

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias Biológicas y Ambientales

**Delimitación genética en el complejo de especies *Eschweilera*
del clado Parvifolia**

Camila Alejandra Contreras Guerrero

Ingeniería en Biotecnología

Trabajo de fin de carrera presentado como requisito
para la obtención del título de
Ingeniera en Biotecnología

Quito, 13 de mayo de 2022

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias Biológicas y Ambientales

HOJA DE CALIFICACIÓN DE TRABAJO DE FIN DE CARRERA

**Delimitación genética en el complejo de especies *Eschweilera*
del clado Parvifolia**

Camila Alejandra Contreras Guerrero

Nombre del profesor, Título académico

Pieter Van 't Hof, PhD.

Quito, 13 de mayo de 2022

© DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica de Educación Superior del Ecuador.

Nombres y apellidos: Camila Alejandra Contreras Guerrero

Código: 00201336

Cédula de identidad: 1722693908

Lugar y fecha: Quito, 13 de mayo de 2022

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en <http://bit.ly/COPETHeses>.

UNPUBLISHED DOCUMENT

Note: The following capstone project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around these publishing available on <http://bit.ly/COPETHeses>

RESUMEN

La región Neotropical alberga una gran cantidad de especies arbóreas que presentan amplia variabilidad y distribución. En las tierras bajas de la Amazonía, la distribución de especies arbóreas hiperdominantes se restringe a un conjunto reducido mientras las especies raras abarcan una mayor proporción. *Lecythidaceae* es una de las familias con mayor número de especies hiperdominantes, donde el género *Eschweilera* se destaca por exhibir una gran cantidad de especies hiperdominantes y raras como miembros del genus. Sin embargo, la identificación botánica basada en rasgos morfológicos y las relaciones filogenéticas en este género son poco concisas debido a la semejanza de especies arbóreas estrechamente relacionadas, existencia de especies crípticas y procesos evolutivos complejos. En Ecuador, la comprensión acerca de la distribución y la delimitación taxonómica de *Eschweilera* es aún limitada, por tanto, el objetivo del presente estudio fue el análisis de la estructura poblacional, diversidad genética y alélica en el complejo de especies de *Eschweilera* del clado Parvifolia. Para esto, se secuenció el genoma de diversas especies de *Eschweilera* a través del método SSR-Seq. Los resultados obtenidos fueron analizados por diversos programas computacionales para determinar la delimitación de especies y verificar su distribución por las poblaciones. El análisis a partir de marcadores microsatélite permitió comprender la diversidad genética a través de una diferenciación geográfica, sin embargo, se requiere mayor investigación para establecer una delimitación genética de especies de *Eschweilera* del clado Parvifolia. Estos resultados pueden ser útiles para estimar la diversidad de especies amazónicas y definir con precisión la taxonomía e historia evolutiva de especies hiperdominantes en Ecuador. De esta manera se podría comprender mejor los procesos ecosistémicos y las implicaciones para la conservación de la Amazonía.

Palabras clave: *Eschweilera*, microsatélite, delimitación de especies, especies hiperdominantes, relaciones filogenéticas

ABSTRACT

The Neotropical region harbors a large number of tree species with wide distribution ranges. In the Amazonian lowlands, the distribution of hyperdominant tree species is restricted to a reduced number, while rare species comprise a larger proportion. *Lecythidaceae* is one of the families with a high number of hyperdominant species, where the genus *Eschweilera* stands out with both hyperdominant and rare species. However, botanical identification based on morphological characteristics and phylogenetic relationships in this genus is not very clear due to the similarity of closely related tree species, the existence of cryptic species, and complex evolutionary processes involved. In Ecuador, the understanding of the distribution and taxonomic delimitation of species complex *Eschweilera* is still limited. Therefore, the objective of the present study has been to explore the genetic delimitation of *Eschweilera* species in Ecuador and French Guiana. For this purpose, the genome of several *Eschweilera* species was sequenced using the SSR-Seq method. The results obtained were analyzed using an array of population genetic software to determine the species delimitation and to verify their population. The analysis of microsatellite markers allowed us to understand the genetic diversity through geographic differentiation; however, further research is required to establish a more detailed genetic delimitation of *Eschweilera* species of the *Parvifolia* clade. These results can be useful to estimate the diversity of Amazonian species and to precisely define the taxonomy and evolutionary history of hyperdominant species, thereby facilitating a better understanding of ecosystem processes with implications for the conservation of the Amazon Rainforest.

Key words: *Eschweilera*, microsatellite, species delimitation, hyperdominant species, phylogenetic relationships

TABLA DE CONTENIDOS

1.	INTRODUCCIÓN	11
2.	METODOLOGÍA	15
2.1	Recolección de muestras y extracción de ADN	15
2.2	Diseño de cebadores	15
2.3	Pruebas de modificación de cebadores y amplificación simple	15
2.4	Amplificación múltiple de microsatélite y construcción de bibliotecas de secuenciación	15
2.5	Análisis bioinformático	16
2.6	Filtración de datos	16
2.7	Análisis de estructura poblacional.....	17
2.8	Análisis de diversidad alélica y genética.....	17
3.	RESULTADOS	18
3.1	Análisis bioinformático	18
3.2	Análisis de datos (2n)	18
3.2.1	Análisis de estructura poblacional.....	18
3.2.2	Análisis de diversidad alélica y genética.....	18
3.3	Análisis de datos (4n).....	19
3.3.1	Análisis de estructura poblacional.....	19
3.3.2	Análisis de diversidad alélica y genética.....	20
4.	DISCUSIÓN	22
5.	CONCLUSIONES	25
6.	TABLAS	26
7.	FIGURAS	29
8.	REFERENCIAS BIBLIOGRÁFICAS.....	32
9.	ANEXOS.....	35

ÍNDICE DE TABLAS

Tabla 1. Índice de diversidad genética y alélica para individuos diploides en las zonas de estudio de <i>Eschweilera</i>	26
Tabla 2. Distancias genéticas de Nei para individuos diploides para cada zona de estudio de <i>Eschweilera</i>	27
Tabla 3. Índice de diversidad genética y alélica para individuos tetraploides en las zonas de estudio de <i>Eschweilera</i>	27
Tabla 4. Distancias genéticas de Nei para individuos tetraploides para cada zona de estudio de <i>Eschweilera</i>	28

ÍNDICE DE FIGURAS

Figura 1. Gráfico de coeficientes de asignación diploide para las zonas de estudio de <i>Eschweilera</i>	29
Figura 2. Análisis de Coordenadas Principales (PCoA) entre las zonas de estudio de <i>Eschweilera</i>	29
Figura 3. Gráfico de coeficientes de asignación tetraploide para las zonas de estudio de <i>Eschweilera</i>	30
Figura 4. Gráfico de coeficientes de asignación tetraploide para las zonas de estudio de <i>Eschweilera</i>	30
Figura 5. Análisis de Coordenadas Principales (PCoA) entre las zonas de estudio de <i>Eschweilera</i>	31

ÍNDICE DE ANEXOS

Anexo 1: Mapa de las zonas de extracción del material foliar de *Eschweilera* en Ecuador... 35

