EMMANUEL IFEANYICHUKWU UGWOR

MOLECULAR INSIGHTS TO THE GENEALOGICAL RELATIONSHIPS IN THE GENUS *Ceiba* (MALVACEAE)

Dissertation presented to the Federal University of Viçosa, as part of the requirements of the Graduate Program in Applied Biochemistry, to obtain the title of *Magister Scientiae*.

Advisor: Luiz Orlando de Oliveira

VIÇOSA - MINAS GERAIS

2023

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T

U27m
2023

Ugwor, Emmanuel Ifeanyichukwu, 1995-
Molecular insights to the genealogical relationships in the genus *Ceiba* (Malvaceae) / Emmanuel Ifeanyichukwu Ugwor. – Viçosa, MG, 2023.
1 dissertação eletrônica (48 f.): il. (algumas color.).

Texto em inglês.
Inclui apêndice.
Orientador: Luiz Orlando de Oliveira.
Dissertação (mestrado) - Universidade Federal de Viçosa, Departamento de Bioquímica e Biologia Molecular, 2023.
Referências bibliográficas: f. 36-39.
DOI: https://doi.org/10.47328/ufvbbt.2023.107 Modo de acesso: World Wide Web.

1. Plantas - Variação. 2. Marcadores genéticos. 3. Ceiba. 4. Malvaceae. I. Oliveira, Luiz Orlando de, 1961-. II. Universidade Federal de Viçosa. Departamento de Bioquímica e Biologia Molecular. Programa de PósGraduação em Bioquímica Aplicada. III. Título.

CDD 22. ed. 581.3

Bibliotecário(a) responsável: Euzébio Luiz Pinto CRB-6/3317

EMMANUEL IFEANYICHUKWU UGWOR

MOLECULAR INSIGHTS TO THE GENEALOGICAL RELATIONSHIPS IN THE GENUS *Ceiba* (MALVACEAE)

Dissertation presented to the Federal University of Viçosa, as part of the requirements of the Graduate Program in Applied Biochemistry, to obtain the title of *Magister Scientiae*.

APPROVED: March 28, 2023**.**

Assent:

Emmanuel Ifeanyichukwu Ugwor
(Author)

Documento assinado digitalmente
LUIZ ORLANDO DE OLIVEIRA
Data: 22/07/2024 12:29:28-0300
Verifique em https://validar.iti.gov.br

Luiz Orlando de Oliveira
Advisor

To God Almighty, the creator of everything
seen and unseen, and to my parents
for their unwavering belief.

ACKNOWLEDGEMENTS

To Almighty God, my Heavenly Father, for the inspiration, ideas and provisions not just during the course of this research work, but also throughout my life.

To my parents, Mr. C. and Mrs. R. Ugwor, who by sacrificing their needs for my wants and instilling in me the fear of God, made my life a continuous testimony.

To the Nigerian Tertiary Trust Fund (TETFund) and the Forum for Agricultural Research in Africa (FARA) for initiating the Agricultural Research and Innovation Fellowship for Africa (ARIFA), which sponsored my master's program.

To the Graduate Program in Applied Biochemistry, Federal University of Viçosa for the opportunity to carry out this master's program.

This work was carried out with the support of the Coordination for the Improvement of Higher Education Personnel – Brazil (CAPES) - Financing Code 001. To the National Council for Scientific and Technological Development (CNPq) and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) for providing financial support for this research.

To the Herbarium of Embrapa Cernagem (CEN) and Herbarium of the Federal University of Mato Grosso do Sul Foundation (CGMS) for the provision of samples.

To the Institute of Biotechnology Applied to Agriculture (BIOAGRO) for the provision of space to carry out this research work.

To my advisor, Professor Luiz Orlando de Oliveira, for his guidance, contributions, time, and support throughout the course of my program and beyond.

To my colleagues at the Molecular Biology and Phylogeography Lab: Vinicius, Samara, Leandro, Yan. I appreciate the camaraderie and support.

To the family ARIFA gave me, Henry, Vivian, Shafiu, Caio, Ola, Deji, Abimbola, and Ejima, to mention a few. May God continue to keep you, until we meet again.

ABSTRACT

UGWOR, Emmanuel Ifeanyichukwu, M.Sc., Universidade Federal de Viçosa, March, 2023. **Molecular insights to the genealogical relationships in the genus *Ceiba* (Malvaceae)**. Adviser: Luiz Orlando de Oliveira.

