclade A2.2 shows a Mesoamerican clade as the sister group of the South American subclade (Figure 5a-b). Lineage B is represented by haplotypes belonging to the South American cluster (Figure 6a).
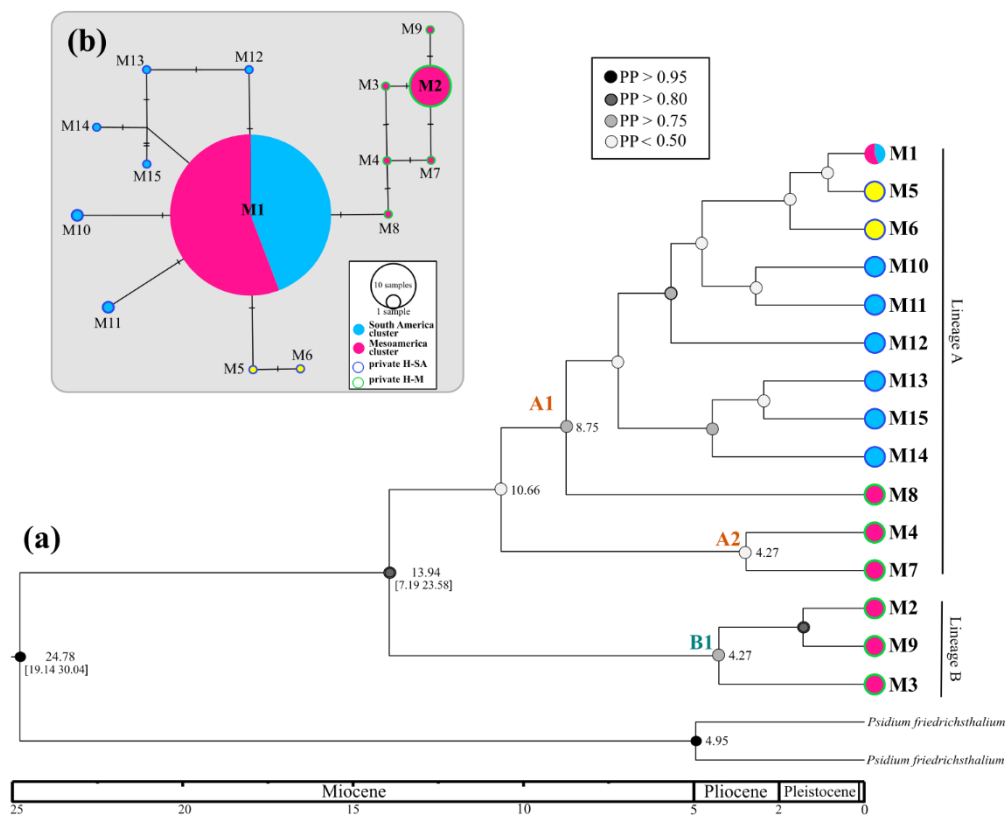
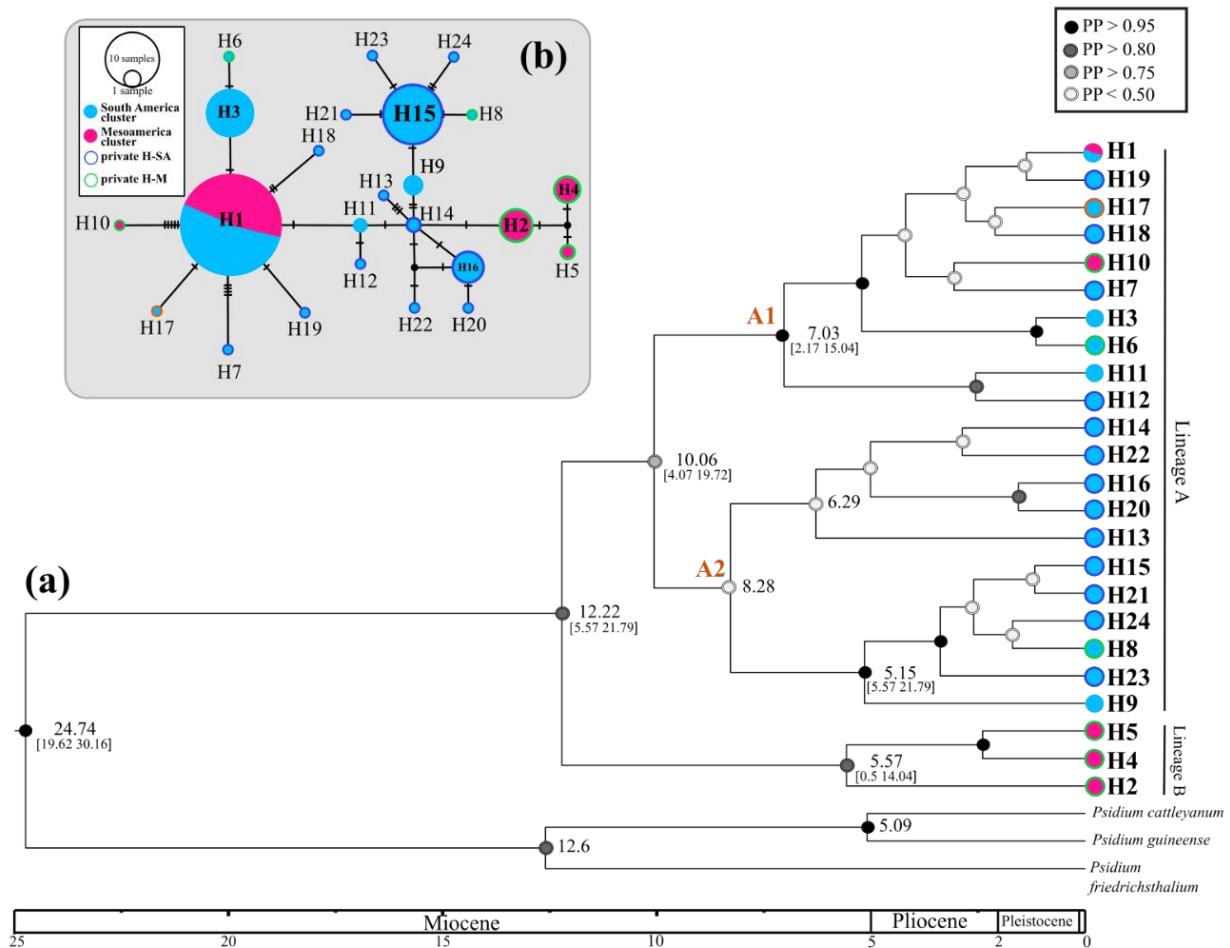**Figure 7.** (a) Chronogram of *Psidium guajava* haplotypes and other *Psidium* based on the maximum credibility tree from the Bayesian dating analysis using the single-copy nuclear gene

MeNu39 haplotypes dataset. The numbers above the branches are time scale (Mya with 95% HPD intervals in square brackets). (b) The haplotype networks. Color coding in a and b indicates the assignment of each ribotype to a genetic group related with the figure 4c.

Based on the haplotypes of the MeNu39 marker, guava has an average age of 13.94 Mya (95% HPD: 7.19 – 23.58 Mya: Figure 7a) and a later diversification into two lineages (A and B). Lineage A diversified around 10 Mya into two clades. Clade A1 contains the widely distributed haplotype M1, the South American genetic groups, and Mesoamerican tip M8 as its sister lineage (Figure 7a-b). Lineage B diversified more recently, around 4.27 Mya, and contains the most haplotypes belonging to Mesoamerican genetic group (Figure 7a-b). However, due to their low statistical support, the diversification dates of the two lineages shown by this marker should be taken with caution.

The phylogenetic analyzes of guava showed polyphyletic relationships for all markers since there is no concordance between the clades and the recovered genetic groups.

3.5 Demography

Lineage A showed an increase in effective population size (Ne) dating back to ~0.4 – 0.2 and ~0.1 Myr for cpDNA and ITS, respectively (Figure 8a-b). In contrast, the BSP analysis showed population stability for lineage B for the same markers (Figure 8a-b). The single-copy MeNu 39 gene suggests a slight expansion over the last ~1.2 Myr years ago for Lineage A and B (Figure 8c). However, the confidence interval indicates that this result should be taken with caution.
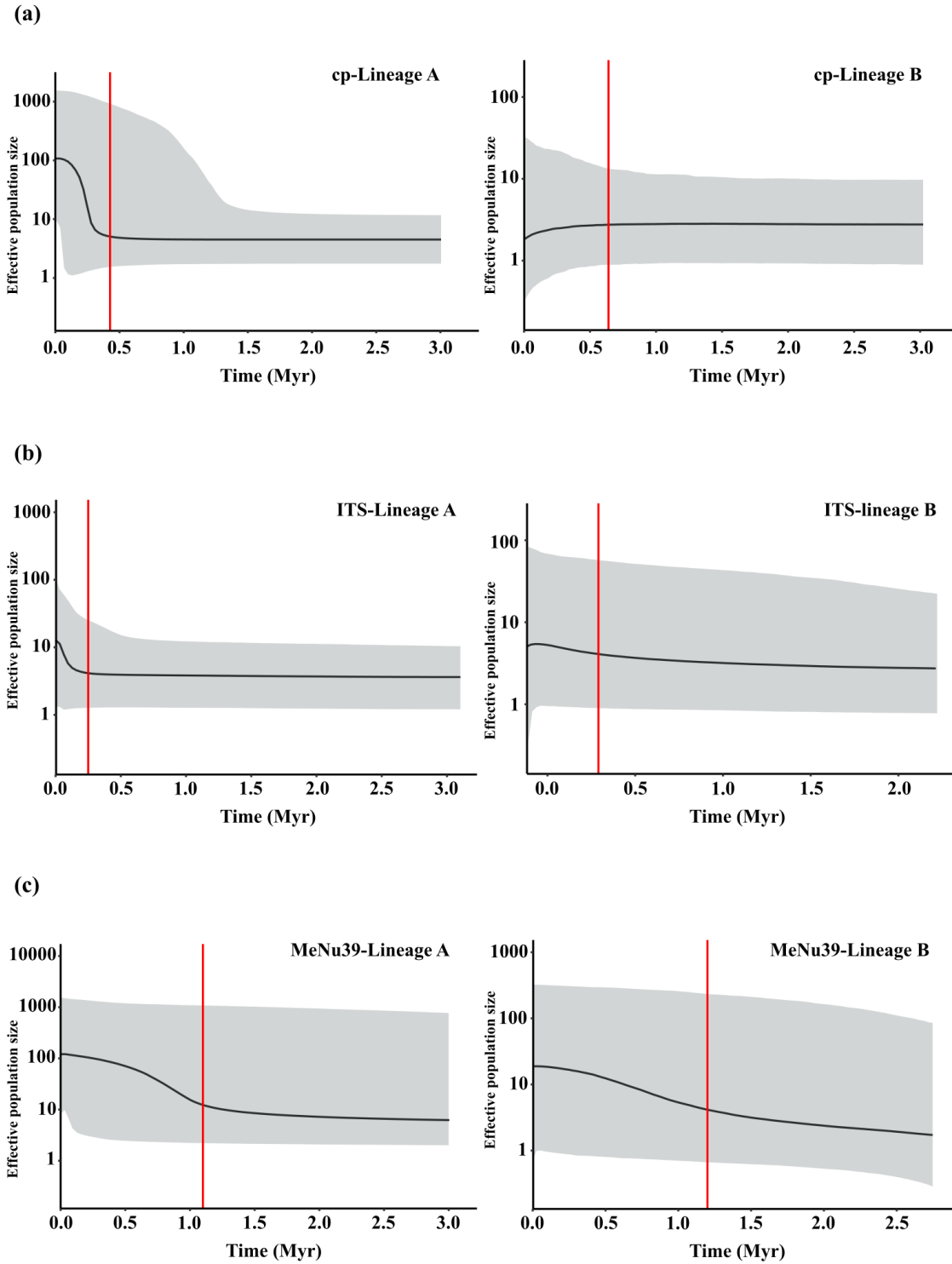
Figure 8. Bayesian skyline plot analysis for the effective population size over time for South American accessions (a) cpDNA, (b) ITS, (c) single-copy gene MeNu39 for Mesoamerican and South American region.

Discussion

Testing hypotheses about the centers of origin of Neotropical fruits has been a complex task because economically and culturally important perennials have been broadly dispersed by humans (Clement et al., 2021; Meyer et al., 2012; van Zonneveld et al., 2018), and their cultivated populations consist of individuals derived from multiple geographic areas (Meyer & Purugganan, 2013). Nevertheless, the impact of geological and paleo-climate events on the distribution and geographical structure of these species is poorly understood. Our study provides a comprehensive analysis of the origin and possible processes that determined the current distribution of genetic diversity of guava, using a dataset comprising its almost entire distribution range.

Phylogeographycal patterns

According to network analyses, the most frequent guava haplotypes and ribotypes are broadly distributed on the continent. However, the gene genealogies of the three markers recovered specific relationships among guava haplotypes and ribotypes from Mesoamerica and South America, agreeing with the genetic clusters found in Bayesian phylogeographic clustering analyses (BPEC). Within population genetics, BPEC finds substructure only in South America with the single copy nuclear marker, which separates the private Amazonian M5 and M6 haplotypes from the rest of the South American haplotypes. Shared haplotypes and ribotypes between regions would suggest historical or recent gene flow between Mesoamerica, South America, and the Antilles. Therefore, the genetic clusters show some degree of overlapping between geographic distributions. This pattern also has been reported to *Palicourea padifolia* (Gutiérrez-Rodríguez et al., 2011), and *Capsicum annum* (Aguilar-Melendez et al., 2009). Therefore, these results suggest an early continuous historical distribution of the guava across the American continent, with two independent evolutionary histories occurring/mainly associated to in Meso- and South America. On the other hand, South American populations, particularly the Andes and Amazonia, bore many private haplotypes and ribotypes for the chloroplast and nuclear markers, suggesting a long-term accumulation of variation and, therefore, a possible region of private haplotypes.

Historical analyses

Our phylogenetic analyses placed the most recent common ancestor of *P. guajava* between ~13 and ~7 Myr (cpDNA and nDNA, respectively) during the middle or late Miocene, with subsequent intraspecific diversifications (Fig. 4-6). Important diversification events during the Miocene have also been reported in other groups of plants in the Neotropics such as Chrysobalanaceae, Melastomataceae, Verbenaceae (Hughes et al., 2013), and Acanthaceae, Buseraceae, Clusisaceae, and Malvaceae (de Queiroz et al., 2017; De-Nova et al., 2012; Dick et al., 2003). Such diversification could have been associated with global climatic and geological changes (Hughes et al., 2013).

The diversification during the Miocene has been recorded in many clades of the Myrteae tribe (to which the genus *Psidium* belongs) as *Eugenia* and *Myrcia* in various biomes of South America as Chaco, Cerrado, Caatinga, and the Mata Atlantica. These regions are recognized as a center of diversity for various Myrtaceae genera, including the *Psidium* genus (De Souza Neto et al., 2022; Landrum, 2017; Lima et al., 2021). Unfortunately, our study does not include guava samples from the southern end of its distribution: the Humid Chaco areas in Paraguay and northern Argentina. However, since the range of distribution of the species extends to this biome (Arévalo-Marín et al., 2021; Landrum, 2017), it would be advisable to test the biogeographic hypothesis of the origin of guava in the Humid Chaco (Arévalo-Marín et al., 2021). In addition, all markers used in this research revealed dates long before human presence in the Americas for the diversification of the main guava lineages (~10 - ~5 Mya; Figures 4-6), agreeing with the pattern explained before. Therefore, the hypothesis about the arrival of guava in Mesoamerica as result those anthropogenic activities (Arévalo-Marín et al., 2021) is rejected.

Additionally, the presence of single haplotypes from Mesoamerica nested within South American associated lineages and the opposite case would corroborate the hypothesis of migration and genetic flow between Meso and South America. This process would also dilute phylogenetic structure by interleaving haplotypes within geographically structured lineages. The estimated dating for diversification of guava lineages is also earlier than the formation of the Isthmus of Panama, which occurred approximately 3 million years ago (O'Dea et al., 2016). However, according to

geophysical, sedimentological, and molecular data demonstrate that the collision of the arc with South America in the Miocene generated narrower and shallower interoceanic seaways (O'Dea et al., 2016). These changes would have allowed successful dispersals of terrestrial animals and plants between the continents over these straits, as it is supported by several intercontinental dispersals that happened in the Miocene according to fossil records (O'Dea et al., 2016). These are for instance the cases of monkeys, ants, and palms, among others (review in O'Dea et al., 2016). Therefore, our data suggest that guava could have dispersed before the formation of the Isthmus of Panama. Likely, the diffusion and the exchange of genetic information could have been promoted through seed dispersal by the megafauna, flying mammals, and birds, affecting the pattern of genetic structure in guava. Several studies have shown evidence of plant material exchange between Meso and South America long before the union of the two continents (Cody et al., 2010; Pennington & Dick, 2004). Species of the genus *Cedrela* (Muellner et al., 2010) and *Casearia* (de Mestier et al., 2022) as well as species of the family Melastomataceae (Michelangeli et al., 2022) are some examples recorded.

In this line of evidence, the lineage B (Figures 4-6) show an early divergence with little diversification, probably associated with small population size. Unfortunately, there are not previous biogeographic studies for other congeneric species to compare, being this report the first biogeographic reconstruction for a *Psidium* species. Regarding the number of migration events between South America and Mesoamerica, all markers revealed both, shared and exclusive independent lineages for Mesoamerica and South America. Such pattern indicates more than one dispersal event in different geological times; consistently, the BPEC analyses suggest the occurrence of two migration events. It is well recognized that the evolution of Neotropical plants is profoundly influenced by climate changes and tectonic events and that radiations are often associated with significant geological events (Richardson et al., 2001). Several phylogeographic studies show that during the Pleistocene, there were long periods of cooling and a rapid expansion of polar ice sheets accompanied with climatic fluctuations that included warmer periods which influenced the diversification of different groups of plants (references in Ornelas et al., 2013; Ramírez-Barahona & Eguiarte, 2013; J. E. Richardson et al., 2001; Turchetto-Zolet et al., 2013). However, these studies show that

not only the Pleistocene climatic fluctuations would have played an important role in the divergence and increase in the effective population size of plants, but also the past orographic processes that occurred during the Neogene in America (i.e. volcanic and tectonic events, mountain uplift, alteration in drainage systems) (Turchetto-Zolet et al., 2013). Therefore, we believe that these processes could also have influenced the increase in the effective population size of guava lineages according to the adaptive radiation model (Stroud & Losos, 2016).