1. INTRODUCCIÓN

El Neotrópico es una región biogeográfica que abarca la región tropical del continente Americano (Zunino & Zunilli, 2003). Se distribuye a lo largo de Sudamérica, América Central, el sur de México, las Antillas y el sur de Florida. La región Neotropical se compone de 3 subregiones (Antillana, Brasileña y Chaqueña), 2 zonas de transición (Mexicana y Sudamericana), siete dominios (Mesoamericano, Pacífico, Boreal Brasileño, Sudoeste Amazónico, Sudeste Amazónico, Chacoano y Paraná) y 53 provincias (Morrone, 2014). La biodiversidad presente en los trópicos americanos es la más rica del mundo (Antonelli et al., 2018), resultado de una evolución de miles de años que involucra diversos procesos bióticos y abióticos, cambios climáticos y geológicos, hábitats heterogéneos e interacciones ecológicas (Antonelli, 2021). Los bosques húmedos neotropicales se caracterizan por ser ecosistemas con gran diversidad de plantas a escala local y entre sitios, siendo la Amazonía una de las áreas más notables (Urquhart, 2020). Una hectárea de selva amazónica es capaz de albergar 900 especies de plantas vasculares y 300 especies de árboles (Balslev et al., 1998).

En la región neotropical, así como es en otras áreas, la distribución de las plantas no es homogénea y el número de individuos por especies puede presentar una alta variabilidad. Según una investigación reciente, las tierras bajas de la Amazonía comprenden alrededor de 16 000 especies de árboles, donde la mitad de los árboles de toda la Amazonía corresponden a 227 especies “hiperdominantes” mientras que 11000 especies raras se distribuyen en una menor proporción (Ter Steege et al., 2013). Los patrones de oligarquía o hiperdominancia en la naturaleza se constituyen por comunidades que alojan un grupo reducido de especies y se disponen de manera abundante en un área específica (Draper et al., 2019). Se ha reportado que los hiperdominantes se extienden por rangos geográficamente amplios, sin embargo, sólo son oligarcas en una o dos regiones y con frecuencia se especializan en hábitats (Ter Steege et al., 2013). En la región amazónica, las especies hiperdominantes cumplen un rol

fundamental, pues se encargan de almacenar la mitad del carbono del ambiente, además de ser importantes para mantener niveles de productividad altos (Fauset et al., 2015). Entre las familias con mayor proporción de especies hiperdominantes predomina: *Arecaceae*, *Myristicaceae* y *Lecythidaceae* (Ter Steege et al., 2013).

La clasificación de especies arbóreas como hiperdominantes tiene su base en la precisa identificación botánica, no obstante, este proceso puede resultar complicado debido a la abundancia de hiperdominantes pertenecientes a géneros taxonómicos complejos y abundantes en especies (Funk et al., 2007). Sin embargo, las limitaciones por establecer una correcta identificación botánica no se limitan a un único factor, pues, existe carencia de caracteres diagnósticos que evidencien una diferenciación clara. en especies arbóreas estrechamente relacionadas cuya morfología tiende a ser similar (Goodwin, 2015). Además, la presencia de diversas especies hiperdominantes crípticas, caracterizadas por integrarse a una errónea clasificación de dos o más especies bajo una sola especie, es un gran problema que puede conllevar a errores en la tipificación de especies hiperdominantes (Cavers et al., 2013; Bickford et al., 2007).

El género *Eschweilera* perteneciente a la familia *Lecythidaceae* se divide en 3 clados parafiléticos: *Integrifolia*, *Tetrapetala* y *Parvifolia* (Mori et al., 2017). Se caracteriza por integrar varias especies hiperdominantes y raras, sin embargo, la asignación de relaciones evolutivas en los clados es poco clara, especialmente en el clado *Parvifolia*, lo cual, dificulta conocer si la asignación taxonómica fundamentada en elementos morfológicos involucra uno o varios linajes (Huang, Mori, & Kelly, 2015). La principal limitante en el clado *Parvifolia* radica en la presencia de una gran variedad de especies en simpatría que comparten haplotipos de ADN plastidial, cuyas secuencias son incapaces de segregar a las especies (Caron et al., 2019; Mori, 1987; Mori & Lepsch-Cunha, 1995). Este suceso sugiere el surgimiento de eventos de ascendencia común recientes, divergencia genética incompleta, clasificación incompleta de linajes o hibridación interespecífica (Mori et al., 2017; Heuertz et al., 2020).

Además, estudios genéticos en *Eschweilera* reportaron la presencia de hasta cuatro alelos por genotipo en ciertos loci, lo cual sugiere la presencia de taxones tetraploides originados de un evento de duplicación del genoma antiguo común en todas las especies y loci reportados (Heuertz et al., 2020). Debido a la variación en la representación morfológica dentro de las especies de *Eschweilera* y la aleatoriedad e indeterminación de eventos reproductivos en especies arbóreas tropicales, la delimitación genética y botánica en especies relacionadas es comúnmente un desafío (Mori & Prance, 1990). Sin embargo, el uso de marcadores moleculares microsatélite para la delimitación de especies complejas puede contribuir enormemente a detectar cambios genéticos que permitan establecer distinción entre especies, además de otorgar una correcta inferencia filogenética (Twyford & Ennos, 2012).

Investigaciones botánicas realizadas en Guayana Francesa y la Amazonía Central ha determinado que especies comunes y raras de *Eschweilera* son capaces de convivir en la misma parcela forestal, lo cual es ideal para el estudio de la distribución y las relaciones evolutivas en este género (Huang, Mori, & Kelly, 2015). En Guayana Francesa la delimitación del complejo de especies de *Eschweilera* ha sido evidenciado en varias investigaciones, especialmente en *Eschweilera coriacea*, la única especie arbórea catalogada como hiperdominante en seis regiones de la Amazonía (Ter Steege et al., 2013). Por otro lado, en Ecuador la investigación de especies hiperdominantes arbóreas locales es poco profundizada. Por tanto, el contraste de investigaciones entre Guayana Francesa y Ecuador podría establecer una base esencial hacia un mayor conocimiento de las especies hiperdominantes, su distribución en la Amazonía y las relaciones filogenética en *Eschweilera* del clado Parvifolia.

El objetivo de este estudio fue analizar los datos obtenidos de la secuenciación con marcadores microsatélite de especies de *Eschweilera* en Guayana Francesa y Ecuador para generar una delimitación genéticamente de las especies presentes. Para lograr este objetivo se secuenció el genoma de varias especies de *Eschweilera* mediante el método SSR-Seq. Los

resultados obtenidos fueron analizados por diversos programas computacionales para determinar la delimitación de especies y verificar su distribución por las poblaciones. Esta información puede ser útil para estimar la diversidad de especies amazónicas y definir con precisión la taxonomía e historia evolutiva de hiperdominantes. De esta manera se podría comprender mejor los procesos ecosistémicos y las implicaciones para la conservación de la Amazonía.

2. METODOLOGÍA

2.1 Recolección de muestras y extracción de ADN

El material vegetal fue recolectado en cuatro zonas de la Amazonía: La Estación de Biodiversidad Tiputini y la Estación científica Yasuní en Ecuador y los inventarios forestales de Piste de St Élie y Paracou en Guayana Francesa (Anexo 1). Se extrajo el material foliar de 190 árboles representativos como miembros de la familia *Lecythidaceae* pertenecientes al género *Eschweilera*. Las muestras vegetales fueron extraídas y preservadas en bolsas con perlas de sílica. Posteriormente, la extracción de ADN fue realizado con el protocolo Invisorb DNA Plant HTS 96 kit según las instrucciones del kit.

2.2 Diseño de cebadores

Se diseñaron 48 cebadores en base a loci previamente desarrollados, utilizado como referencia base las secuencias genómicas de especies estrechamente relacionadas a *Eschweilera*. Las secuencias de los cebadores se mapearon respecto al genoma de referencia en bowtie 2 v2.3.4.1 (Langmead & Salzberg , 2012) y mediante bedtools v2.25.0 (Quinlan, 2014) se extrajo la secuencia genómica que comprende desde 200 pb aguas abajo del cebador directo hasta 200 pb aguas arriba de la posición del cebador inverso.

2.3 Pruebas de modificación de cebadores y amplificación simple

Los pares de cebadores se probaron mediante amplificación simple utilizando el kit Qiagen Multiplex (Lepais & Bacles, 2011) con un volumen final de 10 μ L y una concentración final de cada cebador de 0,2 μ M. Las condiciones de la amplificación fueron: 95 °C por 15 minutos, seguido de 35 ciclos a 95 °C por 20 segundos, 59 °C durante 60 segundos y 72 °C por 30 segundos, con una extensión final de 72 °C durante 10 minutos. La amplificación de los locus se visualizó bajo luz ultravioleta.

2.4 Amplificación múltiple de microsatélite y construcción de bibliotecas de secuenciación

Se realizó una PCR multiplex de tres rondas para obtener la amplificación simultánea de los loci. En la tercera ronda se añadieron los adaptadores de secuenciación Illumina y los códigos de barras correspondientes. Se realizó el control de calidad de los amplicones con el kit Agilent TapeStation D1000 y el sistema fluorométrico Qubit (Thermo Fisher Scientific), mientras la cuantificación total se realizó con el kit de cuantificación de bibliotecas Kapa en un PCR cuantitativo Roche LightCycler 480. Se secuenciaron 12 muestras, 6 de Guayana Francesa y 6 de Ecuador en Illumina MiSeq -1/3 V2 PE.

2.5 Análisis bioinformático

La calidad de las secuencias se analizó mediante FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), se removieron las lecturas con valores menores a 70 pb mediante cutadapt (Martin, 2011) y las lecturas se ensamblaron en contigs utilizando pear (Zhang et al., 2014). A continuación, se clasificaron las secuencias en archivos fastq mediante fqgrep (<https://github.com/indraniel/fqgrep>). Luego, se utilizó el pipeline FDSTools v1.1.1 (Hoogenboom et al., 2017) para determinar las secuencias correspondientes a los alelos microsatélite e identificar los genotipos de cada individuo. Se tabuló la tabla de genotipos resultantes y se analizó la cobertura de los locus, las características de los locus, información de los alelos y genotipos encontrados

2.6 Filtración de datos

Se utilizó Excel para generar una primera matriz tetraploide donde se conservaron todos los individuos de la matriz cruda y se eliminaron los loci con >25% de datos faltantes, estableciendo un tamaño de muestras final de 212 individuos, 10 loci y 20 especies de *Esxhweilera*. Se generó una segunda matriz tetraploide donde se eliminaron los individuos con réplicas cuyos datos no correspondieran entre sí. Adicionalmente, se excluyeron los individuos con >80% de datos faltantes (139 individuos) y los loci con >50% de datos faltantes (6 loci), estableciendo un tamaño de muestras final de 71 individuos, 17 loci y 13

especies de *Eschweilera*.

Para la elaboración de la matriz diploide se tomó como base el tamaño de muestra final de la matriz tetraploide y se realizó un subconjunto diplotizado de datos mediante la eliminación de los individuos con < 55% de datos con más de dos alelos (24 individuos) y los loci con < 55% de datos con más de dos alelos (8 loci), determinando un tamaño de muestra final de 49 individuos y 9 loci.

2.7 Análisis de estructura poblacional

Las matrices fueron modificadas a un formato .txt para ser compatibles con el Software STRUCTURE (Pritchard, Stephens, & Donn, 2000). Los parámetros incorporados a STRUCTURE fueron: modelo ancestral “admixture” con frecuencias alélicas correlacionadas. El análisis de datos diploides y tetraploides fue realizado en una única corrida y se generó bajo los siguientes criterios: burning de 100000, MCMC de 100000, rango de, K=1 a K=10 por 10 iteraciones. Los resultados obtenidos fueron analizados en el programa Structure Harvester (Earl & vonHoldt , 2012) y Clumpak (Earl & vonHoldt , 2012).

2.8 Análisis de diversidad alélica y genética

Las matrices se modificaron a un formato .txt para ser compatibles con R studio (R studio team, 2015). Para ambos archivos se obtuvo el número de alelos, heterocigocidad observada (H_o) y heterocigocidad esperada (H_e), número de alelos privados y la riqueza alélica. Además, se ejecutó un Análisis de Coordenadas Principales (PCoA) para obtener la distribución de la diversidad alélica y la estructura poblacional de las muestras.

3. RESULTADOS

3.1 Análisis bioinformático

Se determinaron 130 muestras provenientes de Ecuador, 155 muestras de Guayana Francesa y 3 muestras indeterminadas, originando un tamaño de muestra de 212 individuos incluidas réplicas. Los datos generaron un total de 48 loci. Se excluyeron los loci con >50% de datos faltantes y un >10% de tasa de error, estableciendo un tamaño de muestras final de 24 loci. Se establecieron 212 individuos pertenecientes a 20 diferentes especies de *Eschweilera* y 5 especies fuera de grupo.

3.2 Análisis de datos (2n)

3.2.1 Análisis de estructura poblacional.

El análisis de estructura poblacional en Structure Harvester apuntó a K óptimo de K=3 ($\Delta K = 87.478$). Esto se contrasta con los datos de STRUCTURE analizados en Clumpak que indicaron la presencia de tres linajes genéticos mostrados en color celeste, azul marino y violeta entre las poblaciones analizadas. La distribución del color azul se observa dominante en Guayana Francesa, sin embargo, posee una distribución más pequeña del color violeta. Mientras Católica y Tiputini muestran una mayor dominancia del color celeste, como se indica en la Figura # 1.