Guava origin

The present sampling of loci did not conclusively resolve the geographic origin of guava. Even main lineages can be associated to Meso and South America, and it is plausible that the area of origin was obscured by gene flow, a result of fruit movement by dispersers. It is also possible that this result is due to the type of sampling. Our study was based mainly on the collection of plant material in germplasm banks, which housed both cultivated and feral trees, and feral individuals collected in Brazil, Peru, and Venezuela. Although P. guajava is widely distributed in the Americas, it is widely cultivated, and there are no precise records of natural populations, although these are likely to be found in areas in southern of Paraguay and northern Argentina in the Humid Chaco (Arévalo-Marín et al., 2021). Thus, we should continue to expand the sampling range in the future and identify and collect as many wild guava samples as possible.

Conclusion

A sound reconstruction of the origin of a perennial Neotropical tree of wide distribution, such as *Psidium guajava*, is a highly challenging task since it requires extensive sampling through the whole area of distribution. Nevertheless, our sample provides the first insight to understand the evolutionary history of this species. Bayesian phylogeographic clustering analyses recovered two genetic clusters within *P. guava* each one associated with a main geographic area (Mesoamerica or South America). However, network analyzes indicated the absence of completely divergent regional patterns, and shared haplotypes and ribotypes that would suggest historical gene flow and migration events between Mesoamerica, South America, and the Antilles. Moreover,

these results also suggest a continuous historical distribution of the species across the whole Neotropics, with independent evolutionary histories occurring in each region of the Americas. Likewise, ancient guava lineages would demonstrate a pre-human distribution in the Americas, while archaeological information available identifies a relatively late use of guava by Mesoamerican people. Therefore, the guava distribution was foremost shaped mainly by biogeographical processes rather than by anthropogenic factors. However, further studies with increased sampling area and molecular markers with higher resolution will increase clarify about the evolutionary history of this interesting plant species.

administrative staff of IIES-UNAM, campus Morelia, Mexico especially, Benjamín Mora, Juan Carlos Mata, and Claudia Lenina Sánchez for their support in the importation of plant material.

References

Aguilar-Melendez, A., Morrell, P. L., Roose, M. L., Kim, S.-C. C., Araceli, A. M., Morrell, P. L., Roose, M. L., Kim, S.-C. C., Aguilar-Meléndez, A., Morrell, P. L., Roose, M. L., & Kim, S.-C. C. (2009). Genetic diversity and structure in semiwild and domesticated chiles (*Capsicum Annuum*; Solanaceae) from Mexico. *American Journal of Botany*, *96*(6), 1190–1202. https://doi.org/10.3732/ajb.0800155

Altendorf, S. (2018). Minor tropical fruits: Mainstreaming a niche market. In *Food Outlook Biannual Report on Global Food Markets* (Issue July, pp. 67–74). http://www.fao.org/3/a-I8080e.pdf

Altendorf, S. (2019). *Major tropical fruits market review 2018*. http://www.fao.org/3/ca5692en/CA5692EN.pdf

Arévalo-Marín, E., Casas, A., Landrum, L., Shock, M. P., Alvarado-Sizzo, H., Ruiz-Sanchez, E., & Clement, C. R. (2021). The Taming of *Psidium guajava*: Natural and Cultural History of a Neotropical Fruit. *Frontiers in Plant Science*, *12*(September), 1–15. https://doi.org/10.3389/fpls.2021.714763

Avise, J. C. (2009). Phylogeography: Retrospect and prospect. *Journal of Biogeography*, *36*(1), 3–15. https://doi.org/10.1111/j.1365-2699.2008.02032.x

Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A., & Saunders, N. C. (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, *18*(1), 489–522.

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., & De Maio, N. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *15*(4), e1006650.

Casas, A., & Blancas, J. (2023). Human Cultures and Plant Diversity in the Mountains of Mexico: An Introduction. In *Ethnobotany of the Mountain Regions of Mexico* (pp. 1–22). Springer.

Chávez-Pesqueira, M., & Núñez-Farfán, J. (2016). Genetic diversity and structure of wild populations of *Carica papaya* in northern Mesoamerica inferred by nuclear microsatellites and chloroplast markers. *Ann Bot*, *118*(7), 1293–1306. https://doi.org/10.1093/aob/mcw183

Clement, C. R., Casas, A., Parra-Rondinel, F. A., Levis, C., Peroni, N., Hanazaki, N., Cortés-Zárraga, L., Rangel-Landa, S., Alves, R. P., Ferreira, M. J., Cassino, M. F., Coelho, S. D., Cruz-Soriano, A., Pancorbo-Olivera, M., Blancas, J., Martínez-Ballesté, A., Lemes, G., Lotero-Velásquez, E., Bertin, V. M., & Mazzochini, G. G. (2021). Disentangling Domestication from Food Production Systems in the Neotropics. *Quaternary*, *4*(1), 4. https://doi.org/10.3390/quat4010004

Clement, C. R., & Cassino, M. F. (2020). Landscape domestication and archaeology. *Encyclopedia of Global Archaeology*, 6431–6438.

Clement, C. R., de Cristo-Araújo, M., Coppens d'Eeckenbrugge, G., dos Reis, V. M., Lehnebach, R., Picanço-Rodrigues, D., Cristo-Araújo, M. de, Coppens d'Eeckenbrugge, G., Reis, V. M. dos, Lehnebach, R., & Picanço-Rodrigues, D. (2017). Origin and dispersal of domesticated peach palm. *Frontiers in Ecology and Evolution*, *5*(NOV), 1–19. https://doi.org/10.3389/fevo.2017.00148

Cody, S., Richardson, J. E., Rull, V., Ellis, C., & Pennington, R. T. (2010). The great American biotic interchange revisited. *Ecography*, *33*(2), 326–332.

Crandall, K. A., & Templeton, A. R. (1993). Empirical Tests of Some Predictions from Coalescent Theory with Applications to Intraspecific Phylogeny Reconstruction. *Genetics*, *134*, 959–969. http://www.genetics.org/content/genetics/134/3/959.full.pdf

Darriba, D., & Posada, D. (2012). jModelTest 2.0 Manual. *Nature Methods2*, *9*(8), 772. https://doi.org/10.1093/molbev/msn083

de Mestier, A., Brokamp, G., Celis, M., Falcón-Hidalgo, B., Gutiérrez, J., & Borsch, T. (2022). Character evolution and biogeography of Casearia (Salicaceae): Evidence for the South American origin of a pantropical genus and for multiple migrations to the Caribbean islands. *Taxon*, *71*(2), 321–347.

de Queiroz, L. P., Cardoso, D., Fernandes, M. F., & Moro, M. F. (2017). Diversity and evolution of flowering plants of the Caatinga domain. *Caatinga: The Largest Tropical Dry Forest Region in South America*, 23–63.

De Souza Neto, J. D., Dos Santos, E. K., Lucas, E., Vetö, N. M., Barrientos-Diaz, O., Staggemeier, V. G., Vasconcelos, T., & Turchetto-Zolet, A. C. (2022). Advances and perspectives on the evolutionary history and diversification of Neotropical Myrteae (Myrtaceae). *Botanical Journal of the Linnean Society*, *199*(1), 173–195.

De-Nova, J. A., Medina, R., Montero, J. C., Weeks, A., Rosell, J. A., Olson, M. E., Eguiarte, L. E., & Magallón, S. (2012). Insights into the historical construction of species-rich Mesoamerican seasonally dry tropical forests: the diversification of *Bursera* (Burseraceae, Sapindales). *New Phytologist*, *193*(1), 276–287.

Dick, C. W., Abdul-Salim, K., & Bermingham, E. (2003). Molecular systematic analysis reveals cryptic tertiary diversification of a widespread tropical rain forest tree. *The American Naturalist*, *162*(6), 691–703.

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, *19*, 11–15.

Drummond, A. J., Rambaut, A., Shapiro, B., & Pybus, O. G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, *22*(5), 1185–1192.

Excoffier, L., Laval, G., & Schneider, S. (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, *1*, 47–50.

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, *131*(2), 479–491.

Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, *52*(5), 696–704.

Gutiérrez-Rodríguez, C., Ornelas, J. F., & Rodríguez-Gómez, F. (2011). Chloroplast DNA phylogeography of a distylous shrub (*Palicourea padifolia*, Rubiaceae) reveals past fragmentation and demographic expansion in Mexican cloud forests. *Molecular Phylogenetics and Evolution*, *61*(3), 603–615.

Hughes, C. E., Pennington, R. T., & Antonelli, A. (2013). Neotropical plant evolution: assembling the big picture. *Botanical Journal of the Linnean Society*, *171*(1), 1–18.

Kantar, M. B., Nashoba, A. R., Anderson, J. E., Blackman, B. K., & Rieseberg, L. H. (2017). The Genetics and Genomics of Plant Domestication. *BioScience*, *67*(11), 971–982. https://doi.org/10.1093/biosci/bix114

Kraft, K. H., Brown, C. H., Nabhan, G. P., Luedeling, E., Ruiz, J. de J. L., d'Eeckenbrugge, G. C., Hijmans, R. J., Gepts, P., Luna Ruiz Jde, J., Coppens d'Eeckenbrugge, G., Hijmans, R. J., & Gepts, P. (2014). Multiple lines of evidence for the origin of domesticated chili pepper, *Capsicum annuum*, in Mexico. *Proceedings of the National Academy of Sciences*, *111*(17), 6165–6170. https://doi.org/10.1073/pnas.1308933111

Landrum, L. R. (2017). The genus *Psidium* (Myrtaceae) in the state of Bahia, Brazil. *Canotia*, *13*, 1–101.

Landrum, L. R. (2021). *Psidium guajava* L.: taxonomy, relatives and possible origin. In S. Mitra (Ed.), *Guava: botany, production and uses* (pp. 1–21). Cab International. https://doi.org/10.1079/9781789247022.0001

Larranaga, N., Albertazzi, F. J., Fontecha, G., Palmieri, M., Rainer, H., van Zonneveld, M., & Hormaza, J. I. (2017). A Mesoamerican origin of cherimoya (*Annona cherimola* Mill.): Implications for the conservation of plant genetic resources. *Molecular Ecology*, *26*(16), 4116–4130. https://doi.org/10.1111/mec.14157

Leigh, J. W., & Bryant, D. (2015). Popart: full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*(9), 1110–1116.

Lewontin, R. C. (1974). *The genetic basis of evolutionary change* (Vol. 560). Columbia University Press New York.

Lima, D. F., Goldenberg, R., Forest, F., Cowan, R. S., & Lucas, E. J. (2021). Phylogeny and biogeography of *Myrcia* sect. Aguava (Myrtaceae, Myrteae) based on phylogenomic and Sanger data provide evidence for a Cerrado origin and geographically structured clades. *Molecular Phylogenetics and Evolution*, *157*, 107043.

Manolopoulou, I., Hille, A., & Emerson, B. (2020). BPEC: An R package for Bayesian phylogeographic and ecological clustering. *Journal of Statistical Software*, *92*(5). https://doi.org/10.18637/jss.v092.i05

Meyer, R. S., DuVal, A. E., & Jensen, H. R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *The New Phytologist*, *196*(1), 29–48. https://doi.org/10.1111/j.1469-8137.2012.04253.x

Meyer, R. S., & Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nature Reviews Genetics*, *14*(12), 840–852.

Michelangeli, F. A., Nicolas, A. N., Ocampo, G., Goldenberg, R., Almeda, F., Judd, W. S., Bécquer, E. R., Skean Jr, J. D., Kriebel, R., & Sosa, K. (2022). Historical biogeography of the Neotropical Miconieae (Melastomataceae) reveals a pattern of progressive colonization out of lowland South America. In *Systematics, Evolution, and Ecology of Melastomataceae* (pp. 629–644). Springer.

Miller, A. J., & Schaal, B. A. (2006). Domestication and the distribution of genetic variation in wild and cultivated populations of the Mesoamerican fruit tree *Spondias purpurea* L. (Anacardiaceae). *Molecular Ecology*, *15*(6), 1467–1480. http://onlinelibrary.wiley.com/doi/10.1111/j.1365-294X.2006.02834.x/abstract

Muellner, A. N., Pennington, T. D., Koecke, A. V., & Renner, S. S. (2010). Biogeography of *Cedrela* (Meliaceae, Sapindales) in central and South america. *American Journal of Botany*, *97*(3), 511–518.

O'Dea, A., Aguilera, O., Aubry, M. P., Berggren, W. A., Budd, A. F., Cione, A. L., Coates, A. G., Collins, L. S., Coppard, S. E., & Cozzuol, M. A. (2016). *Building Bridges. Response to Erkens and Hoorn: "The Panama Isthmus,'old','young'or both?"*.

Ornelas, J. F., Sosa, V., Soltis, D. E., Daza, J. M., González, C., Soltis, P. S., Gutierrez-Rodriguez, C., de los Monteros, A. E., Castoe, T. A., & Bell, C. (2013). Comparative phylogeographic analyses illustrate the complex evolutionary history of threatened cloud forests of northern Mesoamerica. *PloS One*, *8*(2), e56283.

Paradis, E. (2010). pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics*, *26*(3), 419–420.

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*(2), 289–290.

Pennington, R. T., & Dick, C. W. (2004). The role of immigrants in the assembly of the South American rainforest tree flora. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *359*(1450), 1611–1622.

Petersen, J. J., Parker, I. M., & Potter, D. (2012). Origins and close relatives of a semi-domesticated neotropical fruit tree: *Chrysophyllum cainito* (Sapotaceae). *Am J Bot*, *99*(3), 585–604. https://doi.org/10.3732/ajb.1100326

Pillon, Y., Johansen, J., Sakishima, T., Chamala, S., Barbazuk, W. B., & Stacy, E. A. (2014). Primers for low-copy nuclear genes in Metrosideros and cross-amplification in Myrtaceae. *Appl Plant Sci*, *2*(10). https://doi.org/10.3732/apps.1400049

Proença, C. E. B., Tuler, A. C., Lucas, E. J., Vasconcelos, T. N. da C., de Faria, J. E. Q., Staggemeier, V. G., de-Carvalho, P. S., Forni-Martins, E. R., Inglis, P. W., da Mata, L. R., & da Costa, I. R. (2022). Diversity, phylogeny and evolution of the rapidly evolving genus *Psidium* L. (Myrtaceae, Myrteae). *Annals of Botany*, *129*(4), 367–388. https://doi.org/10.1093/aob/mcac005

QGIS.org. (2021). *QGIS Geographic Information System.* QGIS Association.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904.

Ramírez-Barahona, S., & Eguiarte, L. E. (2013). The role of glacial cycles in promoting genetic diversity in the Neotropics: the case of cloud forests during the Last Glacial Maximum. *Ecology and Evolution*, *3*(3), 725–738.

Richardson, J. E., Pennington, R. T., Pennington, T. D., & Hollingsworth, P. M. (2001). Rapid diversification of a species-rich genus of neotropical rain forest trees. *Science*, *293*(5538), 2242–2245.

Rozas, J., & Rozas, R. (1999). DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics (Oxford, England)*, *15*(2), 174–175.

Shaw, J., Shafer, H. L., Leonard, O. R., Kovach, M. J., Schorr, M., & Morris, A. B. (2014). Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: the tortoise and the hare IV. *Am J Bot*, *101*(11), 1987–2004. https://doi.org/10.3732/ajb.1400398

Simmons, M. P., & Ochoterena, H. (2000). Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology*, *49*(2), 369–381.

Slatkin, M. (1987). Gene flow and the geographic structure of natural populations. *Science*, *236*(4803), 787–792.

Stephens, M., & Scheet, P. (2005). Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *The American Journal of Human Genetics*, *76*(3), 449–462.

Stephens, M., Smith, N. J., & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *The American Journal of Human Genetics*, *68*(4), 978–989.

Stroud, J. T., & Losos, J. B. (2016). Ecological opportunity and adaptive radiation. *Annual Review of Ecology, Evolution, and Systematics*, *47*, 507–532.

Torres, M. de L., & Gutiérrez, B. (2018). A Preliminary Assessment of the Genetic Diversity and Population Structure of Guava, *Psidium guajava*, in San Cristobal. In *Understanding Invasive Species in the Galapagos Islands* (pp. 3–17). https://doi.org/10.1007/978-3-319-67177-2_1

Turchetto-Zolet, A. C., Pinheiro, F., Salgueiro, F., & Palma-Silva, C. (2013). Phylogeographical patterns shed light on evolutionary process in S outh A merica. *Molecular Ecology*, *22*(5), 1193–1213.

van Zonneveld, M., Larranaga, N., Blonder, B., Coradin, L., Hormaza, J. I., & Hunter, D. (2018). Human diets drive range expansion of megafauna-dispersed fruit species. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(13), 3326–3331. https://doi.org/10.1073/pnas.1718045115

Vasconcelos, T. N., Proenca, C. E., Ahmad, B., Aguilar, D. S., Aguilar, R., Amorim, B. S., Campbell, K., Costa, I. R., De-Carvalho, P. S., Faria, J. E., Giaretta, A., Kooij, P. W., Lima, D. F., Mazine, F. F., Peguero, B., Prenner, G., Santos, M. F., Soewarto, J., Wingler, A., & Lucas, E. J. (2017). Myrteae phylogeny, calibration, biogeography and diversification patterns: Increased understanding in the most species rich tribe of Myrtaceae. *Mol Phylogenet Evol*, *109*, 113–137. https://doi.org/10.1016/j.ympev.2017.01.002

Wright, S. (1969). *Evolution and the genetics of populations: Vol. 2. The theory of gene frequencies*.

Zarrillo, S., Gaikwad, N., Lanaud, C., Powis, T., Viot, C., Lesur, I., Fouet, O., Argout, X., Guichoux, E., & Salin, F. (2018). The use and domestication of *Theobroma cacao* during the mid-Holocene in the upper Amazon. *Nature Ecology & Evolution*, *2*(12), 1879–1888.

SUPPORTING INFORMATION

Evolutionary history of *Psidium guajava* (Myrtaceae) in the Americas inferred by chloroplast and nuclear markers

E. Arévalo-Marín, A. Casas, H. Alvarado-Sizzo, E. Ruiz-Sanchez, A. Cosacov, I. Manolopoulou, G. Fermín, José Saúl Padilla-Ramírez, C. R. Clement

Table S1. Sampling sites, including country, locality, geographical coordinates of sample collection, geographical region, and cultivation status (wild and cultivated) are indicated. The number of collected individuals, successful amplifications, and of nuclear and plastid haplotypes and ribotypes are shown.