3.2.2 Análisis de diversidad alélica y genética.

Los datos obtenidos de R studio indican un número totales de 318 alelos distribuidos entre las poblaciones de Guayana Francesa, Católica y Tiputini. En la Tabla #1 se aprecia los datos de diversidad alélica y genética distribuidos a partir de los marcadores seleccionados. El locus 38 alcanzó el mayor número de alelos por locus (47 alelos) mientras el locus 4 obtuvo el número más bajo de alelos (12 alelos). Por otro lado, Guayana Francesa obtuvo los valores más elevados respecto al número de alelos (N_a), número de alelos privados (PA), riqueza alélica (AR). Mientras Tiputini adquirió los valores más bajos en los análisis realizados. El

valor p de la prueba de equilibrio Hardy-Weinberg en los loci muestra valores $p < 0.05$ en casi todos los locus presentados, excepto por el locus 43 ($p = 1.00$). Por otro lado, a nivel poblacional, Guayana Francesa ($p = 0.0000$) y Católica ($p = 0.0000$) presentan valores $p < 0.05$ mientras Tiputini no es significativo ($p = 0.2438$). La heterocigocidad esperada (H_e) entre poblaciones fue superior en Guayana Francesa ($H_e = 0,90$) e inferior en Tiputini ($H_e = 0,69$). Por otro lado, la heterocigocidad observada (H_o) se determina con un valor más alto en Tiputini ($H_o = 0,72$) y con un valor más bajo en Católica ($H_o = 0,68$). El Análisis de Coordenadas Principales (PCoA) mostró 3 agrupaciones marcadas definidas por las poblaciones de Guayana Francesa, Católica y Tiputini como se observa en la Figura #2. Las agrupaciones de Católica y Tiputini se superponen entre sí, mientras Guayana Francesa se ubica poco alejada de estas poblaciones. Las agrupaciones se manifiestan cercanas entre sí, especialmente entre Católica y Guayana Francesa. La dispersión de las muestras indica que dos muestras de Guayana Francesa se encuentran cercanas a la agrupación de Católica. Mientras una única muestra de Católica se vincula cerca de la agrupación de Guayana Francesa. Este contenido se respalda en la Tabla # 2 donde se determina las distancias genéticas entre poblaciones. El valor de la distancia genética de Nei entre Católica y Tiputini es bajo (0.174) comparado con los valores de las distancias genéticas entre Guayana Francesa y Católica (0.813), y Guayana Francesa y Tiputini (0.749).

3.3 Análisis de datos (4n)

3.3.1 Análisis de estructura poblacional.

El análisis poblacional de la primera matriz tetraploide manifestó un K óptimo de $K = 8$ ($\Delta K = 248.361$). Esto se contrasta con los datos de STRUCTURE analizados en Clumpak que indicaron la presencia de ocho linajes genéticos entre las poblaciones analizadas, mostrados en color celeste, azul marino, azul, violeta, blanco, turquesa, verde agua y lila. La distribución de colores es muy diversa a través de las poblaciones, como se indica en la Figura #3. Mientras el análisis de estructura poblacional en la segunda matriz tetraploide apuntó a K

óptimo de $K=7$ ($\Delta K = 33.430$). Esto se contrasta con los datos de STRUCTURE analizados en Clumpak que indicaron la presencia de siete linajes genéticos mostrados en color azul marino, azul, violeta, gris, turquesa, verde agua y lila entre las poblaciones analizadas. La distribución del color azul se observa dominante en Guayana Francesa, mientras Católica y Tiputini muestran una mayor dominancia del color azul marino, como se indica en la Figura #4.

3.3.2 Análisis de diversidad alélica y genética.

Los datos obtenidos de R studio de la segunda matriz tetraploide indican un número totales de 785 alelos distribuidos entre las poblaciones de Guayana Francesa, Católica y Tiputini. En la Tabla #3 se aprecia los datos de diversidad alélica y genética distribuidos a partir de los marcadores seleccionados. El locus 23 alcanzó el mayor número de alelos por locus (68 alelos) mientras el locus 4 obtuvo el número más bajo de alelos (10 alelos). Por otro lado, al igual que en el análisis diploide, la población en Guayana Francesa obtuvo los valores más elevados respecto al número de alelos (N_a), número de alelos privados (PA), riqueza alélica (AR). Mientras Tiputini adquirió los valores más bajos en los análisis realizados. La heterocigocidad esperada (H_e) entre poblaciones fue superior en Guayana Francesa ($H_e = 0,91$) e inferior en Tiputini ($H_e = 0,71$). Sin embargo, la heterocigocidad observada (H_o) mantuvo un mismo valor para todas las poblaciones ($H_o = 1$). El Análisis de Coordenadas Principales (PCoA) mostró 3 agrupaciones marcadas definidas por las poblaciones de Guayana Francesa, Católica y Tiputini como se observa en la Figura #5. Las agrupaciones se manifiestan cercanas entre sí, especialmente Católica y Guayana Francesa. La dispersión de las muestras indica que una muestra de Guayana Francesa se encuentra vinculadas a las agrupaciones de Católica y Tiputini, y una muestras de Católica se vincula a la agrupación de Guayana Francesa. Este contenido se respalda en la Tabla #4 donde se determina las distancias genéticas entre poblaciones. El valor de la distancia genética de Nei entre Católica y Tiputini es bajo (0.171) comparado con los valores de las distancias genéticas entre

Guayana Francesa y Católica (0.838), y Guayana Francesa y Tiputini (0.874).

Cabe destacar que el análisis de la primera matriz no se realizó ya que el propósito de dichos datos se destinó únicamente para la comparación de diversidad genética entre los resultados obtenidos de las dos matrices tetraploides

4. DISCUSIÓN

La distinción de especies putativas se basa en un diagnóstico de características morfológicas, sin embargo, el uso de análisis genéticos es indispensable para confirmar una correcta clasificación botánica. El uso de marcadores microsatélite para la delimitación genética de especies del clado Parvifolia ha sido un gran avance para el descubrimiento de las relaciones filogenéticas en *Eschweilera*, no obstante, el panorama sobre la constitución genética y los procesos de hibridación desarrollados en el clado Parvifolia son aún un reto (Heuertz et al., 2020). Por lo tanto, es importante implementar nuevas metodologías y tecnologías capaces de establecer un escenario resolutivo ante la compleja relación evolutiva establecida en el clado Parvifolia.

La delimitación genética en el complejo de especies de *Eschweilera* del clado Parvifolia permitió establecer diferenciación genética entre especies de Ecuador y Guayana Francesa. El análisis de estructura poblacional en la matriz diploide permitió comprender el proceso evolutivo de diplotización determinado en las especies de *Eschweilera* del clado Parvifolia, donde únicamente se detectó 2 acervos genéticos mayoritarios en toda la población, que se destacan por mostrar una diferenciación notable entre la población de Ecuador y Guayana Francesa. Los diferentes patrones mostrados en cada acervo genético pueden deberse a eventos climáticos o geohistóricos capaces de generar barreras de dispersión geográfica a través del tiempo (Hughes, Pennington, & Antonelli, 2012). Además de considerar diferentes zonas de muestreo en la Amazonía, lo cual indica diferenciación genética asociadas a barreras físicas (Jinga & Ashley, 2018). Esto se confirma también en la figura de Análisis de Coordenadas Principales (PCoA), donde existen dos agrupamientos establecidos entre Ecuador y Guayana Francesa. Los resultados obtenidos en el PcoA se respaldan con los valores de las distancias genéticas de Nei, donde se indica que los individuos de Ecuador son genéticamente más distantes que los individuos de Guayana

Francesa. Respecto al análisis de diversidad genética, se observó diversidad genética moderadamente alta en Guayana Francesa frente a Ecuador, lo cual indica gran potencial adaptativo y persistencia de la población a largo plazo (Greenbaum et al., 2014). Un mayor número de alelos en Guayana Francesa puede relacionarse con un mayor número de especies de *Eschweilera* recolectadas frente a Ecuador. Sin embargo, es posible que intervengan posibles eventos fundadores en la Amazonía Ecuatoriana (Greenbaum et al., 2014).

Por otro lado, el análisis de los datos tetraploides permitió comprender la diversidad genética y alélica en la paleoploidía presente de *Eschweilera* del clado *Parvifolia*. El análisis de estructura poblacional resultó en 8 y 7 acervos genéticos dentro de la población. Ambos resultados presentan gran congruencia entre sí al tratarse de datos netamente comparativos, sin embargo, indica una baja resolución de los marcadores ya que el poder resolutivo no permitió la delimitación genética de las especies *Eschweilera* del clado *Parvifolia*. La baja diferenciación genética se relaciona con la introgresión y la hibridación manifestada en los procesos reproductivos del clado *Parvifolia*, lo cual aún representa un desafío (Gonzalez et al., 2009). Aunque también puede vincularse una identificación incorrecta de especies basada en la morfología dentro de los datos.

En Guayana francesa se presentan 3 acervos genéticos dominantes distinguidos en color lila, gris y verde, mientras que Ecuador posee un único acervo genético mayoritario. Esto puede deberse a las zonas de extracción de muestras, ya que la mayoría de las especies hiperdominantes se caracterizan por ser especialistas en hábitats, especialmente dominantes en ciertos tipos de bosques (Ter Steege et al., 2013). Lo cual posiblemente indicaría mayor abundancia de especies en Guayana Francesa frente a Ecuador. Además, al comparar resultados diploides y tetraploides se observa una diferencia entre los valores de acervos genéticos, lo cual indicaría una posible diferencia en base a la diversidad y el número de individuos homocigotos (Cazco, 2018). Por otro lado, la información tetraploide de la

estructura poblacional se respalda con la figura de Análisis de Coordenadas Principales (PCoA), donde existen dos agrupamientos establecidos entre Ecuador y Guayana Francesa. Los resultados obtenidos en el PcoA se confirman con los valores de las distancias genéticas de Nei, donde se indica que los individuos de Ecuador son genéticamente más distantes que los individuos de Guayana Francesa, similares a los resultados obtenidos con la matriz diploide. Respecto al análisis de diversidad genética, se observó diversidad genética moderadamente alta en Guayana Francesa frente a Ecuador. Por otro lado, se ha reportado que especies tetraploides muestra mayores valores de riqueza alélica frente a especies diploides, lo cual es acorde a los resultados obtenidos en este estudio. (Brown & Young, 2000).

En cuanto a la determinación de secuencias microsatélite a través del método SSR- seq demostró ser una herramienta importante para determinar especies con una falta de diagnósticos fenotípicos, como es el caso de *Eschweilera*. No obstante, se requiere de un mayor conjunto de marcadores microsatélite y un patrón de muestreo más amplio y homogéneo que permita determinar la constitución genética en el clado *Parvifolia*, además de generar una caracterización acerca de la historia evolutiva de especies abundantes y raras pertenecientes al complejo de especies de *Eschweilera* para comprender los elementos genéticos detrás de la hiperdominancia de especies arbóreas en la Amazonía (Heuertz et al., 2020)

5. CONCLUSIONES

En el presente estudio, se logró analizar por primera vez la diversidad genética a través de una diferenciación geográfica mediante el uso de marcadores microsatélite, dentro de Ecuador y Guayana Francesa. El método SSR-seq permitió una amplificación y secuenciación adecuada para establecer un genotipado de microsatélite preciso y confiable, siendo eficaz para la identificación de marcadores microsatélite en el complejo de especies de *Eschweilera*. En total se detectaron 3 acervos genéticos en Guayana Francesa y 1 acervo genético en Ecuador dentro de la matriz tetraploides y en la matriz diploide se encontró 1 acervo genético mayoritario en Guayana Francesa y 1 acervo genético en Ecuador. Estos hallazgos son una base para comprender las relaciones evolutivas dentro del clado Parvifolia. Los marcadores microsatélite identificados representan elementos útiles para profundizar en el análisis de estructuras poblacionales y el desarrollo de nuevos estudios en *Eschweilera*. Para lograr esto, se recomienda realizar análisis más detallados que sean capaces de establecer delimitación genética a nivel de especie en el complejo de especies de *Eschweilera*. Así se podrá generar mayor conocimiento sobre los posibles eventos evolutivos vinculados al clado Parvifolia. Además, se propone llevar a cabo investigaciones que involucren marcadores microsatélite con mayor poder resolutivo o la incorporación de diferentes metodologías, con el fin de mejorar la resolución obtenida de los resultados. También, se debe considerar un enfoque mixto que integre información de rasgos morfológicos, caracterización de nicho ecológico y secuenciación genómica de alto rendimiento. Esto permitirá un mejor entendimiento sobre los patrones de hiperdominancia y su importancia dentro de los procesos ecosistémicos en la Amazonía.

6. TABLAS

Tabla 1. Índice de diversidad genética y alélica para individuos diploides en las zonas de estudio de *Eschweilera*. A) Tabla de diversidad alélica presente en las muestras. B) Tabla de diversidad genética en las muestras.

A.

Marcador	Alelos por locus	Heterocigocidad esperada (He) por locus	Heterocigocidad observada (Ho) por locus	p-value prueba de desviación del equilibrio Hardy-Weinberg
ssrseq eschweilera 004	12	0.633	0.612	0.018
ssrseq eschweilera 009	31	0.883	0.755	0.000
ssrseq eschweilera 018	24	0.892	0.673	0.000
ssrseq eschweilera 023	45	0.946	0.830	0.000
ssrseq eschweilera 028	18	0.861	0.660	0.009
ssrseq eschweilera 030	19	0.691	0.531	0.036
ssrseq eschweilera 038	47	0.955	0.755	0.000
ssrseq eschweilera 043	17	0.848	0.776	1.000
ssrseq eschweilera 047	23	0.910	0.714	0.000

B.

Grupo	Muestras analizadas (n)	Número de alelos (Na)	Número de alelos privados (PA)	Riqueza alélica (AR)	Riqueza alélica por rarefacción (ARr)	Heterocigocidad esperada (He)	Heterocigocidad observada (Ho)	Valor p equilibrio Hardy-Weinberg
Guyana Francesa	19	158	114	20.38	8.23	0.88	0.71	0.0000
Católica	23	105	51	16.29	5.88	0.73	0.68	0.0000
Tiputini	7	55	12	8.68	5.79	0.71	0.72	0.2438

Tabla 2. Distancias genéticas de Nei para individuos diploides para cada zona de estudio de *Eschweilera*

	Católica	Guyana Francesa	Tiputini
Católica	0.000000		
Guyana Francesa	0.8126372	0.000000	
Tiputini	0.1744028	0.7487354	0.000000

Tabla 3. Índice de diversidad genética y alélica para individuos tetraploides en las zonas de estudio de *Eschweilera*. A) Tabla de diversidad alélica presente en las muestras. B) Tabla de diversidad genética en las muestras.

A.