Source (Germplasm bank or in situ collection)	Country	Locality	Latitude	Longitude	Cult. Status	No. Samples	No. successful amplification*	cp	ITS	MeNu39
Mesoamerica										
CATIE	Costa Rica	Cartago	9.890327	-83.6557	Cultivated	12	12	H1, H3, H4	R1, R2, R5, R8, R14, R20, R22	M1, M8, M9
CATIE	El Salvador	La Montañita	13.9400333	-89.7822083	Cultivated	2	2	H1, H4	R2, R11	M1
CATIE	Guatemala	Alta Verapaz	15.8281827	-90.1927846	Cultivated	6	5	H1, H9	R2, R8, R20	M1, M2
CATIE	Honduras	Ojojona	13.940869	-87.30106	Cultivated	2	2	H1, H4	R1, R8, R13	M1, M2
INIFAP	Mexico	Aguascalientes	21.927489	-102.721297	Cultivated	5	5	H1, H5	R2, R10	M1
INIFAP	Mexico	Campeche	18.616422	-90.992626	Cultivated	4	4	H1, H2	R1, R2, R14, R22	M1
INIFAP	Mexico	Chiapas	16.767764	-92.639609	Cultivated	5	5	H1, H2, H9	R1, R2, R8, R13, R22	M1, M2, M3
INIFAP	Mexico	Colima	18.8258333	-103.8166666	Cultivated	5	5	H1, H2	R1, R2, R8, R13	M1, M2
INIFAP	Mexico	Estado de México	18,841868	-99,769493	Cultivated	5	5	H1, H2, H11	R2, R5, R19, R20	M1, M2
INIFAP	Mexico	Guanajuato	20.0905556	-101.4441666	Cultivated	5	5	H1, H2, H6, H8	R1, R2, R9, R10, R11	M1, M2
INIFAP	Mexico	Guerrero	18.685636	-99.562506	Cultivated	5	5	H1. H3	R1, R2, R4. R5, R12	M1, M4

INIFAP	Mexico	Jalisco	20.428789	-104.427889	Cultivated	5	5	H1, H3	R1, R2, R5, R12, R15	M1, M2, M7
INIFAP	Mexico	Michoacán	19.403402	-102.13775	Cultivated	5	5	H1	R1, R2	M1
INIFAP	Mexico	Nayarit	21.9508333	-105.160	Cultivated	4	4	H1	R1, R2, R11	M1
INIFAP	Mexico	Puebla	20.193046	-98.046105	Cultivated	3	3	H1, H2	R2	M1, M2
INIFAP	Mexico	Querétaro	21.170469	-99.48995	Cultivated	2	2	H4	R2, R3	M1
INIFAP	Mexico	San Luis Potosí	22.016234		Cultivated	5	5	H1, H4	R1, R2, R3	M1, M2
INIFAP	Mexico	Sinaloa	25.939549	-109.147159	Cultivated	5	5	H1	R1, R2, R5, R6, R7, R12, R21	M1
INIFAP	Mexico	Sonora	27.030591	-109.487185	Cultivated	5	3	H1	R2	M1
INIFAP	Mexico	Tabasco	17.986888	-92.895879	Cultivated	4	4	H1, H2, H10	R1, R2, R7	M1
INIFAP	Mexico	Zacatecas	21.733601	-103.007573	Cultivated	5	5	H1, H5	R1, R2, R8, R16, R17, R18	M1
South America										
INIFAP	Bolivia	Santa Cruz	-16.7261	-64.435	Cultivated	1	1	H1	R1	M1
INIFAP	Brazil	São Paulo (Jaboticaba)	-21.2347222	-48.2858333	Cultivated	5	5	H1, H7,	R1, R5	M1, M5, M6
AGROSAVIA	Brazil	São Paulo (Piracicaba)	-22.7252808	-47.6491699	Cultivated	2	2	H3	R5	M1
Collected	Brazil	Amazonas (Manaus)	-3.095222	-59.995023	Feral/Cultivated	10	9	H1, H3, H16	R1, R5, R43, R44	M1
Collected	Brazil	Amazonas (Manicoré)	-5.329275979	-60.42794396	Cultivated	24	21	H1, H3, H16	R1, R5, R10, R26, R28, R41, R42	M1
Collected	Brazil	Amazonas (Maraã)	-2.48529397	-64.74743597	Cultivated	9	9	H1, H15	R1, R5, R10, R36	M1
AGROSAVIA	Colombia	Atlántico	10.804112	-74.921853	Cultivated	6	6	H15	R1, R29	M1, M10, M11
AGROSAVIA	Colombia	Boyacá	5.030505	-73.2082752	Cultivated	3	3	H1, H3	-	M1

AGROSAVIA	Colombia	Caldas	5.0455556	-75.51527778	Cultivated	1	1	H3	R5	M1
AGROSAVIA	Colombia	Cauca	3.220608	-76.415466	Cultivated	2	2	H1, H18	R2, R8, R16	M1
AGROSAVIA	Colombia	Cundinamarca	4.90933	-73.9408	Cultivated	1	1	H1	R2	M1
AGROSAVIA	Colombia	Risaralda	4.8891694	-75.8707416	Cultivated	1	1	H1	R32, R33	M1
AGROSAVIA	Colombia	Valle del Cauca	3.537971	-76.327617	Cultivated	14	13	H1, H15, H19	R1, R2, R8, R25, R31, R34	M1
Collected	Peru	Loreto (Barrio Florido1)	-3.590701	-73.157125	Feral	12	12	H1, H3, H16, H20, H21	R1, R5, R7, R8, R12, R16, R35, R36, R37	M1
Collected	Peru	Loreto (Barrio Florido2)	-3.603265	-73.132726	Feral	8	8	H1, H16, H22,	R1, R5, R7, R8, R16, R26, R38, R39	M1
Collected	Peru	Huánuco	-10.0805556	-76.2222222	Cultivated	20	20	H11, H15, H23, H24	R1, R8, R16, R18, R40	M1, M12, M13, M14, M15
AGROSAVIA	Surinam	Paramaribo	5.852051	-55.203828	Cultivated	1	1	H1	R23, R30	M1
Collected	Venezuela	Amazonas	5.5072222	-67.6138888	Feral	3	3	H16	R5, R23, R28,	M1
Collected	Venezuela	Apure	7.1466667	-70.0233333	Feral	2	2	H1	R5, R20, R23	M1
Collected	Venezuela	Bolívar	7.5005556	-64.5899999	Feral	2	2	H1	R23, R24	M1
Collected	Venezuela	Guárico	9.390777	-66.580161	Feral	3	3	H1, H14, H15	R5, R26, R27	M1
Collected	Venezuela	Lara-Falcón	10.62	-69.4136111	Feral	1	1	H1	R1	M1
Collected	Venezuela	Mérida	8.569174	-71.176143	Cultivated	1	1	H1	R1	M1
Collected	Venezuela	Miranda	10.409827	-66.771195	Cultivated	1	1	H9	-	M1

Collected	Venezuela	Monagas	8.4825	-62.68555556	Feral	3	3	H13, H14	R23	M1
Collected	Venezuela	Monagas (Sucre)	9.5530556	-63.1755555	Feral	1	1	H3	R5, R25	M1
Collected	Venezuela	Táchira	8.5911111	-71.7605555	Feral	1	1	H9	R8	M1
Collected	Venezuela	Zúlia	10.787401	-71.709799	Feral	1	1	H15	R8	M1
The Antilles										
INIFAP	Cuba	Pinar del Río	22.970207	-82.385867	Cultivated	1	1	H1	R20	M1
AGROSAVIA	Puerto Rico	San Juan	18.38925	-66.05525	Cultivated	4	4	R1, R17	R1, R2, R16, R20	M1
Total samples						**243**	**235**			

*number of samples with successful amplification for at least one marker

Table S2. Plastidial sequences developed for *P. guajava*

Primer	Sequences
ycf4-cemA_F	TCTTGCGCGTACCGATTGAA
ycf4-cemA_R	TGAACCCACCTCATTTGTTGA
ndhf-rpl32_F	TCTTTCTAATTCACCGGCTCTT
ndhf-rpl32_R	ATCTCTGTCTATTTTCGTCGTGC
petA-psbj_F	CGTGTCCAAGGCCTTTTGTT
petA-psbj_R	TGGTTCATACTCCGGGTTGG

Table S3. Results of independent hierarchical analysis of molecular variance (AMOVA) based on cpDNA, ITS, and MeNu39 sequences for *P. guajava* populations from Mesoamerica, the Antilles, and South America.

	df	Sum of Squares	Estimated variance	%	Fixation indices
cpDNA					
No groups defined					
Among populations	12	53.251	0.2184	15.6	FST = 0.15***
Within populations	208	245.894	1.1821	84.4	
Total	220	299.145	1.4006		
Three groups					
Between groups	2	26.672	0.1631	11.17	FCT = 0.11
Among pop. Within groups	10	26.579	0.1154	7.9	FSC = 0.08**
Within populations	208	245.894	1.1821	80.93	FST = 0.19***
Total	220	299.145	1.1821		
ITS					
No groups defined					
Among populations	12	227.371	0.5668	24.03	FST = 0.24***
Within populations	441	790.134	1.7916	75.97	
Total	453	1017.504	2.3585		
Three groups					
Between groups	2	74.899	0.0591	2.49	FCT = 0.02
Among pop. Within groups	10	152.472	0.5289	22.23	FSC = 0.22***
Within populations	441	790.134	1.7916	75.29	FST = 0.24***
Total	453	1017.504	2.3797		
MeNu39					
No groups defined					
Among populations	12	25.179	0.0586	20.03	FST = 0.20***
Within populations	459	107.467	0.2341	79.97	
Total	471	132.646	0.2927		
Three groups					
Between groups	2	12.139	0.0308	10.14	FCT = 0.10**
Among pop. Within groups	10	13.04	0.0392	12.91	FSC = 0.14***

Within populations	459	107.467	0.2341	76.95	FST = 0.23***
Total	471	132.646	0.3042		

**P < 0.05; ***P < 0.0001

CAPÍTULO III

GENETIC DIFFERENTIATION OF GUAVA ACROSS THE AMERICAS

E Arévalo-Marín, A. Casas, H. Alvarado-Sizzo, E. Ruiz-Sanchez, G. Castellanos-Morales, G. Fermin, J. S. Padilla-Ramírez, C. R. Clement

GENETIC DIFFERENTIATION OF GUAVA ACROSS THE AMERICAS

Edna Arévalo-Marín[1,2*], Alejandro Casas[1*], Hernán Alvarado-Sizzo[3], Eduardo Ruiz-Sanchez[4], Gabriela Castellanos-Morales[5], Gustavo. Fermin[6], José Saúl Padilla-Ramírez[7], Charles R. Clement[8]

[1]Instituto de Investigaciones en Ecosistemas y Sustentabilidad - IIES, Universidad Nacional Autónoma de México, Morelia, Michoacán, México

[2]Posgrado en Ciencias Biológicas, Unidad de Posgrado, Edificio D, 1º Piso, Circuito de Posgrado, Ciudad Universitaria, Coyoacán, CDMX, México

[3]Laboratorio de Biogeografía y Sistemática, Departamento de Biología Evolutiva, Facultad de Ciencias, Universidad Nacional Autónoma de México, Ciudad de México, México

[4]Departamento de Botánica y Zoología, Centro Universitario de Ciencias Biológicas y Agropecuarias, Universidad de Guadalajara, Zapopan, Jalisco, México

[5]Departamento de Conservación de la Biodiversidad, El Colegio de la Frontera del Sur, Unidad Villahermosa (ECOSUR-Villahermosa), Villahermosa, Tabasco México

[6]Instituto Jardín Botánico de Mérida, Facultad de Ciencias, Universidad de Los Andes. Mérida, Mérida, Venezuela

[7]Campo Experimental Pabellón, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP), Pabellón de Arteaga, Aguascalientes, México

[8]Coordenação de Tecnologia e Inovação, Instituto Nacional de Pesquisas da Amazônia, Manaus, Amazonas, Brasil

[*]**Corresponding Authors**

Edna Arévalo-Marín earevalo@cieco.unam.mx

Alejandro Casas acasas@cieco.unam.mx

Abstract

Guava (*Psidium guajava* L.) is a Neotropical semi-domesticated fruit species of medium importance in the Americas and especially adapted to anthropized areas. Its main importance is the fresh consumption of the fruit and the elaboration of different products. It is source of micronutrient and medicinal benefits. To characterize the genetic structure of guava in almost its entire geographical distribution, we studied 215 individuals from 11 countries in Mesoamerican, Andean, and Amazonian regions by using 25 nuclear microsatellite loci developed for *P. guajava*. Analyses of genetic provided evidence that Amazon guavas from Brazil and Peru, and the Peruvian Andes formed specific clusters. In contrast, guava samples from Colombia and Venezuela overlap with those from the Antilles. Similarly, we defined a Central American cluster and a Mexican cluster. These differentiation patterns were consistent in all tests used. We observed intermediate levels of genetic differentiation (F_{ST}=0.2079). The pairwise F_{ST} values suggested moderate to high genetic differentiation between guava populations, with the Peruvian Andes (PER-AND), the Peruvian Amazon (PER-AM), and the Brazilian Amazon (BRA-AM) showing greater degree of differentiation when compared with the rest of the samples. We found excess of homozygosity within populations ($H_E = 0.78 > Ho = 0.30$), which could be due to inbreeding. There is high genetic diversity in Mesoamerican (Mexico and Central America) populations, with moderate genetic structure due to possible historical events, and gene flow events with northern South American populations. Finally, we tested several diffusion scenarios with coalescent analyses, and the scenario with highest probability suggest that guava could have spread from the Brazilian Amazon to the Peruvian and from there to the northern of South America, Central America, and Mexico.

Keywords: crop evolution, domestication, genetic resources conservation; microsatellites

1. Introduction

Guava (*Psidium guajava* L.) is a Neotropical semi-domesticated fruit species of medium importance in the Americas (Altendorf, 2018, 2019). It is distributed from Mexico and the Antilles to northwestern Argentina (Landrum, 2017). The guava fruit is the most frequently used part of the plant. It is consumed fresh or used to make candies, dried fruits, jams, jellies, juices, pastes, soup bases, and syrup (Landrum, 2017, 2021). It is a good source of calcium, iron, niacin, pantothenic acid, phosphorus, riboflavin, and thiamine (Hiwale, 2015). In folk medicine, the decoction of roots, bark, and leaves is used to treat respiratory discomfort and help to expel the placenta after childbirth (Gutiérrez et al., 2008). The aqueous extracts of guava leaves are used to treat gastrointestinal problems such as vomiting, diarrhea, and gastric pain and to reduce blood glucose levels in diabetic patients (Gutiérrez et al., 2008; Hiwale, 2015). Guava grows in dry tropical forests and savannah-like vegetation. It also grows in disturbed areas (roadsides and grasslands), small agroecological environments (homegardens and orchards), and larger-scale production systems (Landrum, 2017, 2021). It adapts easily to different rainfall conditions and soil types. However, it does not tolerate flooded soils and is sensitive to low temperatures (Menzel, 1985). Guava cultivation has spread outside its natural distribution to countries such as India, Pakistan, China, and Egypt (Altendorf, 2018, 2019), thus reflecting the current global importance of this species.

During post-Columbian times, guava was the fruit tree more widely recorded by European chroniclers of the 16th century, documenting its presence in Mesoamerica and South America in both wild and cultivated populations. There are also additional reports of its occurrence in the Greater and Lesser Antilles (Arévalo-Marín et al., 2021; Patiño, 2002). The European conquerors learned to use fruits and leaves of the guava as medicine and food (Patiño, 1963), uses that have prevailed among indigenous peoples until now. The oldest archaeological macro remains records place guava in pre-Columbian contexts in Southwestern Amazonia (date between 9490 and 6505 years calibrated before present [cal. BP]) and the human settlements of the Peruvian coast (7.000 cal. BP). In Mexico, the earliest macro remains found were fruit fragments much more recent, dating to ca. 670 cal. BP (references in Arévalo-Marín et al., 2021).

Despite its economic, cultural, and historical importance, and even when efforts have increased to study genetic diversity and understand the history under management and domestication of various Neotropical fruit trees, guava has received little attention from genetic studies (Arévalo-Marín et al., 2021). This may be due in part to the fact that under domestication, perennial trees are often propagated both asexually (by clonal or grafting techniques) and sexually (Diez et al., 2015; McClure et al., 2014; Miller & Gross, 2011). The result is a reduction in sexual reproduction that separates domesticated individuals from their wild progenitors (Gaut et al., 2015) and therefore, slow rates of evolution. These characteristics often make it hard to study the evolutionary history of the domestication of tree crops (Miller & Gross, 2011). In addition, currently, guava wild populations are unknown, being found mainly as tolerated or feral individuals that with the passage of time may form small populations (personal observations) or in cultivated files.

Therefore, it is necessary to expand studies on this fruit species to obtain a clear image of their structure and genetic diversity, which will contribute to the knowledge about the domestication process and the recent impact of human practices on this important plant. Thus, our objective was to characterize the genetic guava structure in its almost whole geographical distribution based on SSR markers. The following questions were posed: (a) what is the level of guava genetic variability among the sampled genotypes? (b) How is this diversity structured? (c) Is there isolation by distance between populations? (d) What is the history of diffusion of this species?

2. Material and Methods

2.1 Study area

We studied 215 individuals from 11 countries. We collected 86 guava samples from Brazil, Colombia, Honduras, Mexico, and Venezuela in the guava germplasm bank of the Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP) in Aguas Calientes, Mexico; 17 individuals from the Guava collection of the Tropical Agricultural Research and Higher Education Center (CATIE) in Turrialba, Costa Rica from Costa Rica, El Salvador, Guatemala, and Honduras; and 26 individuals from Brazil, Colombia, and Puerto Rico in the guava collection of the Corporación Colombiana de

Investigación Agropecuaria (Agrosavia) in Palmira, Colombia. We also included samples collected outside germplasm banks from the Brazilian Amazon (38 individuals), Peruvian Amazon (15 individuals), Peruvian Andes (19 individuals), and 14 samples from different localities in Venezuela. We considered samples collected in the Peruvian Andes and the Brazilian Amazon as tolerated or planted individuals because these were collected in orchards and gardens. Samples from Venezuela and the Peruvian Amazon were collected in areas far from plantations or crops; however, since it is difficult to distinguish between wild and feral guavas, each of these samples were considered as feral accessions.

2.2 Molecular methods

DNA was extracted from young leaves using a modified CTAB protocol (Doyle & Doyle, 1987). Initially, all 215 individuals were genotyped using 25 nuclear microsatellite loci developed for *P. guajava* (Guavamap, 2008; Risterucci et al., 2005) We pooled together five primers in each of the five multiplex reactions (see Table S1 for primers and multiplex reactions details). PCRs were performed using the Platinum Multiplex PCR Master Mix (Thermo-Fisher, USA) following the manufacturer's instructions for reaction assembly and program. Every reaction was driven to a 5.5 µL final volume containing 2.0 µL Platinum Multiplex PCR Master Mix, 2 µL PCR grade H_2O, 0.5 µL G/C enhancer volume, 1.0 µL DNA template (50–200 ng/µL), and primer concentrations between 50 to 70 nM according to each product RFU. Multiplex reactions required an annealing temperature of 55°C for all primers; 40 cycles were used in every PCR reaction. When amplification was not successful, we repeated the PCR reactions using 0.04 µL Kapa polymerase (Kapa Taq HotStart), 2.0 µL Buffer Kapa, 2.0 µL PCR grade H2O, and 1.0 µL DNA template (50–200 ng/µL). The annealing temperature and the number of cycles were maintained. Additional cycling conditions were implemented following manufacturer directions. To control for possible contamination, we used negative controls for each reaction. All products were verified in 2% agarose gels electrophoresis. PCRs were carried out in a MultiGene OptiMax (Labnet International, Inc., Edison, NJ, USA) or in a 2700 thermal cycler (Applied Biosystems, Foster City, CA, USA). Genotyping was achieved using the Microsatellite plugin (v. 1.4.7) of Geneious Prime 2022 (Dotmatics,

NZ). Allele scoring was performed manually using as guidelines the supplementary of Selkoe and Toonen (2006)

2.3 Null alleles, Hardy-Weinberg equilibrium, and linkage disequilibrium tests

We tested the presence and frequency of null alleles following the method of Brookfield (1996) using the *PopGenReport* v.3.0.7 package (Adamack & Gruber, 2014) in R. We calculated deviations from Hardy-Weinberg equilibrium (HWE) for each locus and separately, for each population. Also, we calculated HWE across all samples using the 'hw.test' function of the R package *pegas* v.1.1 (Paradis, 2010), with 1,000 Monte Carlo permutations. Alpha levels to determine statistical significance for Hardy–Weinberg proportions and independent sorting of genotypes were adjusted using the false discovery rate (FDR) approach developed by Benjamini & Hochberg (1995), using 0.05 alpha level. P-values were corrected for multiple comparisons using the Benjamini-Hochberg method (Benjamini & Hochberg, 1995). We calculated a measure of correlation (rbarD; Agapow & Burt, 2001) using the function 'ia' (Kamvar et al., 2014) in the R package *poppr* v.2.9.3, for testing overall linkage disequilibrium. Using the function 'genotype curve' of the same package, we described the genotypic diversity in relation to different combinations of loci by a genotype accumulation curve. The curve was generated by sampling x loci randomly and counting the number of multilocus genotypes (MLG) observed. This sampling was repeated r times from 1 to n-1 loci, creating n-1 distributions of observed MLGs.

2.4 Genetic diversity and genetic differentiation

Genetic differentiation was examined using several complementary approaches. First, as an exploratory method, we performed a Principal Components Analysis (PCA) to summarize the genetic variation based on the microsatellite data set. Subsequently, we performed a Discriminant Analysis of Principal Components (DAPC; Jombart et al., 2010). DAPC is an approach that optimizes the separation of individuals into predefined groups using a discriminant function of the principal component (Jombart et al., 2010). Based on DAPC, the membership probability was calculated for the overall genetic background of an individual. We used the components identified in the PCA analysis as

predefined groups for the DAPC implementation. For implementing the PCA, we used the 'dudi.pca' function from *ade4* v.1.7-22 R package (Dray & Dufour, 2007) and visualized it with the *factoextra* v.1.0.7 R package (Wickham, 2016). For DAPC, we used the *adegenet* v.2.2.10 package (Jombart, 2008) implemented in R.

We assessed standard measures of genetic diversity for the entire dataset and genetics groups according to DAPC results. The number of individuals (*N*), number of alleles (*A*), and the expected (*H$_E$*) and observed (*H$_O$*) heterozygosities were calculated using *poppr* v.2.9.3 package (Kamvar et al., 2014) in R. We estimated rarefied allelic richness using the 'allel.rich' function of the *PopGenReport* v.3.0.7 package (Adamack & Gruber, 2014) in R. Private allele richness (*AP*) were calculated using a rarefaction approach (Hurlbert, 1971; Kalinowski, 2004) implemented in the ADZE 1.0 software (Szpiech et al., 2008).

As an additional test to calculate the population assignment probability for each individual, we performed genetic structure analysis using the Bayesian approach implemented in STRUCTURE 2.3.4 (Falush et al., 2003; Pritchard et al., 2000), based on an admixture model with correlated allele frequencies and information on the origin of populations (popinfo = 1). The admixture model was tested for K-values ranging from 1 to 8, with 10 independent runs per K value for the entire dataset. We used 1 000,000 Markov Chain Monte Carlo iterations with a burn-in length of 100,000. To determine the most probable value of K, we used the Evanno ΔK method (Evanno et al., 2005) and mean LnP(K) (Janes et al., 2017) implemented in Structure Harvester v0.6.94 (Earl & VonHoldt, 2012). We used CLUMPP 1.1.2 (Jakobsson & Rosenberg, 2007) program with the Greedy algorithm to infer the optimal K-cluster affiliations of individuals. StructuRly v.0.1.0 package (Criscuolo & Angelini, 2020), in R, was used to generate bar graphs of STRUCTURE results.

Wright's F statistics (Wright, 1949; *F$_{IS}$*, *F$_{IT}$*, and *F$_{ST}$*) were estimated using the methods of Weir & Cockerham, (1984). We also calculated the genetic differentiation among populations through a pairwise *F$_{ST}$* matrix. Both the F statistics and the paired *F$_{ST}$* matrix were calculated with values with 95% confidence intervals from 10,000 bootstrapping, using the 'diffCalc' function of the *diveRsity* package (Keenan et al., 2013). A Mantel test (Mantel, 1967) was used to test for isolation by distance (IBD)

between guava populations. We used the geographic distance matrix transform from coordinates in Euclidean distance and calculated using the function 'dist' in the *stats* v.4.3.1 package in R and a linearized pairwise F_{ST} matrix (F_{ST} /1 - F_{ST}) as genetic distance. The function 'mantel.rtest' from *ade4* v.1.7-22 package was used to calculated the Mantel test, and scatter-plots were then generated with the R package *adegenet* v.2.2.10 *package*.

We also tested the degree of genetic differentiation between clusters (determined here) and locations, performing the analysis of molecular variance (AMOVA) followed by estimating the extent of genetic differentiation with phi-statistics, both using the 'poppr.amova' function in the *poppr* v.2.9.3 package (Kamvar et al., 2014). The significance of variance components was assessed using a permutation test implemented through the 'randtest' function in the *ape4* v.5.7 package (Bougeard & Dray, 2018; Chessel et al., 2004; Dray et al., 2007; Dray & Dufour, 2007) with 999 permutations.

2.5 Demographic history of guava

We used nuclear microsatellite data to run the Approximate Bayesian Computation (ABC) framework (Beaumont, 2010; Beaumont et al., 2002) implemented in DIYABC-RF GUI (Collin et al., 2021). We considered five scenarios (1) Mexico as a probable domestication area with dissemination to South America; (2) South America (Brazil-AM) as an probable area of domestication and a posterior dissemination to Mexico via Peru; (3) Two independent events of domestication (Peru-An and Mexico); (4) Peru-AM and Brazil-AM as independents areas of domestication and dissemination towards northern South America with Central America and Mexico being of admixed origin; and (5) domestication in northern South America and dissemination to three areas (Mexico and Central America, Venezuela and Antilles, and Peru and Brazil) (Supplementary Figure 1). The priors and conditions for each parameter can be found in Supplementary File 2; we considered a generation time of 10 years (probable fruiting time in natural conditions) (Crane & Balerdi, 2005). We conducted previous runs to adjust the tested scenarios and the parameters (Bertorelle et al., 2010). For the final run, we obtained 500,000 simulated datasets, 500 trees, and 424 summary statistics. To

identify the best supported scenario, we performed model check based on 500 pseudo-observed data sets (PODs) under each scenario to assess confidence in scenario choice, and to estimate the class specific error rates which is the mean classification error rate (Collin et al., 2021; Cornuet et al., 2010; Robert et al., 2011).

3. Results

3.1 Null alleles, Hardy-Weinberg equilibrium, and linkage disequilibrium tests

We removed 18 accessions and discarded the locus mPgCIR08 with more than 40% of missing data. We found no evidence of null alleles in our data set. All loci showed significant deviations from Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) in more than one population (Table S2-S3). Because the loci in HWE and LD were not the same for all populations, markers were retained for further analyses. The genotype accumulation curve (Fig 1) shows that the set of loci teste is sufficient to accurately resolve the different haplotypes in our sample as the curve revealed that more than 100% of the genotypes could be detected with 8 markers, hence the loci accurately estimated the diversity of our sample (Fig 1).
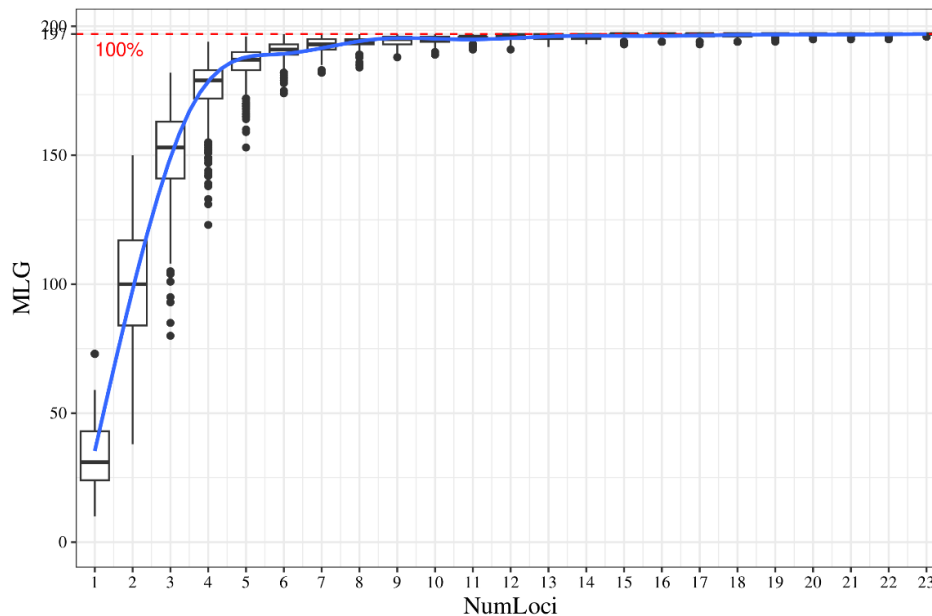

Fig 1. Genotype accumulation curve for 197 samples of *P. guajava* over 24 loci. The horizontal axis represents the number of loci randomly sampled without replacement up to n-1

loci, and the vertical axis shows the number of unique multilocus genotypes observed in the data set. The level of 100% of the unique multilocus genotype detected is indicated with a dotted red line.

3.2 Genetic diversity and genetic differentiation

We analyzed a total of 188 individuals of guava. The PCA provided evidence of genetic structure of guava across its geographical range. The first two principal components explained 18.5% of the total variation (Fig 2). Amazonian guavas from Brazil and Peru (BRA-AM and PER-AM, respectively) formed well-defined clusters. In contrast, guava samples from Colombia (COL) and Venezuela (VEN) overlap with those from the Antilles (ANT), documenting the close relationship between these regions. Given that the centroids of the Colombian and Venezuelan clusters do not co-occur within their respective standard deviation ellipses, we decided to define the Colombian (COL) and Venezuela-Antilles (VEN-ANT) clusters separately. Similarly, we defined a Peruvian Andes (PER-AND) cluster, Central American (CenAme) cluster and a Mexican (MEX) cluster. For subsequent analyses, we decided to discard the accessions from southern Brazil (BRA-SP), because the origin of these samples is uncertain (Fig 2).
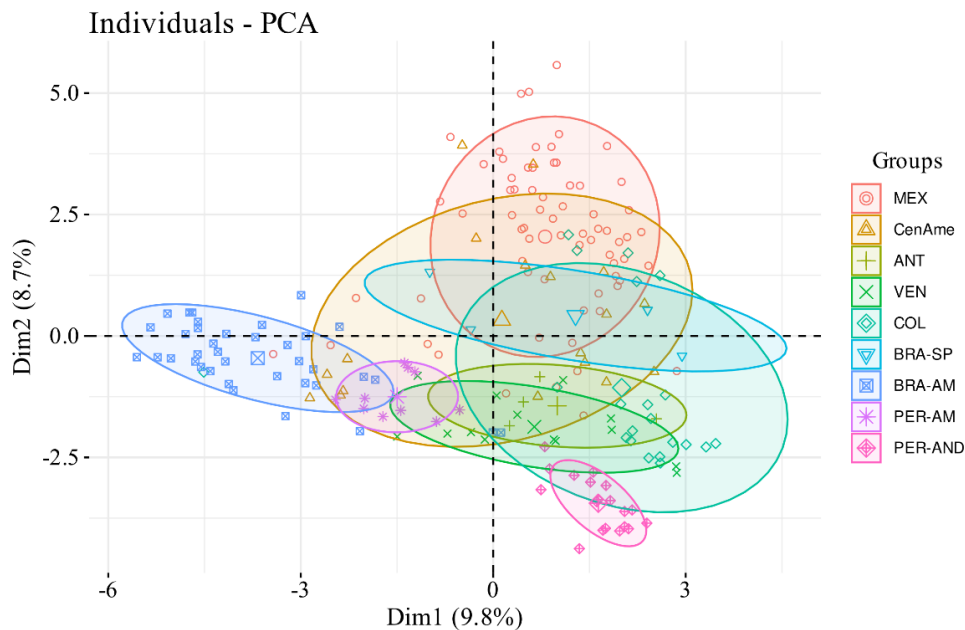
Fig 2. Principal component analysis (PCA) of microsatellite genotype data from *P. guajava* individuals showing the clustering along principal component axis 1-2. Populations: MEX

(Mexico), CenAme (Central America), ANT (The Antilles), VEN (Venezuela), COL (Colombia), BRA-SP (Sao Paulo, Brazil), BRA-AM (Brazilian Amazon), PER-AM (Peruvian Amazon), PER-AND (Peruvian Andes).

The DAPC analysis with 7 predefined groups from the results of the PCA revealed a clear differentiation of guavas from the Brazilian and Peruvian Amazonia and the Peruvian Andes (Fig 3a). These results were consistent with the patterns obtained using PCA. Likewise, the Mesoamerican (MEX+CenAme) and the northern South American clusters (COL+VEN-ANT) appeared as an admixed population (Fig 3a). We performed a second DPAC analysis excluding the Peruvian and Brazilian well-differentiated groups to depict the relationships between Mesoamerican and northern South American populations. This analysis showed that VEN-ANT cluster was well differentiated from COL, CenAme and MEX, which are more closely related (Fig 3b).

Fig 3. Discriminant analysis of principal components (DAPC) a) for 8 guava populations, and b) for only 4 guava populations. Populations ID described in PCA Fig 2.

113

The samples from The Antilles were not considered for diversity analyses because of their low sample numbers. The Mexican (MEX) and Central American (CenAme) samples showed highest H_E values (0.70 ± 0.022 and 0.70 ± 0.027, respectively). While the lowest H_E value was found in the Peruvian Andes and Peruvian Amazon (Table 1). Total expected heterozygosity was 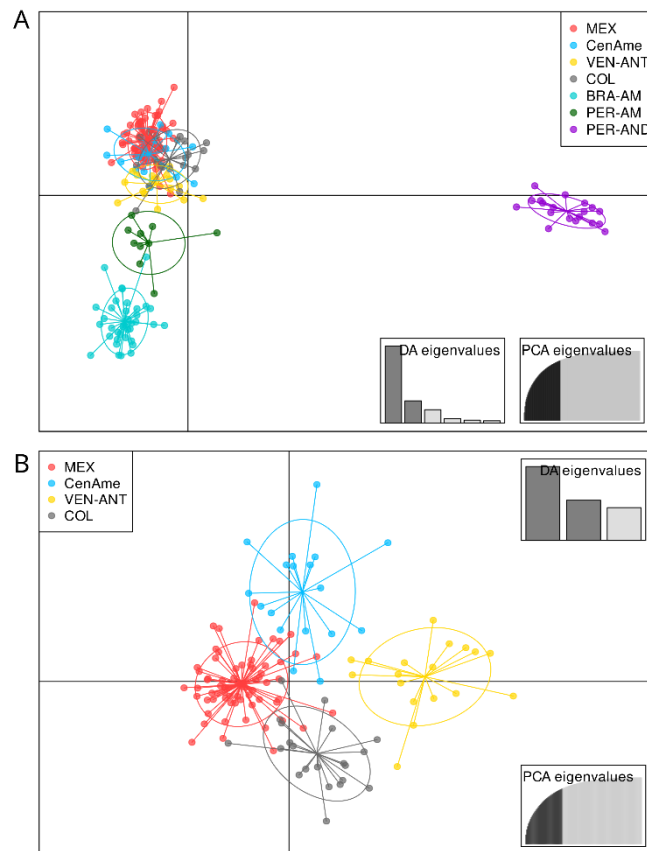0.78 ± 0.020. When considering all the samples, total expected heterozygosity (H_T) was 0.79 ± 0.021. The averaged observed heterozygosity (H_O) was 0.30 and it ranged from 0.21 ± 0.023 in COL to 0.44 ± 0.062 in PER-AMA. Total observed heterozygosity was 0.30 ± 0.015. Inbreeding coefficient (F) averaged 0.50 and ranged from 0.06 to 0.61 in PER-AMA and COL, respectively. The overall F was 0.46 ±0.025.