Marcador	Alelos por locus	Heterocigocidad esperada (He) por locus	Heterocigocidad observada (Ho) por locus
ssrseq eschweilera 001	54	0.971	1
ssrseq eschweilera 004	10	0.634	1
ssrseq eschweilera 009	35	0.903	1
ssrseq eschweilera 013	35	0.813	1
ssrseq eschweilera 015	38	0.951	1
ssrseq eschweilera 016	24	0.863	1
ssrseq eschweilera 018	30	0.917	1
ssrseq eschweilera 023	68	0.968	1
ssrseq eschweilera 024	30	0.890	1
ssrseq eschweilera 026	64	0.976	1
ssrseq eschweilera 028	20	0.854	1
ssrseq eschweilera 030	25	0.778	1
ssrseq eschweilera 037	46	0.942	1
ssrseq eschweilera 038	55	0.960	1
ssrseq eschweilera 043	18	0.826	1
ssrseq eschweilera 045	51	0.950	1
ssrseq eschweilera 047	28	0.914	1

B.

Grupo	Muestras analizadas (n)	Número de alelos (Na)	Número de alelos privados (PA)	Riqueza alélica (AR)	Riqueza alélica por rarefacción (ARr)	Heterocigicidad esperada (He)	Heterocigicidad observada (Ho)
Guayana Francesa	31	420	346	29,81	8,67	0.91	1
Católica	29	238	132	21,28	6,33	0.78	1
Tiputini	11	127	41	12,16	5,90	0.71	1

Tabla 4. Distancias genéticas de Nei para individuos tetraploides para cada zona de estudio de *Eschweilera*.

	Católica	Guyana Francesa	Tiputini
Católica	0.000000		
Guyana Francesa	0.8377848	0.000000	
Tiputini	0.1713139	0.8740991	0.000000

7. FIGURAS

K=3

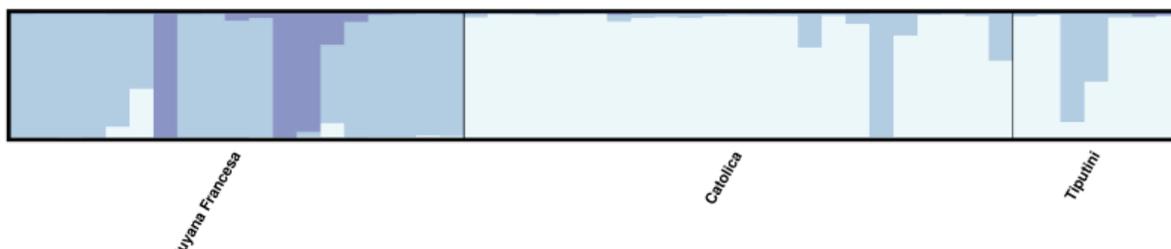


Figura 1. Gráfico de coeficientes de asignación diploide para las zonas de estudio de *Eschweilera*. El K óptimo asignado es K=3. La variación de colores azul, violeta y celeste entre las zonas establece la representación de linajes presentes y su distribución.

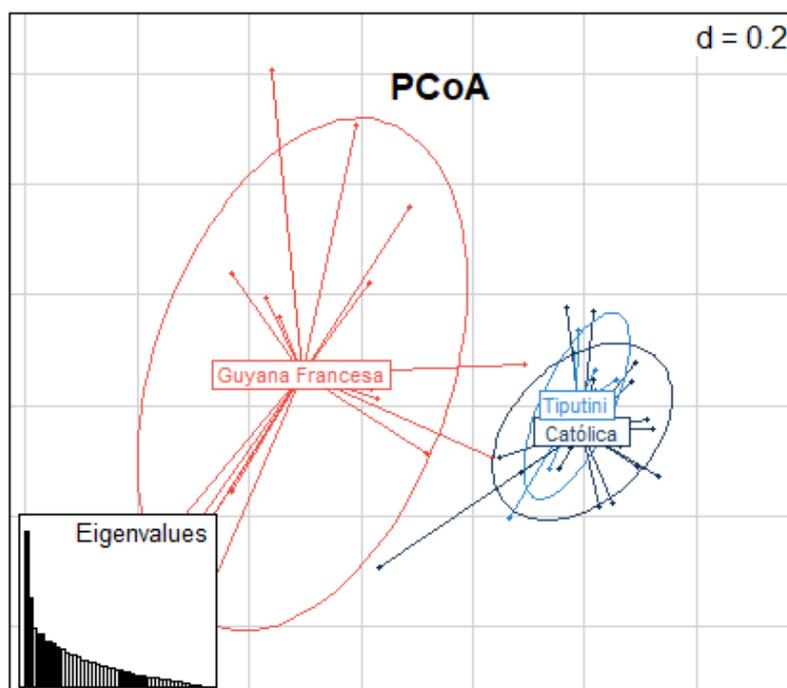


Figura 2. Análisis de Coordenadas Principales (PCoA) entre las zonas de estudio de *Eschweilera*. Representación de la distribución de muestras de *Eschweilera* en Guayana Francesa, Tiputini y Católica.

K=8

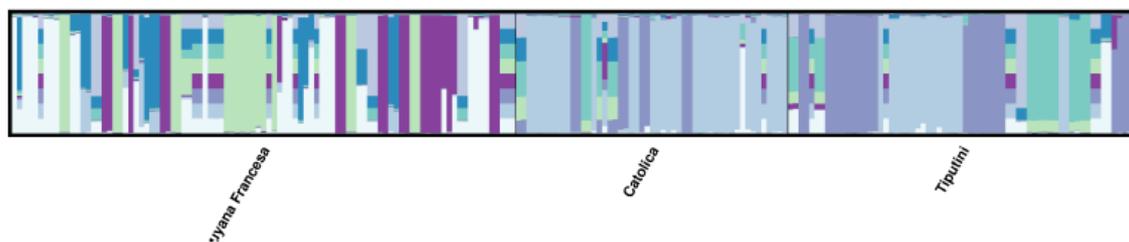


Figura 3. Gráfico de coeficientes de asignación tetraploide para las zonas de estudio de *Eschweilera*. El K óptimo asignado es K=8. La variación de colores celeste, azul marino, azul, violeta, blanco, turquesa, verde agua y lila entre las zonas establece la representación de linajes presentes y su distribución.

K=7

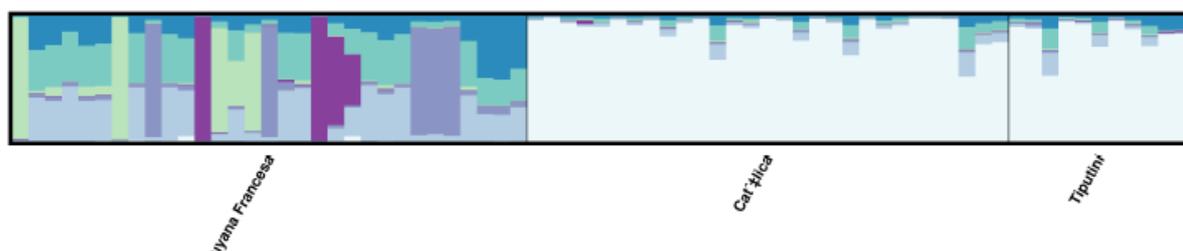


Figura 4. Gráfico de coeficientes de asignación tetraploide para las zonas de estudio de *Eschweilera*. El K óptimo asignado es K=7. La variación de colores azul marino, azul, violeta, gris, turquesa, verde agua y lila entre las zonas establece la representación de linajes presentes y su distribución.