Table 1. Genetic diversity obtained for 24 nuclear microsatellite loci for 188 accessions of *P. guajava*

	N	A	A_r	P_{Ar}	H_O	H_E	uH_E	F
Species	192	351	14.25	5.14	0.30 ±0.015	0.78 ± 0.020	0.79 ± 0.021	0.46 ±0.025
DAPC								
MEX	69	227	1.70	0.51	0.34 ± 0.024	0.70 ± 0.022	0.70 ± 0.022	0.52 ±0.028
CenAme	18	152	1.71	0.44	0.34 ± 0.027	0.70 ± 0.026	0.71 ± 0.027	0.50 ±0.043
VEN	18	151	1.67	0.44	0.26 ± 0.026	0.65 ± 0.046	0.68 ± 0.047	0.50 ±0.052
COL	21	142	1.60	0.35	0.21 ± 0.023	0.60 ± 0.041	0.61 ± 0.042	0.61 ±0.044
BRA-AMA	37	158	1.61	0.59	0.25 ± 0.026	0.60 ± 0.046	0.61 ± 0.046	0.53 ±0.044
PER-AMA	10	83	1.50	0.62	0.44 ± 0.062	0.47 ± 0.047	0.53 ± 0.055	0.06 ±0.093
PER-AND	19	81	1.45	0.56	0.22± 0.042	0.44 ± 0.048	0.45 ± 0.050	0.50 ±0.071

Number of individuals (N), number of alleles (A), rarefied allelic richness (Ar), number of private alleles P_{Ar}), observed heterozygosity (Ho), expected heterozygosity (He), unbiased heterozygosity (uH_E) and fixation index (F)

The results of the Structure analysis were consistent with the results from PCA and DAPC. Evanno and Jane's methods indicated an optimal value of K = 3 and K = 5 as the most likely numbers of genetic clusters (Fig S1). For K =3, the samples from the Peruvian Andes and the Brazilian Amazon show well-defined clusters (cluster 3 and 2, respectively; Figure 3). The admixture of these two genetic groups forms the rest of the South American samples and The Antilles, with the cluster associated with the Peruvian

114

Andes being more relevant (cluster 3). A Mesoamerican cluster (MEX and CenAme) is also identified (cluster 1). This cluster is also found in the Colombian (COL) and Antillean samples, although in smaller proportions than the South American genetic groups (Fig 4).

The K = 4 value shows clusters defined for Colombia (cluster 1) and Mexico (cluster 3). The individuals from Venezuela (VEN) and the Antilles (ANT) showed a mixture mainly with the Peruvian Andes (PER-AM) and Amazonia (BRA-AM), and Colombia (COL) clusters. Central American (CenAme) samples are admixed with all genetic groups (Fig.4).

For K = 5, we found five geographically well-defined genetic groups: cluster 1 (purple) contains individuals from the Peruvian Amazon (PER-AM); cluster 2 (red), individuals from the Brazilian Amazon (BRA-AM); cluster 3 (yellow), individuals from the Peruvian Andean (PER-AND); cluster 4 (blue), individuals from Colombia; and cluster 5 (turquoise), individuals from Mexico (MEX). Within South America, most individuals from Venezuela (VEN), and Antilles (ANT) show admixture with the Peruvian Amazon (PER-AM) and the Colombia (COL) clusters. Nearly 30% of the Mexican individual trees show some level of admixture with the South American clusters (Fig 4). The Mexican cluster (MEX) is represented in some individuals of the CenAme and COL clusters. Within South America, most individuals from VEN, and ANT are likely outcome of admixture between the PER-AM cluster and the COL cluster. About 30% of the Mexican guavas studied show some level of admixture with the South American genetic groups (Fig 4). Also, the MEX genetic group (cluster 5) is represented in CenAme and some individuals of the COL cluster. The PER-AM cluster is highly represented in some CenAme individuals, although most of the samples from this region are admixed individuals formed by the COL and MEX clusters.
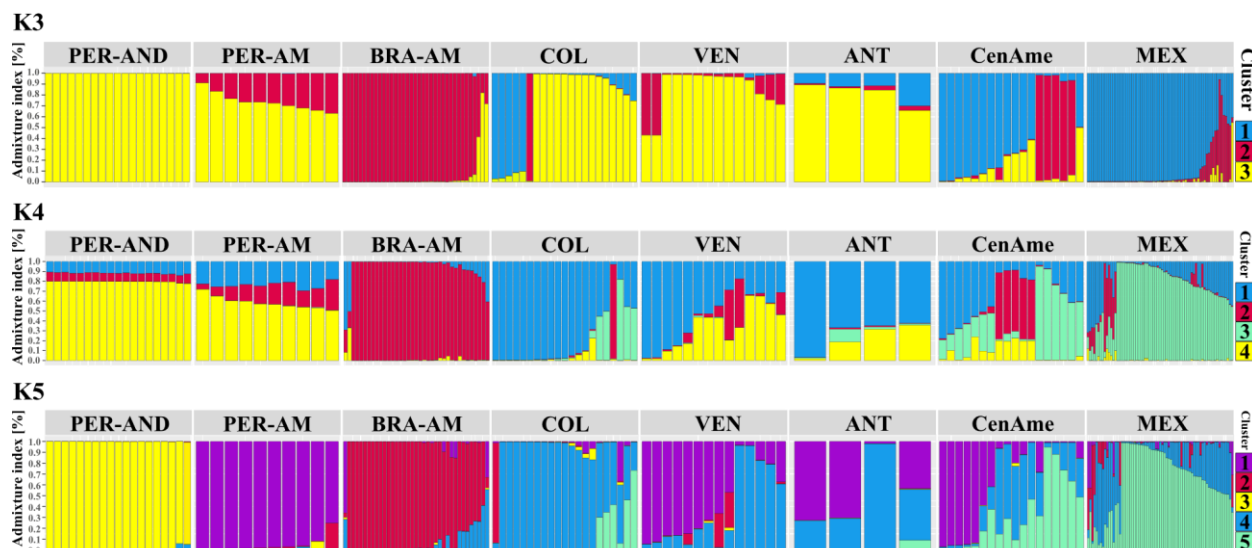
Fig 4. Assignment probabilities of each of the 192 guava samples analyzed to each cluster inferred by STRUCTURE for K = 3, 4 and 5. Each accession is represented by a vertical bar, and color indicates the probability of belonging to each cluster. Accessions are ordered according to the geographic region from southern to northern parts of the Americas.

Estimates of Wright's F (Wright, 1949) among the sampling localities and regions indicated that guava diversity is more concentrated within populations than among populations; nevertheless, we found intermediate levels of genetic differentiation with F_{ST}=0.2079 (Table 2). F_{IT} and F_{IS} estimates were higher compared to F_{ST} (Table 2). All fixation indexes were statistically significant. These results suggest that the frequency of heterozygotes is lower than expected under HWE (Table 1).

Table 2. Results of Wright's F-statistics estimates obtained for 188 accessions of *P. guajava* using 24 specific microsatellite loci.

DAPC Groups	F_{IS}	F_{ST}	F_{IT}
All loci	0.5433	0.2079	0.6383
Upper (CI95%)	0.5844	0.2319	0.6702
Lower (CI95%)	0.5013	0.1859	0.6055

CI95% = 95% confidence interval.

The pairwise F$_{ST}$ values suggested moderate to high genetic differentiation between populations (Fig 6). The largest difference was observed for the Peruvian Andes (PER-AND), Peruvian Amazon (PER-AM), and Brazilian Amazon (BRA-AM). Samples from Colombia (COL), Venezuela (VEN), Central America (CenAme), and Mexico (MEX). Mesoamerica (MEX and CenAme) and northern South America (COL and VEN-ANT) showed less pairwise genetic differentiation (Fig 5).
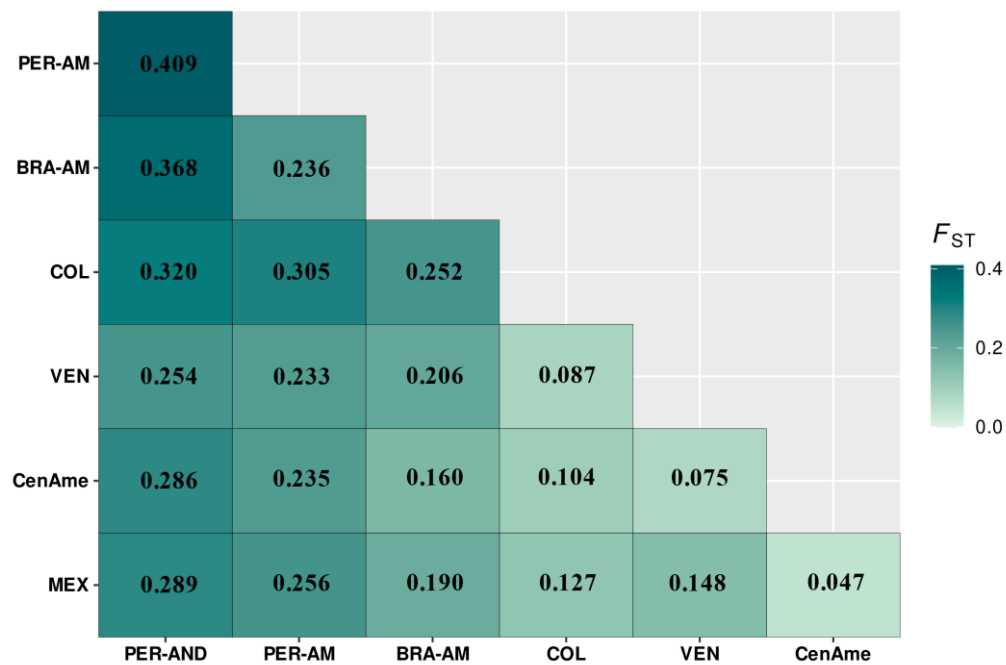
Fig 5. F_{ST} genetic differentiation values among the 192 accessions of guava grouped by country.

The Mantel test revealed no significant correlation between genetic and geographic matrices (R^2= -0.2870773, p= 0.779), indicating a lack of isolation by distance. However, we found that PER-AND, PER-AM, and BRA-AM populations, which are geographically closer, are genetically less similar (Figure 7; Table S5). Pairwise F_{ST} could indicate possible long-distance gene flow between MEX, CenAme, VEN, and COL populations (Table Sx4-5).

According to AMOVA the variation between samples within countries (41%, Φ = 0.52) is higher than between countries (21,04%, Φ = 0.62; Table 3).
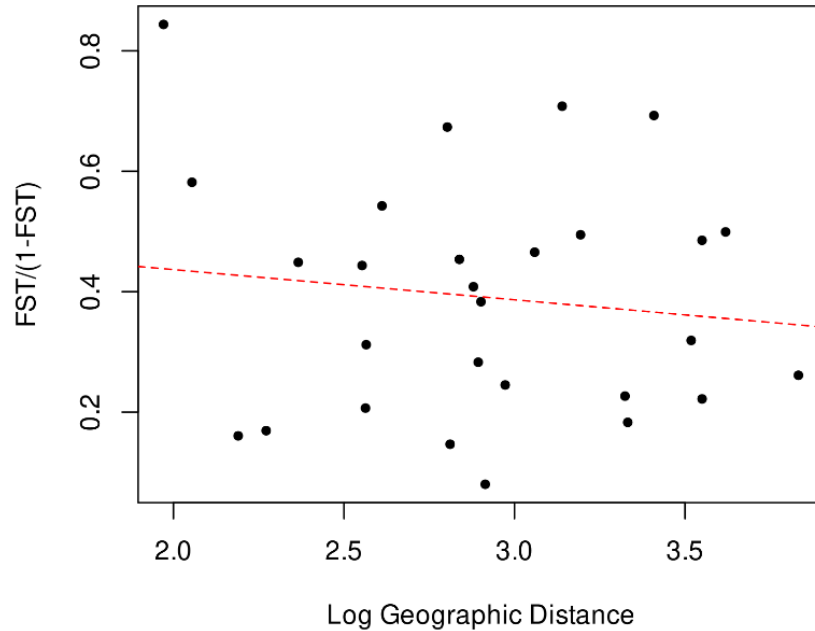
Fig 6. Test of isolation by distance (IBD) obtained through a Mantel test. The result was non-significant

Table 3. Results of the analysis of molecular variance (AMOVA) testing for differentiation between populations in *P. guajava*

	df	**Sum of Squares**	**% variance**	**Φ-statistic**	***p*-value**
Between countries	6	1235.802	21.04	0.627	0.001
Between samples Within countries	1811	3884.253	41.64	0.527	0.001
Within samples	188	1248.343	37.32	0.210	0.001
Total	375	6368.399	100		

Approximate Bayesian Analyses Computation (ABC) analyses, implemented in DIYABC-RF indicated that the supported diffusion hypothesis was scenario 2 (Figure 7) with posterior probability of 0.999 and, non-overlapping confidence intervals. This scenario showed low type I and type II error rates (0.00013; Table 4). This scenario considers that the domestication could have started in South America, specifically in the

Brazilian Amazon (Brazil-AM), with dissemination to Mexico via the Peruvian Amazon (PER-AMA) and northern of South America (COL and VEN).

Scenario 2

- N1 = MEX
- N2 = CENAME
- N3 = ANT
- N4 = VEN
- N5 = COL
- N6 = BRA-AM
- N7 = PER-AM
- N8 = PER-AND

t5
t4
t3
t2
t1
0

N8 N6 N7 N3 N4 N5 N1 N2
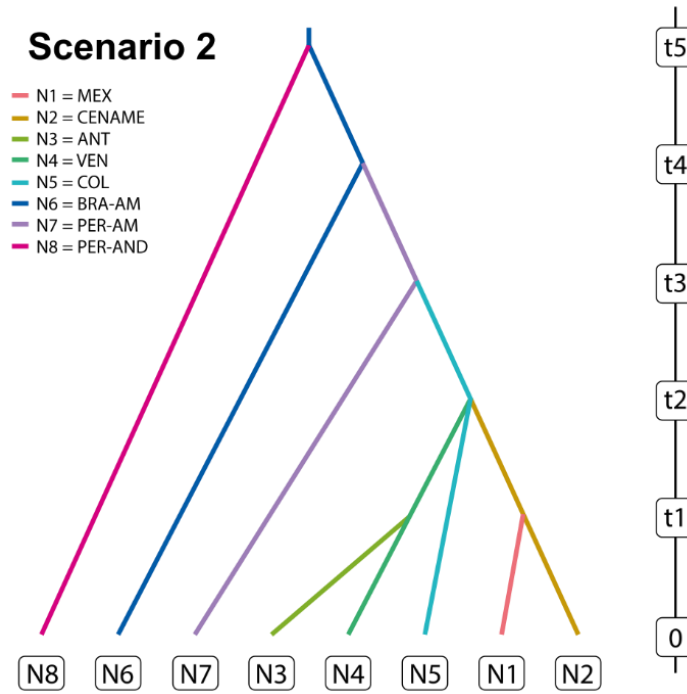
Fig 7. Highest-probability scenario tested for diffusion of *Psidum guajava* in The Americas.

Table 4. Class specific prediction errors for 5 scenarios tested with DIYABC-RF. Model performance indicated by type I (rows) and type II (columns) errors in percentage of times each predicted scenario was predicted

	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Class error
Predicted 1	99.96	0.00	0.04	0.00	0.00	0.00035
Predicted 2	0.00	99.99	0.00	0.00	0.00	7.97×10^{-05}
Predicted 3	0.00	0.00	100.00	0.00	0.00	9.97×10^{-06}
Predicted 4	0.00	0.01	0.00	99.98	0.01	0.00021
Predicted 5	0.00	0.00	0.00	0.00	100.00	0

4. Discussion

4.1 Genetic diversity and genetic differentiation

Perennial trees are typically expected to exhibit weak population structure due to several factors, including their extended lifespan with delayed sexual reproduction, outcrossing, hybridization, clonal propagation, gene flow between wild and cultivated populations, and dispersal mechanisms, among others (Hamrick & Godt, 1990; Loveless & Hamrick,

1984)**.** Values of genetic diversity expressed as He ranged from 0.44-0.70. These levels are comparable to those of other perennials such as *Annona*, *Olea*, and *Prunus* (Bourguiba et al., 2020; Breton et al., 2006; Larranaga et al., 2017). However, compared to *Psidium* species, *P. guava* holds high genetic diversity, for example, in a single population in South-East Brazil the maximum H_E values (H_E = 0.71) are comparable to those of *P. guineense* (*He* = 0.74) and *P. macahense* (*He* = 0.63) (de Oliveira Bernardes et al., 2022). An insular *Psidium* species (*P. galapageium*) has moderate to low **(**H_E = 0.275-0.570) genetic diversity (Urquía et al., 2019). *P. cattleianum* also showed lower diversity values (H_E = 0.117-0.326) (Machado et al., 2021). Other studies have demonstrated that large population sizes tend to increase genetic diversity (Cao et al., 2014; Leimu et al., 2006). Therefore, the high levels of genetic diversity found in guava could be due to the larger population sizes of this species than other species of the genus.

Observed heterozygosity (*Ho*) systematically showed lower values than expected heterozygosity, suggesting heterozygous deficiency due to generalized inbreeding. In different islands of the Galapagos (Urquía et al., 2019) and guava samples from germplasm bank (Sitther et al., 2014) have registered similar results. These findings may be explained, in part, by the self-fertilization and vegetative propagation, which can occur in *P. guajava* (Crane & Balerdi, 2005**)** and long-longevity in perennial trees as well as in their clones (Diez et al., 2015). Likewise, the Robertson's (1964) hypothesis, could be useful to explain the heterozygosity decline in guava. He proposes that subdividing a population into several isolated groups would allow maximum genetic diversity (minimum global co-ancestry) to be achieved in the long term since different allelic variants will develop and become fixed in each group, becoming a genetic reservoir of variation. However, complete isolation leads to higher rates of local inbreeding with the possible consequence of inbreeding depression. Therefore, he also suggests that occasional mixing of these subpopulations would minimize the overall rate of inbreeding. In support of this hypothesis, we found lower rates of global inbreeding for guava. Likewise, Mantel's analysis suggests long-distance flow, especially between the populations of northern South America (COL and VEN, Table S4-5) and Mesoamerica (MEX and CenAme, Table S4-5), which would allow the reduction of inbreeding and its

effects. However, further studies are needed to examine the cause of the heterozygote deficiency in guava.

On the other hand, contrary to our south-to-north decrease in diversity hypothesis, the genetic diversity pattern, expressed as H_E is just the opposite. A decreasing trend in H_E was observed from the North in Mexico and Central America (H_E = 0.70) to the South in the Peruvian Andes (PER-AND = 0.44). This pattern can be explained by higher degrees of the mixture of guavas from different regions occurring in Central America and Mexico. In these areas, most of the individuals show signatures of a mixture of well-defined South American genetic groups. The emergence of the land bridge of the Isthmus of Panama, which allowed the connection of the biota of North and South America (Bartoli et al., 2005; Landrum, 2017; O'Dea et al., 2016), and the anthropic dispersion may have favored the increasing of guava genetic diversity in Central America and Mexico. Besides, the diversity of environmental conditions, new biotic interactions, and selection pressures in Mesoamerica could have contributed to the maintenance of genetic variants that were present in the gene pool possibly due to mutations that occurred after the Panama Isthmus closed. These events would explain an increased guava genetic diversity in response to new environmental conditions and challenges, a hypothesis that is testable by using niche analyses. In addition, whether this pattern points out towards a center of genetic diversity, or it is the result of admixture among clusters is a matter to be evaluated by rating explicit demographic scenarios. The effect of human activities on the genetic diversity of Mesoamerican populations must be supported by robust archaeological evidence. Therefore, archaeobotanical studies on the presence and use of guava in archaeological sites in Mesoamerica, and paleogenomics research on macro and micro remains found in Mesoamerica and South America, are desirable.

In our study, the genetic differentiation of *Psidium guajava* populations yielded an F_{ST} value of 0.207, indicating a moderate differentiation, considering the wide geographical range across which the species is distributed. Likewise, the molecular variance is higher between (%) individuals/within populations than between populations (%). Similar findings have been reported for other perennials trees like *Annona*, *Diospyros*, *Juglans*, *Mangifera*, *Olea*, and *Prunus* (Bourguiba et al., 2020; Breton et al.,

2006; Cao et al., 2014; Larranaga et al., 2017; Pollegioni et al., 2015; Warschefsky & von Wettberg, 2019; Xu et al., 2021). In the case of guava, the observed F_{ST} value may be attributed to limited gene flow among the sampled populations, which span different regions of the Neotropics. Similarly, the pattern of variance identified here can be due to outcrossing and its invasive (successional) character (Hamrick et al., 1992 and therein). In cultivated and invasive populations of guava, the genetic variation pattern is also similar, with higher genetic variance among individuals/between populations and clearly defined genetic groups (Fagundes et al., 2016; Kumar et al., 2020; Urquía et al., 2019).

Regarding the genetic clustering found in our study, the most geographically isolated populations from South America (Peruvian Andes [PER- AND], and Brazilian and Peruvian Amazon [BRA-AM; PER-AMA]) belong a distinct genetic group each and show greater differentiation between them (F_{ST}; Figure 6) in relation to other groups. Populations in northern South America, Central America, and Mesoamerica show lower values genetic differentiation with some individuals being admixed. This scenario suggests a pattern of North-South differentiation (greater differentiation in the South), given the isolation between guava populations, and those in the Amazonian and Andean regions. Additionally, the populations evaluated here would have likely been exposed to specific evolutionary processes, considering their geographical origin's climatic and ecological characteristics, thus promoting a differentiation between them. Guava needs open areas and exposure to direct sunlight to ensure its good growth, development, and fruitfulness in temperatures between 20 and 30°C (Paull & Duarte, 2012; Sugahara & Takaki, 2004). The Peruvian Andes and the Amazon of Peru and Brazil probably have more rigorous environmental conditions limiting the establishment of guava populations. On the other hand, the pattern of genetic clustering found in guava, with well-defined genetic groups across Central, Meso, and South America as well as the Antilles, is consistent with people using a subset of the genetic pool from geographically near individuals. Variable admixture levels among populations may also be the outcome of diverse trade routes and human migrations over time (Arévalo-Marín et al., 2021), as is the case of *Juglans* (Pollegioni et al., 2015) and Mangifera (Warschefsky & von Wettberg, 2019).

According to the best-supporting demographic scenario, the first guava diffusion route likely was through the Brazilian Amazon toward the Peruvian Andes. Although our sampling does not include individuals or populations from the southwestern Amazon, and therefore our results do not directly support the hypothesis of this place as a probable area of guava domestication (Arévalo-Marín et al., 2021). This result draws attention given that the oldest archaeological guava macro remains were found in the Southwestern Amazonia, in the Teotonio archeological site in a layer dated between 9490 and 6505 cal. BP (Watling et al., 2018). In South America, the lowlands of southwestern Amazonia are recognized as an important center of domestication (Clement, 1999; Clement et al., 2021) and the place from where crops as important as manioc and peanuts were dispersed towards the Peruvian dry coast (Piperno, 2011). Indeed, a significant number of archaeological guava remains dating from 6975 to 450 BP have been reported in the Peruvian dry coast (see Figure 4 in Arévalo-Marín et al., 2021). Therefore, there is also evidence that guava could also have spread through the Andes, from the Amazon to the Peruvian coast as the best supported scenario suggest.

Conclusion

This study allowed obtained an overview of the genetic structure of guava fruit tree in the Americas. The microsatellite markers and Bayesian clustering approaches allowed recognizing the presence of one gene pool in Mesoamerican (Mexico) and three in South America (Brazilian Amazon, Peruvian Amazon and Andes, and Colombia). The high genetic differentiation between the Brazilian and Peruvian Amazon and Peruvian Andes guava samples could be due to environmental differences, since guava subpopulations in distinct geographic settings may reflect divergent local adaptation. Likewise, this population split could explain high values of inbreeding found in guava. Niche analyses are needed to understand whether climatic events could explain these hypotheses, and genomic analyses would allow testing for hypotheses of local adaptation. On the other hand, the ABC approach identified the Brazilian Amazon to be the potential area of the domestication for guava, with subsequent dissemination to northern South America and Mesoamerica through the Andes. Follow-up studies that include defined populations of feral and cultivated guavas, and focused sampling in the

northeastern Amazon and the Andes could help to unravel the guava domestication process.

Acknowledgments

References

Adamack, A. T., & Gruber, B. (2014). PopGenReport: simplifying basic population genetic analyses in R. *Methods in Ecology and Evolution*, *5*(4), 384–387.

Agapow, P., & Burt, A. (2001). Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes*, *1*(1-2), 101–102.

Altendorf, S. (2018). Minor tropical fruits: Mainstreaming a niche market. In *Food Outlook Biannual Report on Global Food Markets* (Issue July, pp. 67–74). http://www.fao.org/3/a-I8080e.pdf

Altendorf, S. (2019). *Major tropical fruits market review 2018*. http://www.fao.org/3/ca5692en/CA5692EN.pdf

Arévalo-Marín, E., Casas, A., Landrum, L., Shock, M. P., Alvarado-Sizzo, H., Ruiz-Sanchez, E., & Clement, C. R. (2021). The Taming of *Psidium guajava*: Natural and Cultural History of a Neotropical Fruit. *Frontiers in Plant Science*, *12*(September), 1–15. https://doi.org/10.3389/fpls.2021.714763

Bartoli, G., Sarnthein, M., Weinelt, M., Erlenkeuser, H., Garbe-Schönberg, D., & Lea, D. W. (2005). Final closure of Panama and the onset of northern hemisphere glaciation. *Earth and Planetary Science Letters*, *237*(1–2), 33–44.

Beaumont, M. A. (2010). Approximate Bayesian Computation in Evolution and Ecology. *Annual Review of Ecology, Evolution, and Systematics*, *41*(1), 379–406. https://doi.org/10.1146/annurev-ecolsys-102209-144621

Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian Computation in Population Genetics. *Genetics*, *162*(4), 2025–2035. https://doi.org/10.1093/genetics/162.4.2025

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300.

Bertorelle, G., Benazzo, A., & Mona, S. (2010). ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology*, *19*(13), 2609–2625.

Bougeard, S., & Dray, S. (2018). Supervised multiblock analysis in R with the ade4 package. *Journal of Statistical Software*, *86*, 1–17.

Bourguiba, H., Scotti, I., Sauvage, C., Zhebentyayeva, T., Ledbetter, C., Krška, B., Remay, A., D'onofrio, C., Iketani, H., & Christen, D. (2020). Genetic structure of a worldwide germplasm collection of *Prunus armeniaca* L. reveals three major diffusion routes for varieties coming from the species' center of origin. *Frontiers in Plant Science*, *11*, 638.

Breton, C., Tersac, M., & Bervillé, A. (2006). Genetic diversity and gene flow between the wild olive (oleaster, *Olea europaea* L.) and the olive: several Plio-Pleistocene refuge zones in the Mediterranean basin suggested by simple sequence repeats analysis. *Journal of Biogeography*, *33*(11), 1916–1928.

Cao, K., Zheng, Z., Wang, L., Liu, X., Zhu, G., Fang, W., Cheng, S., Zeng, P., Chen, C., & Wang, X. (2014). Comparative population genomics reveals the domestication history of the peach, Prunus persica, and human influences on perennial fruit crops. *Genome Biology*, *15*, 1–15.

Chessel, D., Dufour, A. B., & Thioulouse, J. (2004). The *ade4* package-I-One-table methods. *R News*, *4*(1), 5–10.

Clement, C. R. (1999). 1492 and the loss of amazonian crop genetic resources. II. Crop biogeography at contact. *Economic Botany*, *53*(2), 203–216. https://doi.org/10.1007/bf02866499

Clement, C. R., Casas, A., Parra-Rondinel, F. A., Levis, C., Peroni, N., Hanazaki, N., Cortés-Zárraga, L., Rangel-Landa, S., Alves, R. P., Ferreira, M. J., Cassino, M. F., Coelho, S. D., Cruz-Soriano, A., Pancorbo-Olivera, M., Blancas, J., Martínez-Ballesté, A., Lemes, G., Lotero-Velásquez, E., Bertin, V. M., & Mazzochini, G. G. (2021). Disentangling Domestication from Food Production Systems in the Neotropics. *Quaternary*, *4*(1), 4. https://doi.org/10.3390/quat4010004

Collin, F., Durif, G., Raynal, L., Lombaert, E., Gautier, M., Vitalis, R., Marin, J., & Estoup, A. (2021). Extending approximate Bayesian computation with supervised machine learning to infer demographic history from genetic polymorphisms using DIYABC Random Forest. *Molecular Ecology Resources*, *21*(8), 2598–2613.

Cornuet, J.-M., Ravigné, V., & Estoup, A. (2010). Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1. 0). *BMC Bioinformatics*, *11*(1), 1–11.

Crane, J. H., & Balerdi, C. F. (2005). Guava growing in the Florida home landscape. *Horticultural Sciences Department Document HS4, Florida Cooperative Extension Service Institute of Food and Agricultural Sciences, University of Florida*, 1–8.

Criscuolo, N. G., & Angelini, C. (2020). StructuRly: A novel shiny app to produce comprehensive, detailed and interactive plots for population genetic analysis. *Plos One*, *15*(2), e0229330.

de Oliveira Bernardes, C., Tuler, A. C., Canal, D., Carvalho, M. S., Ferreira, A., & da Silva Ferreira, M. F. (2022). Genetic Diversity and Population Structure of Psidium Species from Restinga: A Coastal and Disturbed Ecosystem of the Brazilian Atlantic Forest. *Biochemical Genetics*, *60*(6), 2503–2514.

Diez, C. M., Trujillo, I., Martinez-Urdiroz, N., Barranco, D., Rallo, L., Marfil, P., & Gaut, B. S. (2015). Olive domestication and diversification in the Mediterranean Basin. *New Phytologist*, *206*(1), 436–447. https://doi.org/10.1111/nph.13181

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, *19*, 11–15.

Dray, S., Dufour, A. B., & Chessel, D. (2007). The *ade4* package-II: Two-table and K-table methods. *R News*, *7*(2), 47–52.

Dray, S., & Dufour, A.-B. (2007). The *ade4* package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, *22*, 1–20.

Earl, D. A., & VonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*, 359–361.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol*, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Fagundes, B. S., da Silva, L. F., Giacomin, R. M., Secco, D., Díaz-Cruz, J. A., & Da-Silva, P. R. (2016). Transferability of microsatellite markers among Myrtaceae species and their use to obtain population genetics data to help the conservation of the Brazilian Atlantic Forest. *Tropical Conservation Science*, *9*(1), 408–422.

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, *164*(4), 1567–1587.

Gaut, B. S., Díez, C. M., & Morrell, P. L. (2015). Genomics and the Contrasting Dynamics of Annual and Perennial Domestication. *Trends in Genetics*, *31*(12), 709–719. https://doi.org/10.1016/j.tig.2015.10.002

Guavamap. (2008). *Improvement of Guava: Linkage mapping and QTL analysis as a basisfor marker-assisted selection. Available*.

Gutiérrez, R. M., Mitchell, S., & Solis, R. V. (2008). *Psidium guajava*: a review of its traditional uses, phytochemistry and pharmacology. *J Ethnopharmacol*, *117*(1), 1–27. https://doi.org/10.1016/j.jep.2008.01.025

Hamrick, J. L., & Godt, M. J. W. (1990). Allozyme diversity in plant species. *Plant Population Genetics, Breeding, and Genetic Resources.*, 43–63.

Hamrick, J. L., Godt, M. J. W., & Sherman-Broyles, S. L. (1992). Factors influencing levels of genetic diversity in woody plant species. *Population Genetics of Forest Trees: Proceedings of the International Symposium on Population Genetics of Forest Trees Corvallis, Oregon, USA, July 31–August 2, 1990*, 95–124.

Hiwale, S. (2015). Guava (Psidium guajava). In *Sustainable Horticulture in Semiarid Dry Lands* (pp. 213–224). Springer India. https://doi.org/10.1007/978-81-322-2244-6

Hurlbert, S. H. (1971). The nonconcept of species diversity: a critique and alternative parameters. *Ecology*, *52*(4), 577–586.

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, *23*(14), 1801–1806.

Janes, J. K., Miller, J. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., Cullingham, C. I., & Andrew, R. L. (2017). The K= 2 conundrum. *Molecular Ecology*, *26*(14), 3594–3602.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405.

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 1–15.

Kalinowski, S. T. (2004). Counting alleles with rarefaction: private alleles and hierarchical sampling designs. *Conservation Genetics*, *5*, 539–543.

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). *Poppr*: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281.

Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., & Prodöhl, P. A. (2013). diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, *4*(8), 782–788.

Kumar, C., Kumar, R., Singh, S. K., Goswami, A. K., Nagaraja, A., Paliwal, R., & Singh, R. (2020). Development of novel g-SSR markers in guava *(Psidium guajava* L.) cv. Allahabad Safeda and their application in genetic diversity, population structure and cross species transferability studies. *PLoS One*, *15*(8), e0237538.

Landrum, L. R. (2017). The genus *Psidium* (Myrtaceae) in the state of Bahia, Brazil. *Canotia*, *13*, 1–101.

Landrum, L. R. (2021). *Psidium guajava* L.: taxonomy, relatives and possible origin. In S. Mitra (Ed.), *Guava: botany, production and uses* (pp. 1–21). Cab International. https://doi.org/10.1079/9781789247022.0001

Larranaga, N., Albertazzi, F. J., Fontecha, G., Palmieri, M., Rainer, H., van Zonneveld, M., & Hormaza, J. I. (2017). A Mesoamerican origin of cherimoya (*Annona cherimola* Mill.): Implications for the conservation of plant genetic resources. *Molecular Ecology*, *26*(16), 4116–4130. https://doi.org/10.1111/mec.14157

Leimu, R., Mutikainen, P. I. A., Koricheva, J., & Fischer, M. (2006). How general are positive relationships between plant population size, fitness and genetic variation? *Journal of Ecology*, *94*(5), 942–952.

Loveless, M. D., & Hamrick, J. L. (1984). Ecological determinants of genetic structure in plant populations. *Annual Review of Ecology and Systematics*, *15*(1), 65–95.

Machado, R. M., de Oliveira, F. A., de Matos Alves, F., de Souza, A. P., & Forni-Martins, E. R. (2021). Population genetics of polyploid complex *Psidium cattleyanum* Sabine (Myrtaceae): preliminary analyses based on new species-specific microsatellite loci and extension to other species of the genus. *Biochemical Genetics*, *59*(1), 219–234.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, *27*(2_Part_1), 209–220.

McClure, K. A., Sawler, J., Gardner, K. M., Money, D., & Myles, S. (2014). Genomics: A potential panacea for the perennial problem. *American Journal of Botany*, *101*(10), 1780–1790. https://doi.org/10.3732/ajb.1400143

Menzel, C. M. (1985). Guava: an exotic fruit with potential in Queensland. *Queensland Agricultural Journal*, *111*(2), 93–98.

Miller, A. J., & Gross, B. L. (2011). From forest to field: perennial fruit crop domestication. *American Journal of Botany*, *98*(9), 1389–1414.

O'Dea, A., Aguilera, O., Aubry, M. P., Berggren, W. A., Budd, A. F., Cione, A. L., Coates, A. G., Collins, L. S., Coppard, S. E., & Cozzuol, M. A. (2016). *Building Bridges. Response to Erkens and Hoorn:"The Panama Isthmus,'old','young'or both?".*

Paradis, E. (2010). pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics*, *26*(3), 419–420.

Patiño, V. M. (1963). *Plantas cultivadas y animales domésticos en América Equinoccial I: Frutales.* CO: Impre. Departamental.

Patiño, V. M. (2002). *Historia y dispersión de los frutales nativos del neotrópico.* Centro Internacional de Agricultura Tropical. https://books.google.com.mx/books?id=U12m7M2VMmsC

Paull, R. E., & Duarte, O. (2012). Guava. In O. Paull, R. E., Duarte (Ed.), *Tropical fruits, Volume 2* (pp. 91–122). CABI International.

Piperno, D. R. (2011). The Origins of Plant Cultivation and Domestication in the New World Tropics. *Current Anthropology*, *52*(S4), S453–S470. https://doi.org/10.1086/659998

Pollegioni, P., Woeste, K. E., Chiocchini, F., Del Lungo, S., Olimpieri, I., Tortolano, V., Clark, J., Hemery, G. E., Mapelli, S., & Malvolti, M. E. (2015). Ancient humans influenced the current spatial genetic structure of common walnut populations in Asia. *PloS One*, *10*(9), e0135980.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.

Risterucci, A. M., Duval, M. F., Rohde, W., & Billotte, N. (2005). Isolation and characterization of microsatellite loci from *Psidium guajava* L. A. *Molecular Ecology Notes*, *5*(4), 745–748. https://doi.org/10.1111/j.1471-8286.2005.01050.x

Robert, C. P., Cornuet, J.-M., Marin, J.-M., & Pillai, N. S. (2011). Lack of confidence in approximate Bayesian computation model choice. *Proceedings of the National Academy of Sciences*, *108*(37), 15112–15117.