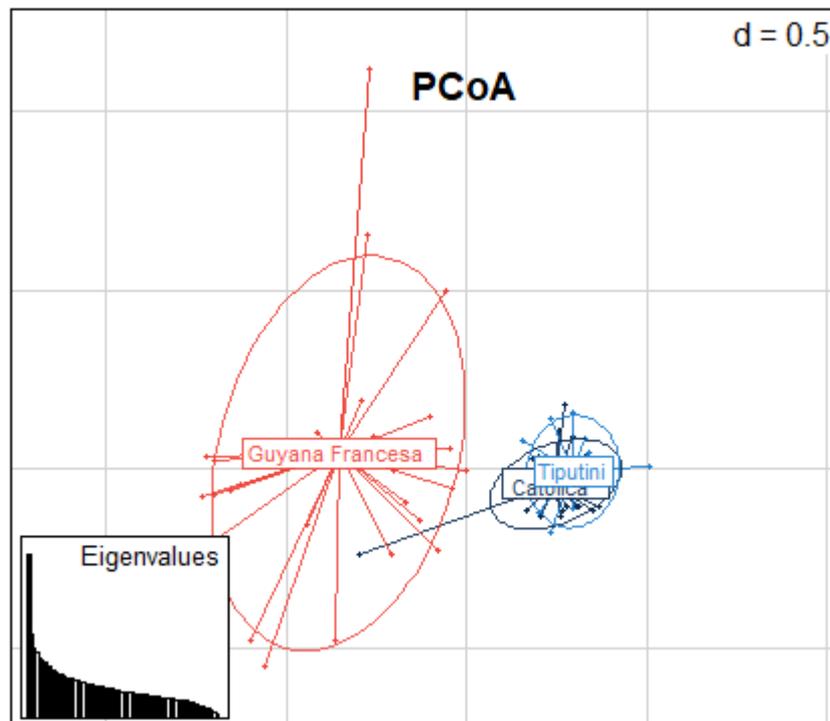


Figura 5. Análisis de Coordenadas Principales (PCoA) entre las zonas de estudio de *Eschweilera*. Representación de la distribución de muestras de *Eschweilera* en Guayana Francesa, Tiputini y Católica.