Robertson, A. (1964). The effect of non-random mating within inbred lines on the rate of inbreeding. *Genetics Research*, *5*(1), 164–167.

Selkoe, K. A., & Toonen, R. J. (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, *9*(5), 615–629. https://doi.org/10.1111/j.1461-0248.2006.00889.x

Sitther, V., Zhang, D., Harris, D. L., Yadav, A. K., Zee, F. T., Meinhardt, L. W., & Dhekney, S. A. (2014). Genetic characterization of guava (*Psidium guajava* L.) germplasm in the United States using microsatellite markers. *Genetic Resources and Crop Evolution*, *61*(4), 829–839. https://doi.org/10.1007/s10722-014-0078-5

Sugahara, V. Y., & Takaki, M. (2004). Effect of light and temperature on seed germination in guava (Psidium guajava L. - Myrtaceae). *Seed Science and Technology*, *32*, 759–764. https://doi.org/10.15258/sst.2008.36.3.30

Szpiech, Z. A., Jakobsson, M., & Rosenberg, N. A. (2008). ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, *24*(21), 2498–2504.

Urquía, D., Gutierrez, B., Pozo, G., Pozo, M. J., Espín, A., & Torres, M. de L. (2019). *Psidium guajava* in the Galapagos Islands: population genetics and history of an invasive species. *PloS One*, *14*(3), e0203737. https://doi.org/10.1371/journal.pone.0203737

Warschefsky, E. J., & von Wettberg, E. J. B. (2019). Population genomic analysis of mango (*Mangifera indica*) suggests a complex history of domestication. *New Phytologist*, *222*(4), 2023–2037. https://doi.org/10.1111/nph.15731

Watling, J., Shock, M. P., Mongeló, G. Z., Almeida, F. O., Kater, T., De Oliveira, P. E., & Neves, E. G. (2018). Direct archaeological evidence for Southwestern Amazonia as an early plant domestication and food production centre. *Plos One*, *13*(7), e0199868. https://doi.org/10.1371/journal.pone.0199868

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 1358–1370.

Wickham, H. (2016). Data analysis. *Ggplot2: Elegant Graphics for Data Analysis*, 189–201.

Wright, S. (1949). The genetical structure of populations. *Annals of Eugenics*, *15*(1), 323–354.

Xu, Y., Cheng, W., Xiong, C., Jiang, X., Wu, K., & Gong, B. (2021). Genetic diversity and association analysis among germplasms of Diospyros kaki in Zhejiang Province based on SSR markers. *Forests*, *12*(4), 422.

SUPPORTING INFORMATION

GENETIC DIFFERENTIATION OF GUAVA ACROSS THE AMERICAS

E. Arévalo-Marín, A. Casas, H. Alvarado-Sizzo, E. Ruiz-Sanchez, G. Castellanos-Morales, G. Fermin, J. S. Padilla-Ramírez, C.R. Clement

Table S1. Characteristics of the 25 microsatellite loci used in the analysis of genetic diversity of guava (*Psidium guajava*) in the Americas

SSR Locus	Repeat Motif	Primer sequences (5' to 3')		Fragment size range
		Forward	**Reverse**	
mPgCIR1 (B)	$(GA)_{17}$	TAGTGCTTTGGTTGCTT	GCAGGTGGATATAAGGTC	236–250
mPgCIR2 (C)	$(GA)_{20}$	AGTGAACGACTGAAGACC	ATTACACATTCAGCCACTT	202–230
mPgCIR4 (D)	$(GA)_{25}$	TTCAGGGTCTATGGCTAC	CAACAAGATACAGCGAACT	126–150
mPgCIR7 (A)	$(CA)_{13}AA(GAA)_3$	ATGGAGGTAGGTTGATG	CGTAGTAATCGAAGAAATG	148–160
mPgCIR 9 (B)	$(GA)_{19}$	GCGTGTCGTATTGTTTC	ATTTTCTTCTGCCTTGTC	156–176
mPgCIR10 (D)	$(CT)12$	GTTGGCTCTTATTTTGGT	GCCCCATATCTAGGAAG	262–320
mPgCIR11 (A)	$(CT)17$	TGAAAGACAACAAACGAG	TTACACCCACCTAAATAAGA	298–314
mPgCIR13 (E)	$AC)_{12}(AT)_4G(GA)_2$	CCTTTTTCCCGACCATTACA	TCGCACTGAGATTTTGTGCT	240–260
mPgCIR14 (B)	$(GA)_{11}$	TAAACACAACAAGGGTCA	CAGTTTTCATATCGTCCTC	184–186
mPgCIR15 (C)	$(GA)_8GG(GA)_9$	TCTAATCCCCTGAGTTTC	CCGATCATCTCTTTCTTT	144–172
mPgCIR17 (D)	$(CT)_{23}$	CCTTTCGTCATATTCACTT	CATTGGATGGTTGACAT	230–240
mPgCIR19 (E)	$(CT)_{16}$	AAAATCCTGAAGACGAAC	TATCAGAGGCTTGCATTA	258–280
mPgCIR20 (C)	$(CT)_{14}(CA)_{17}$	TATACCACACGCTGAAAC	TTCCCCATAAACATCTCT	270–298
mPgCIR21 (E)	$(AG)_{15}GG(AG)_7$	TGCCCTTCTAAGTATAACAG	AGCTACAAACCTTCCTAAA	150–164
mPgCIR22 (C)	$(GT)_9(GA)_{14}$	CATAAGGACATTTGAGGAA	AATAAGAAAGCGAGCAGA	236–252
mPgCIR25 (A)	$(GA)_{24}$	GACAATCCAATCTCACTTT	TGTGTCAAGCATACCTTC	104–130
mPgCIR26 (D)	$(GT)_2(GA)_{17}$	CTACCAAGGAGATAGCAAG	GAAATGGAGACTTTGGAG	180–198
mPgCIR101 (E)	$(GT)_{10}$	ATGGCTGTAAGAAGCAAAAG	GAAGAAATGTAGGTGCGTTC	110
mPgCIR137 (A)	$(GA)_{18}$	GGGGAATGCAGAGATTGT	AGATGATGGTCTCGCTTTT	100
mPgCIR209 (C)	$(GA)_{15}$	CTAAAGCCACATCCAGCA	CTAACATTTGCCTTCTACAGC	139
mPgCIR237 (D)	$(GA)_{16}$	AGATTCCATCTGCGATTGT	GCGGATCAAAACCTAATCT	101

mPgCIR243 **(E)**	(GA)$_{29}$	ACAGCAGGACACAAAGGA CAACTTTGCTAGAGATGAAG	GCTCTGAGGTGGTTTTCAT	174
mPgCIR420 **(A)**	(CA)$_{15}$	C	ATGTAGTAATCGAAGAAATGGTT	239
mPgCIR 437 **(B)**	(AC)$_{10}$	ACAACAGTTCTGATCCCAAA	CTCGGAGACACAGAGGTCTA	153

Colors in the table indicate the fluorophore used: blue = 6-FAM, yellow = NED, red = PET, and gree = VIC. Letters in parentheses indicated the mix to which each microsatellite belongs.

Table S2. P-values of the Hardy–Weinberg equilibrium test with the FRD-based correction for each combination of sample and locus for country

Chisq	MEX	CenAme	VEN-ANT	COL	BRA-AM	PER-AM	PER-AND
mPgCIR137	**0.00**	**0.01**	**0.00**	**0.00**	**0.00**	0.13	**0.06**
mPgCIR25	**0.00**	0.18	**0.00**	**0.00**	**0.00**	0.30	**0.00**
mPgCIR7	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**
mPgCIR420	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**	**0.04**	**0.00**
mPgCIR11	**0.00**	**0.05**	**0.00**	**0.00**	**0.00**	**0.03**	0.98
mPgCIR 437	**0.00**	**0.00**	0.16	**0.00**	0.75	0.81	**0.00**
mPgCIR 9	**0.00**	**0.00**	0.20	**0.00**	**0.00**	0.66	**0.01**
mPgCIR14	**0.00**	0.26	0.96	1.00	**0.00**	0.91	1.00
mPgCIR1	**0.00**	**0.00**	0.36	**0.27**	**0.00**	0.38	0.48
mPgCIR209	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	1.00	0.10
mPgCIR15	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.97	**0.00**
mPgCIR2	**0.00**	**0.00**	1.00	**0.00**	0.97	0.94	**0.00**
mPgCIR22	**0.00**	0.11	**0.00**	**0.00**	**0.00**	0.18	**0.00**
mPgCIR20	**0.00**	**0.04**	**0.00**	0.45	**0.00**	**0.01**	0.46
mPgCIR237	**0.00**	0.20	**0.01**	**0.00**	**0.00**	0.66	**0.00**
mPgCIR4	**0.00**	**0.00**	**0.03**	**0.00**	0.74	0.89	0.30
mPgCIR26	**0.00**	0.48	**0.00**	**0.00**	**0.00**	0.96	**0.02**
mPgCIR17	**0.00**	**0.00**	**0.04**	**0.00**	**0.00**	0.21	**0.01**
mPgCIR10	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.88	**0.01**
mPgCIR101	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.20	0.96
mPgCIR21	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**	0.45	1.00
mPgCIR243	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.15	**0.00**

mPgCIR19	**0.00**	**0.00**	**0.00**	**0.03**	**0.00**	0.94	**0.00**
mPgCIR13	**0.00**	**0.00**	1.00	**0.00**	**0.00**	**0.80**	**0.00**

MC	MEX	CenAme	VEN-ANT	COL	BRA-AM	PER-AM	PER-AND
mPgCIR137	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.28	**0.06**
mPgCIR25	**0.00**	**0.02**	**0.00**	**0.00**	**0.00**	1.00	**0.00**
mPgCIR7	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**	**0.02**
mPgCIR420	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**	**0.00**	**0.00**
mPgCIR11	**0.00**	0.28	**0.00**	**0.00**	**0.00**	**0.01**	0.81
mPgCIR 437	**0.00**	**0.00**	**0.01**	**0.00**	**0,68**	1.00	**0.00**
mPgCIR 9	**0.00**	**0.00**	**0.02**	**0.01**	**0.00**	1.00	**0.00**
mPgCIR14	**0.00**	0.29	1.00	1.00	**0.00**	1.00	1.00
mPgCIR1	**0.00**	**0.00**	**0.02**	**0.08**	**0.00**	1.00	0.18
mPgCIR209	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	1.00	**0.01**
mPgCIR15	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	1.00	**0.00**
mPgCIR2	**0.00**	**0.00**	1.00	**0.00**	1.00	1.00	**0.00**
mPgCIR22	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.14	**0.06**
mPgCIR20	**0.00**	**0.00**	**0.00**	**0.00**	**0.02**	**0.06**	0.19
mPgCIR237	**0.00**	0.21	**0.00**	**0.00**	**0.00**	1.00	**0.00**
mPgCIR4	**0.00**	**0.00**	**0.08**	**0.00**	0.41	0.96	0.39
mPgCIR26	**0.00**	0.35	**0.00**	**0.00**	**0.00**	1.00	**0.02**
mPgCIR17	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.18	**0.00**
mPgCIR10	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.66	**0.04**
mPgCIR101	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.18	1.00
mPgCIR21	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.27	1.00
mPgCIR243	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.06**	**0.00**
mPgCIR19	**0.00**	**0.00**	**0.00**	**0.06**	**0.00**	1.00	**0.00**
mPgCIR13	**0.00**	**0.00**	1.00	**0.00**	**0.00**	0.62	**0.00**

Chisq: p-values calculated using the exact χ2-test, MC:p- values calculated using the Monte Carlo permutation test. P-values ≤ 0.05 are in bold, andcells with p-values <0.05 for both HWE tests are boxed.

Table S3. The proportions of loci that are out of HWE for each population and standardized index of association (LD)

	Hardy-Weinberg equilibrium				Linkage equilibrium	
	Chisq	**MC**	**Chisq.fdr**	**MC.fdr**	**rbarD**	**p.rD**
DAPC clusters						
MEX	1.00	1.00	1.00	1.00	0.08	0.001
CenAme	0.79	0.92	0.75	0.92	0.19	0.001
VEN	0.75	0.83	0.75	0.83	0.06	0.009
COL	0.87	0.92	0.87	0.87	0.10	0.001
BRA-AMA	0.87	0.87	0.87	0.87	0.06	0.001
PER-AMA	0.17	0.21	0.17	0.12	0.03	0.182
PER-AND	0.67	0.67	0.65	0.67	0.05	0.001

Chisq: values calculated using the exact χ2- test; *MC*: values calculated using the Monte Carlo permutation test; *.fdr indicates testing with Benjamini-Hochberg correction; significance of rbarD (p < 0.01)

Table S4. Geographic distances *Psidium guajava* populations

	MEX	CenAme	ANT	VEN	COL	BRA-AM	PER-AM
CenAme	18.433242						
ANT	33.717281	19.545013					
VEN	34.810855	16.629832	12.975978				
COL	27.995193	9.698104	18.057741	8.933477			
BRA-AM	46.173604	27.769857	24.376937	13.002596	18.205155		
PER-AM	34.805722	17.086784	23.098807	10.653623	7.801922	12.847361	
PER-AND	37.295852	21.309452	30.230732	17.806803	13.618934	16.49345	7.177258

Table S5. Genetic distances (F_{ST}) *Psidium guajava* populations

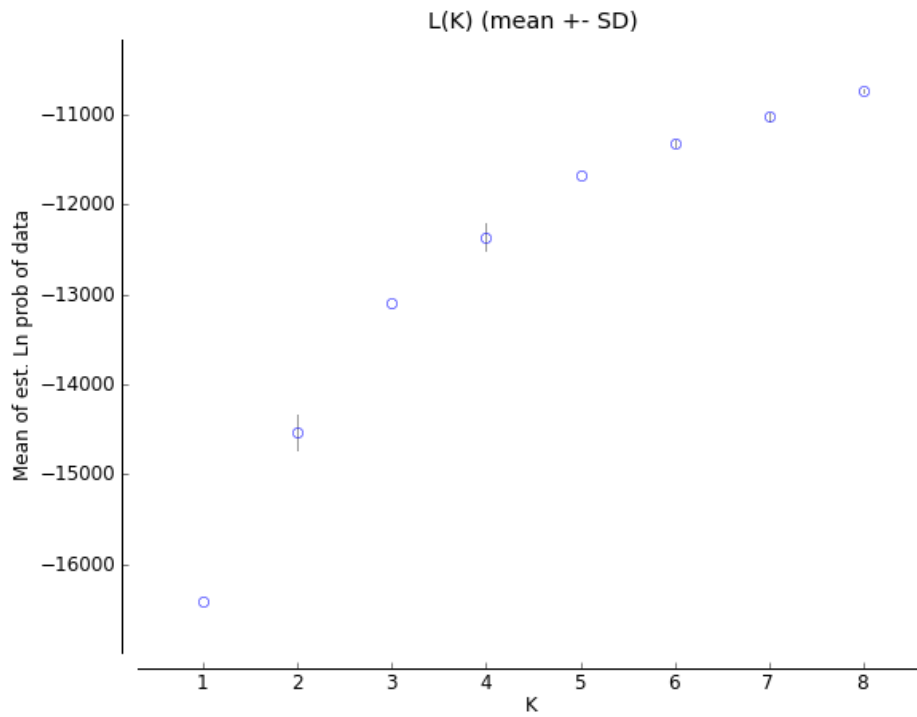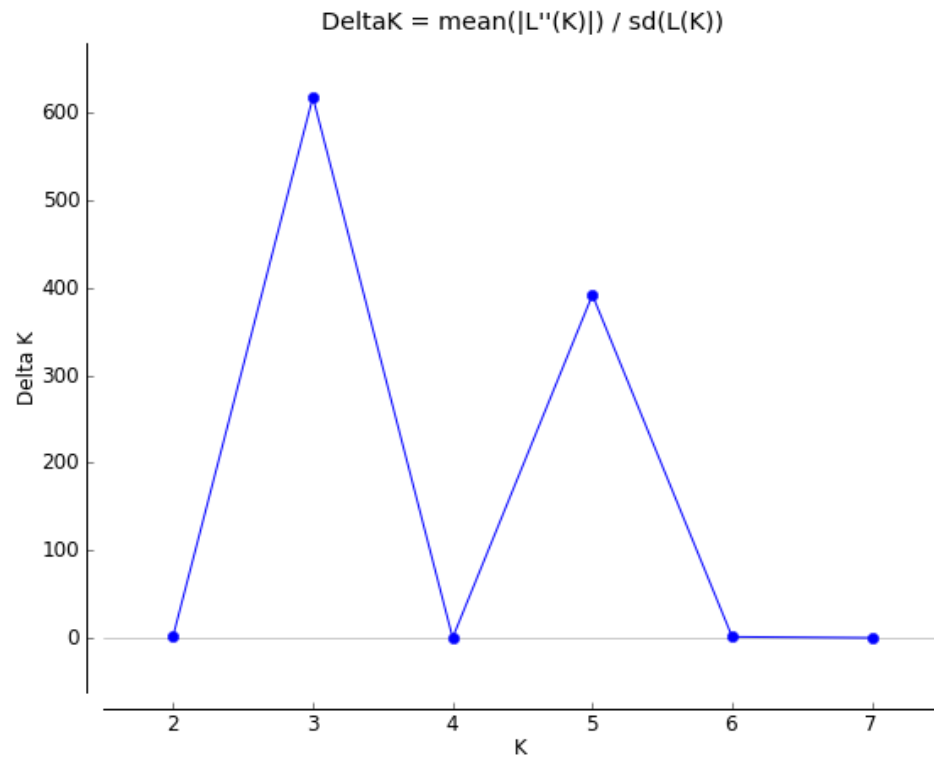|	MEX	CenAme	ANT	VEN	COL	BRA-AM	PER-AM
CenAme	2.914156						
ANT	3.51801	2.97272					
VEN	3.549929	2.811198	2.5631				
COL	3.332033	2.27193	2.893574	2.189806			
BRA-AM	3.832408	3.323951	3.193637	2.565149	2.901705		
PER-AM	3.549782	2.838305	3.139781	2.3659	2.05437	2.553138	
PER-AND	3.618882	3.059151	3.408859	2.879581	2.611461	2.802963	1.970917

Fig. S1. Graphical method allowing the detection of the number of groups K using **a)** ΔK, and **b)** LnP(K).
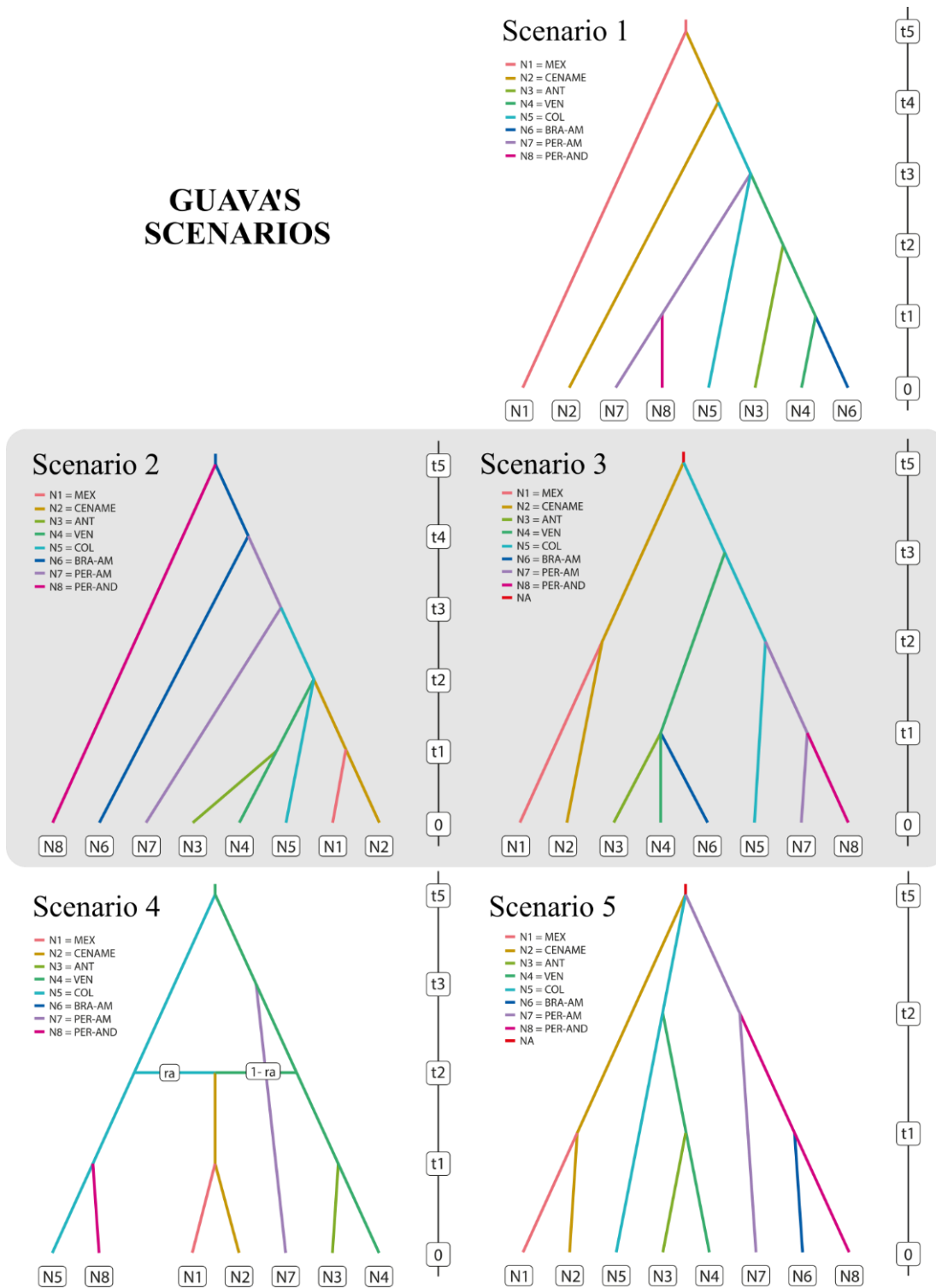
Fig. S2. Evolutionary scenarios under comparison using ABC analyses. **Scenario 1**: Mexico as a probable domestication; **scenario 2:** South America (Brazil-AM) as a probable area of domestication; **scenario 3:** Two independent events of domestication (Peru-An and Mexico); **scenario 4:** Peru-AM and Brazil-AM as independents areas of domestication; and **scenario 5**: domestication in northern South America.

CONCLUSIONES GENERALES

La guayaba es una planta ampliamente utilizada por sus frutos y múltiples beneficios alimenticios y medicinales (Gutiérrez et al., 2008) desde tiempos pre-colombinos (Aranguren et al., 2008; Patiño, 2002; Valdés-Infante et al., 2007). En la actualidad, ha adquirido importancia en algunos países fuera de su área de distribución natural (Kanupriya et al., 2011; Kumari et al., 2018; Mehmood et al., 2016; Sitther et al., 2014; Viji et al., 2010). Los estudios sobre esta especie se han enfocado principalmente al mejoramiento de sus cultivos (Briceno et al., 2010; Rodríguez et al., 2010; Rodríguez-Medina et al., 2010; Valdés-Infante et al., 2007, 2010), sobre sus propiedades químicas para el tratamiento de algunas enfermedades y sus beneficios nutrimentales (Gutiérrez et al., 2008). Pero su origen, su historia evolutiva y proceso de domesticación no han sido abordados.

Reconstruir la historia evolutiva de árboles frutales frutas neotropicales ha sido una tarea compleja porque estas especies perennes domesticadas y semi-domesticadas a menudo tienen orígenes difusos, y sus poblaciones cultivadas consisten en individuos derivados de múltiples áreas geográficas (Gaut et al., 2015; Miller & Gross, 2011). Por ello, la confianza en la prueba de hipótesis sobre los procesos que moldearon la distribución, el área más probable de origen y de domesticación, y las rutas de difusión y de un frutal semi domesticado requiere un respaldo riguroso de múltiples líneas de evidencia.

Para el caso de la guayaba, la información arqueobotánica revisada en este estudio permitió establecer una cronología sobre los restos arqueológicos hasta ahora reportados para esta especie en América (Arévalo-Marín et al., 2021). Esta fuente de información indica que la costa peruana presenta un gran número de reportes de macro restos fechados con una antigüedad de hace 7000 años calibrados antes del presente (cal. AP). Sin embargo, Watling et al. (2018) ubican semillas carbonizadas de *Psidium*, probablemente guayaba, en la Amazonía brasileña con una fecha entre 9490 y 6505 cal. AP. Restos de semillas también fueron encontrados en el norte de América del Sur, las Antillas y Mesoamérica (Belize), en estratos más recientes (2.600 cal. AP). En México, se han registrado solo 4 macro restos relativamente de 670 cal. AP.

Adicionalmente, hay registros de restos de polen encontrados en Uruguay y en México con fechas más antiguas (11.000 cal. AP), aunque su identidad no es suficientemente clara. Esta cronología muestra que la guayaba fue ampliamente utilizada por los pueblos precolombinos del continente (Arévalo-Marín et al., 2021). Pero también sugiere que los estudios arqueológicos llevados a cabo en Mesoamérica han sido insuficientes para reconstruir la historia que sugiere la evidencia genética reportada en este estudio.

Sumado a lo anterior, la caracterización biológica de la especie (hábitat, distribución y dispersores actuales, y otros aspectos ecológicos), permitieron obtener las primeras propuestas sobre el origen y la dispersión de la guayaba en el continente americano (Capítulo I). La síntesis presentada sugiere que sistematizar la información sobre la historia natural y cultural de una especie bajo domesticación es relevante para sustentar hipótesis valiosas para las investigaciones a realizar. Estas grandes fuentes de evidencia pueden tener efectos significativos en los patrones de distribución y variabilidad genética y esclarecerlo es un reto importante para la investigación sobre probables áreas de origen, de domesticación y difusión de las especies cultural y económicamente importantes.

Origen y dispersión de la guayaba

Los resultados obtenidos en este trabajo indican que la guayaba se habría originado durante el Mioceno (~13-8 Millones de años [Ma]). Así mismo, las muestras de *Psidium guajava* provenientes de 9 países (México, Costa Rica, Honduras, El Salvador, Puerto Rico, Cuba, Colombia, Perú y Brasil) mostraron estructura genética que revela dos linajes principales, un linaje suramericano más antiguo y diversificado y un linaje mesoamericano más reciente y con menor diversificación. El presente muestreo de loci no resolvió de manera concluyente el origen geográfico de la guayaba. Es posible que el área de origen haya sido oscurecida por el flujo de genes, como resultado del movimiento de frutos por parte de los dispersores, o que sea una especie ampliamente distribuida. Contrariamente a nuestra hipótesis sobre la dispersión de la guayaba por humanos hacia Mesoamérica, las evidencias que se muestran en el capítulo II indican que la guayaba se habría dispersado a Mesoamérica mucho antes

del cierre del Istmo de Panamá (O'Dea et al., 2016) y de la presencia de los humanos en América (Ardelean et al., 2020; Becerra-Valdivia & Higham, 2020). Por lo tanto, la distribución actual de la guayaba, desde México hasta la Argentina y Paraguay (Landrum, 2017), parece deberse primeramente a procesos biogeográficos y no a procesos antrópicos.

Diversidad y estructura genética de guayaba en el continente americano

La diversidad genética de la guayaba es mayor cuando se compara con otras especies de *Psidium* (capítulo III). No obstante, de acuerdo con los análisis genéticos usando microsatélites, la guayaba presenta una clara estructuración genética, pero, sorprendentemente, encontramos mayor diversidad en México y Centro América, seguidas estas regiones por Colombia y Venezuela (capitulo III). El flujo entre poblaciones de guayaba de Centro América, México y las Antillas, con el norte de América del Sur (Colombia y Venezuela), por el movimiento de dispersores naturales y humanos, podrían haber favorecido este aumento de la diversidad en Mesoamérica. Sin embargo, los datos arqueológicos revelan, hasta ahora, un uso más antiguo de la guayaba en América del Sur. Este resultado concuerda con los resultados del análisis de Computación Bayesiana Aproximada, el cual muestra que el proceso de domesticación de la guayaba probablemente habría iniciado en la Amazonia brasileña y se habría difundido desde allí hacia los Andes peruanos y más tarde hacia el norte de América del Sur y Mesoamérica. Así mismo, aunque no encontramos diferenciación debida a la distancia para todo el conjunto de datos, las poblaciones suramericanas de guayaba de los Andes y la Amazonía presentan mayores valores F_{ST}, por lo que probablemente condiciones ambientales podrían estar jugando un papel importante en la diferenciación de estas poblaciones, al contrario de las poblaciones de Venezuela, Colombia, Centroamérica y México que estarían manteniendo flujo a larga distancia.

Para una mejor compresión sobre el proceso de domesticación de la guayaba, es necesario investigar más regiones en América del Sur. Son particularmente relevantes las poblaciones de Argentina y Paraguay que no fueron muestreadas en este trabajo y un número mayor de poblaciones en Centroamérica y las selvas del sur de México. Si bien nuestros resultados apuntan a la Amazonía como área probable de domesticación,

es necesario ahondar en esta hipótesis y descartar o corroborar eventos múltiples de domesticación. Por lo tanto, trabajos más detallados de las poblaciones amazónicas y andinas y mesoamericanas podrían arrojar mayor luz sobre los detalles del proceso. Así mismo, la reconstrucción de rutas de intercambio a través de la descripción de las migraciones humanas de Sur a Norte, datos lingüísticos, y análisis demográficos más específicos de la guayaba podrían ayudar a esclarecer mejor estos patrones.

Por último, la incorporación de nuevas técnicas de investigación como la paleogenómica, permitiría precisar la identidad de los restos de polen registrados en México y Uruguay. Sin duda, la incorporación de estos enfoques serán clave en la determinación de el o los centros de domesticación y diversificación de la guayaba, los patrones manejo, cultivo, dispersión, consumo y comercio de la guayaba en diferentes períodos y culturas de Mesoamérica y Suramérica.

Referencias

Aranguren, Y., Briceño, A., & Fermin, G. (2008). Assessment of the variability of Venezuelan guava landraces by microsatellites. In W. Rohde & G. Fermin (Eds.), *II International Symposium on Guava and Other Myrtaceae* (Vol. 849, pp. 147–154).

Ardelean, C. F., Becerra-Valdivia, L., Pedersen, M. W., Schwenninger, J.-L., Oviatt, C. G., Macías-Quintero, J. I., Arroyo-Cabrales, J., Sikora, M., Ocampo-Díaz, Y. Z. E., & Rubio-Cisneros, I. I. (2020). Evidence of human occupation in Mexico around the Last Glacial Maximum. *Nature*, *584*(7819), 87–92.

Arévalo-Marín, E., Casas, A., Landrum, L., Shock, M. P., Alvarado-Sizzo, H., Ruiz-Sanchez, E., & Clement, C. R. (2021). The Taming of *Psidium guajava*: Natural and Cultural History of a Neotropical Fruit. *Frontiers in Plant Science*, *12*(September), 1–15. https://doi.org/10.3389/fpls.2021.714763

Becerra-Valdivia, L., & Higham, T. (2020). The timing and effect of the earliest human arrivals in North America. *Nature*, *584*(7819), 93–97.

Briceno, A., Aranguren, Y., & Fermin, G. (2010). Assessment of Guava-Derived SSR Markers for the Molecular Characterization of Myrtaceae from Different Ecosystems in Venezuela. In W. Rohde & G. Fermin (Eds.), *Ii International Symposium on Guava and Other Myrtaceae* (Vol. 849, pp. 139–146).

Gaut, B. S., Díez, C. M., & Morrell, P. L. (2015). Genomics and the Contrasting Dynamics of Annual and Perennial Domestication. *Trends in Genetics*, *31*(12), 709–719. https://doi.org/10.1016/j.tig.2015.10.002

Gutiérrez, R. M., Mitchell, S., & Solis, R. V. (2008). *Psidium guajava*: a review of its traditional uses, phytochemistry and pharmacology. *J Ethnopharmacol*, *117*(1), 1–27. https://doi.org/10.1016/j.jep.2008.01.025

Kanupriya, Latha, P. M., Aswath, C., Reddy, L., Padmakar, B., Vasugi, C., & Dinesh, M. R. (2011). Cultivar identification and genetic fingerprinting of guava (*Psidium guajava*) using microsatellite markers. *International Journal of Fruit Science*, *11*(2), 184–196. https://doi.org/10.1080/15538362.2011.578521

Kumari, S., Nagaraja Arumugam, R. S., Srivastav, M., Banoth, S., Mithra, A. C., GoswamiI, A. K., & Khan, Y. J. (2018). Diversity analysis of guava (*Psidium guajava*) germplasm collection. *Indian Journal of Agricultural Sciences*, *88*(3), 489–497.

Landrum, L. R. (2017). The genus *Psidium* (Myrtaceae) in the state of Bahia, Brazil. *Canotia*, *13*, 1–101.

Mehmood, A., Luo, S., Ahmad, N. M., Dong, C., Mahmood, T., Sajjad, Y., Jaskani, M. J., & Sharp, P. (2016). Molecular variability and phylogenetic relationships of guava (*Psidium guajava* L.) cultivars using inter-primer binding site (iPBS) and microsatellite (SSR) markers. *Genetic Resources and Crop Evolution*, *63*, 1345–1361.

Miller, A. J., & Gross, B. L. (2011). From forest to field: perennial fruit crop domestication. *American Journal of Botany*, *98*(9), 1389–1414.

O'Dea, A., Aguilera, O., Aubry, M. P., Berggren, W. A., Budd, A. F., Cione, A. L., Coates, A. G., Collins, L. S., Coppard, S. E., & Cozzuol, M. A. (2016). *Building Bridges. Response to Erkens and Hoorn: "The Panama Isthmus, 'old', 'young' or both?".*

Patiño, V. M. (2002). *Historia y dispersión de los frutales nativos del neotrópico*. Centro Internacional de Agricultura Tropical. https://books.google.com.mx/books?id=U12m7M2VMmsC

Rodríguez, N. N., Valdés, J., Rodríguez, J. a, Velásquez, J. B., Rivero, D., Martínez, F., González, G., & Sourd, D. G. (2010). Genetic resources and breeding of guava (*Psidium guajava* L.) in Cuba. *Journal of Genetics*, 2–5.

Rodríguez-Medina, N. N., Valdés-Infante, J., Velásquez, B., Rivero, D., Martínez, F., Risterucci, A. M., Billotte, N., Becker, D., Ritter, E., & Rohde, W. (2010). Individual versus combined data set for molecular characterization of Cuban guava *(Psidium guajava)* germplasm. *Acta Horticulturae*, *849*, 163–172. https://doi.org/10.17660/ActaHortic.2010.849.18

Sitther, V., Zhang, D., Harris, D. L., Yadav, A. K., Zee, F. T., Meinhardt, L. W., & Dhekney, S. A. (2014). Genetic characterization of guava (*Psidium guajava* L.) germplasm in the United States using microsatellite markers. *Genetic Resources and Crop Evolution*, *61*(4), 829–839. https://doi.org/10.1007/s10722-014-0078-5

Valdés-Infante, J., Rodríguez, N. N., Becker, D., Velázquez, B., Sourd, D., Espinosa, G., & Rohde, W. (2007). Microsatellite characterization of guava (*Psidium guajava* L.) germplasm collection in Cuba. *Cultivos Tropicales*, *28*(3), 5–11.

Valdés-Infante, J., Rodríguez-Medina, N. N., Velásquez, B., Rivero, D., Martínez, F., Risterucci, A. M., Billotte, N., Becker, D., Ritter, E., & Rohde, W. (2010). Comparison of the polymorphism level, discriminating capacity and informativeness of morpho-agronomic traits and molecular markers in guava (*Psidium guajava* L.). *Acta Horticulturae*, *849*, 121–132. https://doi.org/10.17660/ActaHortic.2010.849.13

Viji, G., Harris, D. L., Yadav, A. K., & Zee, F. T. (2010). Use of microsatellite markers to characterize genetic diversity of selected accessions of guava (*Psidium guajava*) in the United States. *Acta Horticulturae*, *859*, 169–176.

Watling, J., Shock, M. P., Mongeló, G. Z., Almeida, F. O., Kater, T., De Oliveira, P. E., & Neves, E. G. (2018). Direct archaeological evidence for Southwestern Amazonia as an early plant domestication and food production centre. *Plos One*, *13*(7), e0199868. https://doi.org/10.1371/journal.pone.0199868