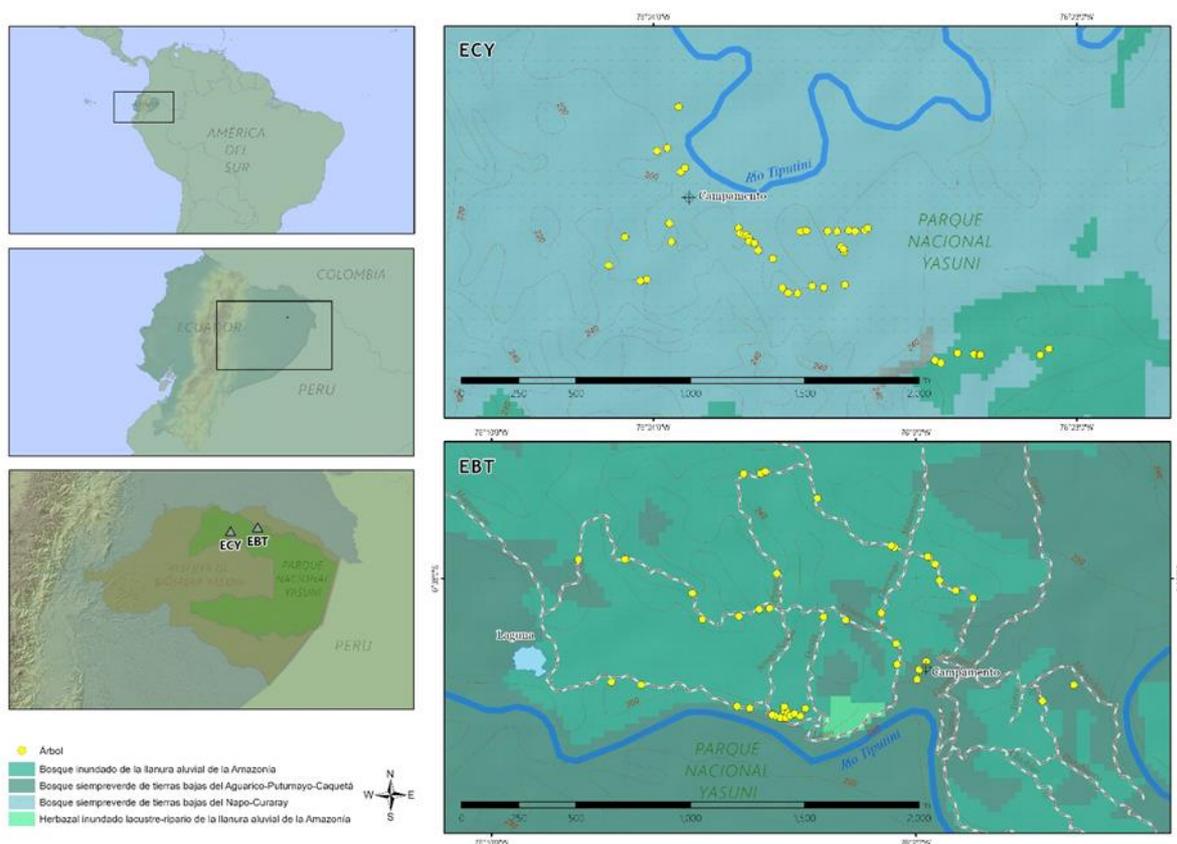
8. REFERENCIAS BIBLIOGRÁFICAS

- Antonelli, A., Zizka, A., Carvalho, F. A., Scharn, R., Bacon, C. D., Silvestro, D., & Condamine, F. L. (2018). Amazonia is the primary source of Neotropical biodiversity. *Proceedings of the National Academy of Sciences*, *115*(23), 6034–6039.
- Antonelli, A. (2021). The rise and fall of Neotropical biodiversity. *Botanical Journal of the Linnean Society*, *199*(1).
- Balslev, H., Valencia, R., Paz y Miño, G., Christensen, H., Nielsen, I. (1998). Species count of vascular plants in one hectare of humid lowland forest in Amazonian Ecuador. *Man and the Biosphere Series*, *21* 585-594.
- Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., ... Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, *22*(3), 148–155.
- Brown, A. H. D., & Young, A. G. (2000). Genetic diversity in tetraploid populations of the endangered daisy *Rutidosis leptorrhynchoides* and implications for its conservation. *Heredity*, *85*(2), 122–129.
- Caron, H., Molino, J., Sabatier, D., Léger, P., Chaumeil, P., Scotti-Saintagne, C., ... Petit, R. J. (2019). Chloroplast DNA variation in a hyperdiverse tropical tree community. *Ecology and Evolution*, *9*(8), 4897–4905.
- Cavers, S., Telford, A., Arenal Cruz, F., Pérez Castañeda, A. J., Valencia, R., Navarro, C., ... Vendramin, G. G. (2013). Cryptic species and phylogeographical structure in the tree *Cedrela odorata* L. throughout the Neotropics. *Journal of Biogeography*, *40*(4), 732–746.
- Cazco Ordoñez, D. C. (2018). Determinación de la diversidad genética y estructura poblacional del guayabillo (*Psidium galapageium*) en la Isla Isabela, Galápagos, Ecuador. Repositorio.usfq.edu.ec. Obtenido de: <https://repositorio.usfq.edu.ec/handle/23000/7969>
- Draper, F. C., Asner, G. P., Coronado, E. N. H., Baker, T. R., García-Villacorta, R., Pitman, N. C. A., ... Del Aguila Pasquel, J. (2019). Dominant tree species drive beta diversity patterns in western Amazonia. *Ecology*, *100*(4), 1–14.
- Earl, D. A., & vonHoldt, B. M. (2011). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359–361.
- Fauset, S., Johnson, M. O., Gloor, M., Baker, T. R., Monteagudo M., A., Brien, R. J. W., ... Allie, E. (2015). Hyperdominance in Amazonian forest carbon cycling. *Nature Communications*, *6*(1).
- Funk, V., Hollowell, T., Berry, P., Kelloff, C., & Alexander, S. N. (2007). Checklist of the Plants of the Guiana Shield (Venezuela: Amazonas, Bolívar, Delta Amacuro; Guyana, Surinam, French Guiana). *Contributions from the United States National Herbarium*, *55*, 1–584.
- Gonzalez, M. A., Baraloto, C., Engel, J., Mori, S. A., Pétronelli, P., Riéra, B., ... Chave, J. (2009). Identification of Amazonian Trees with DNA Barcodes. *PLoS ONE*, *4*(10), e7483.

- Goodwin, Z. A., Harris, D. J., Filer, D., Wood, J. R. I., & Scotland, R. W. (2015). Widespread mistaken identity in tropical plant collections. *Current Biology*, 25(22), R1066–R1067.
- Greenbaum, G., Templeton, A. R., Zarmi, Y., & Bar-David, S. (2014). Allelic Richness following Population Founding Events – A Stochastic Modeling Framework Incorporating Gene Flow and Genetic Drift. *PLoS ONE*, 9(12), e115203.
- Heuertz, M., Caron, H., Scotti-Saintagne, C., Pétronelli, P., Engel, J., Tysklind, N., ... Budde, K. B. (2020). The hyperdominant tropical tree *Eschweilera coriacea* (Lecythidaceae) shows higher genetic heterogeneity than sympatric *Eschweilera* species in French Guiana. *Plant Ecology and Evolution*, 153(1), 67–81.
- Hoogenboom, J., van der Gaag, K. J., de Leeuw, R. H., Sijen, T., de Knijff, P., & Laros, J. F. J. (2017). FDSTools: A software package for analysis of massively parallel sequencing data with the ability to recognise and correct STR stutter and other PCR or sequencing noise. *Forensic Science International: Genetics*, 27, 27–40.
- Huang, Y.-Y., Mori, S. A., & Kelly, L. M. (2015). Toward a phylogenetic-based Generic Classification of Neotropical Lecythidaceae—I. Status of *Bertholletia*, *Corythophora*, *Eschweilera* and *Lecythis*. *Phytotaxa*, 203(2), 85. <https://doi.org/10.11646/phytotaxa.203.2.1>
- Hughes, C. E., Pennington, R. T., & Antonelli, A. (2012). Neotropical Plant Evolution: Assembling the Big Picture. *Botanical Journal of the Linnean Society*, 171(1), 1–18.
- Jinga, P., & Ashley, M. V. (2017). A mountain range is a strong genetic barrier between populations of *Afzelia quanzensis* (pod mahogany) with low genetic diversity. *Tree Genetics & Genomes*, 14(1).
- Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15(5), 1179–1191.
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359.
- Lepais, O., Chancerel, E., Boury, C., Salin, F., Manicki, A., Taillebois, L., ... Guichoux, E. (2020). Fast sequence-based microsatellite genotyping development workflow. *PeerJ*, 8, e9085.
- Lepais, O., & Bacles, C. F. E. (2011). Comparison of random and SSR-enriched shotgun pyrosequencing for microsatellite discovery and single multiplex PCR optimization in *Acacia harpophylla* F. Muell. Ex Benth. *Molecular Ecology Resources*, 11(4), 711–724.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1), 10.
- Mori, S., Kiernan, E., Smith, N., Kelley, L., Huang, Y.-Y., Prance, G., & Thiers, B. (2016). Observations on the phytogeography of the Lecythidaceae clade (Brazil nut family). *Phytoneuron* 2017-30: 1-85.
- Mori, S., & Nadja, C. (1995). The Lecythidaceae of a Central Amazonian Moist Forest. *Memoirs of the New York Botanical Garden*
- Mori, S., & Prance, C. (1990). Lecythidaceae, Part 2. The Zygomorphic-Flowered New World Genera (*Couroupita*, *Corythophora*, *Bertholletia*, *Couratari*, *Eschweilera*, &

- Lecythis), With a Study of Secondary Xylem of Neotropical Lecythidaceae. *Flora Neotropical*, 20(2) 1-373.
- Morrone, J. (2014). Biogeographical regionalisation of the Neotropical region. *Zootaxa*, 3782(1), 1.
- Parisod, C., Holderegger, R., & Brochmann, C. (2010). Evolutionary consequences of autopolyploidy. *New Phytologist*, 186(1), 5–17.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, 155(2), 945–959. <https://doi.org/10.1093/genetics/155.2.945>
- Quinlan, A. R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Current Protocols in Bioinformatics*, 47(1), 11.12.1–11.12.34.
- R studio team. (2015). RStudio: Integrated Development Environment for R. Boston, MA. Obtenido de <http://www.rstudio.com/>
- ter Steege, H., Pitman, N. C. A., Sabatier, D., Baraloto, C., Salomao, R. P., Guevara, J. E., ... Terborgh, J. (2013). Hyperdominance in the Amazonian Tree Flora. *Science*, 342(6156), 1243092–1243092.
- Twyford, A. D., & Ennos, R. A. (2011). Next-generation hybridization and introgression. *Heredity*, 108(3), 179–189.
- Urquhart, G. (2020). The Neotropical Rainforest. In *Encyclopedia of the World's Biomes* (pp. 56–65). Elsevier.
- Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2013). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, 30(5), 614–620.
- Zunino, M., & Zullini, A. (2004). *Biogeografia : la dimensione spaziale dell'evoluzione*. Milano: Cea. <https://doi.org/10.1093/bioinformatics/btt593>

ANEXO 1 : MAPA DE LAS ZONAS DE EXTRACCIÓN DEL MATERIAL FOLIAR DE *ESCHWEILERA* EN ECUADOR



Descripción: El mapa muestra las zonas de colecta de muestras determinadas en la Estación Científica Yasuní (ECY) y la Estación de Biodiversidad Tiputini (EBT) dentro de Ecuador.