The geographical and morphological diversity among the congeners of *Ceiba* makes the genus a suitable model to appraise the evolutionary dynamics of arboreal flora in the Neotropical realm. This study sought to investigate the genealogical relationships of the different species within the genus *Ceiba.* Sequences of the nuclear internal transcribed spacer (*ITS*) and chloroplast gene (t*rn*T-*trn*L, intron *trn*L, *trn*S-*trn*G, *mat*K, and *trn*H-*psb*A) regions from 15 congeners of *Ceiba* were analysed using Bayesian inference and Median Joining Network, with *Pseudobombax* as the outgroup. Results indicate that *Ceiba* and *Pseudobombax* were separated into two distinct lineages. However, based on *ITS* data alone, the genus *Ceiba* is not monophyletic, with specimens of *Ceiba crispiflora* and *Ceiba ventricosa* included in the outgroup, most likely due to incomplete lineage sorting or a hybridisation event. The chloroplast genes-based tree recovered *Ceiba* as monophyletic, but the phylogenetic relationships were unclear as the tree exhibited relatively low support (posterior probability < 90%) for all major nodes. Phylogenetic analysis of the combined *ITS* and chloroplast genes dataset produced a better resolution of the relationships within *Ceiba*, with the congeners divided into three main clades. The first clade brought together specimens of *Ceiba speciosa, Ceiba samauma,* and *Ceiba pentandra,* while the second comprised specimens collected outside Brazil (*Ceiba acuminata, Ceiba aesculifolia,* and *Ceiba schottii*). The third clade was the largest, most diversified, and likely originated from the Atlantic Forest, transitioning into the Cerrado and Mesoamerican subregions. The consensus tree (from the combined dataset) presented a pattern of phylogeographic congruence, in which the phylogenetic placement of the members of distinct clades and subclades paralleled the geographic distribution of the specimens. Our results provide further insights into the genealogical relationships among *Ceiba* species and contribute to the existing knowledge on the processes maintaining the Neotropical diversity.

Keywords: *Ceiba;* Malvaceae; Neotropics; Phylogenetics; Phylogeography

RESUMO

UGWOR, Emmanuel Ifeanyichukwu, M.Sc., Universidade Federal de Viçosa, março de 2023. **Percepções moleculares para as relações genealógicas no gênero *Ceiba* (Malvaceae)**. Orientador: Luiz Orlando de Oliveira.

A diversidade geográfica e morfológica entre os congêneres de *Ceiba* torna o gênero um modelo adequado para avaliar a dinâmica evolutiva da flora arbórea na região Neotropical. Este estudo buscou investigar as relações genealógicas das diferentes espécies dentro do gênero *Ceiba.* Sequências das regiões do espaçador transcrito interno nuclear (*ITS*) e do gene do cloroplasto (*trn*T- *trn*L, intron *trn*L, *trn*S -*trn*G, *mat*K e *trn*H-*psb*A) de 15 congêneres de *Ceiba* foram analisadas usando Inferência Bayesiana e Median Joining Network, com *Pseudobombax* como grupo externo. Os resultados indicam que *Ceiba* e *Pseudobombax* foram separados em duas linhagens distintas. No entanto, com base apenas nos dados *do ITS*, o gênero *Ceiba* não é monofilético, com espécimes de *Ceiba crispiflora* e *Ceiba ventricosa* incluídos no grupo externo, provavelmente devido à classificação incompleta da linhagem ou a um evento de hibridação. A árvore baseada em genes de cloroplasto recuperou *Ceiba* como monofilética, mas as relações filogenéticas não eram claras, pois a árvore exibia suporte relativamente baixo (probabilidade posterior <90%) para todos os nós principais. A análise filogenética do conjunto de dados combinados *de ITS* e genes de cloroplasto produziu uma melhor resolução das relações dentro de *Ceiba*, com os congêneres divididos em três clados principais. O primeiro clado reuniu espécimes de *Ceiba speciosa, Ceiba samauma* e *Ceiba pentandra,* enquanto a segunda compreende espécimes coletados fora do Brasil (*Ceiba acuminata, Ceiba aesculifolia* e *Ceiba schottii*). O terceiro clado foi o maior, mais diversificado e provavelmente originado da Mata Atlântica, fazendo a transição para as sub-regiões do Cerrado e da Mesoamérica. A árvore de consenso (do conjunto de dados combinados) apresentou um padrão de congruência filogeográfica, em que a colocação filogenética dos membros de clados e subclados distintos é paralela à distribuição geográfica dos espécimes. Nossos resultados fornecem mais informações sobre as relações genealógicas entre as espécies *de Ceiba* e contribuem para o conhecimento existente sobre os processos de manutenção da diversidade Neotropical.

Palavras-chave: *Ceiba;* Malvaceae; Neotrópicos; Filogenética; Filogeografia

LIST OF ILLUSTRATIONS

LIST OF TABLES

SUMMARY

1. INTRODUCTION

The Neotropics are one of the world's major biogeographic realms, encompassing the tropical land ecoregions of the Americas and the whole South American temperate region. This extensive realm spans South America, Central America, the Caribbean islands, and the southern reaches of North America, as corroborated by authoritative sources (Urtubey *et al.*, 2010; Antonelli *et al.*, 2018). Renowned for its unparalleled richness in species and distinguished as one of the most biodiverse areas on Earth, the Neotropics emerge as a complex mosaic of ecological diversity and evolutionary dynamism. The geological and climatic vicissitudes that have characterized the Neotropics throughout evolutionary epochs have given rise to an array of distinct biomes, ranging from arid deserts in the Andean subregion to luxuriant tropical rainforests in the Amazon basin (Antonelli *et al.*, 2018). This remarkable heterogeneity is a testament to the historical interplay of diverse ecological factors and dynamic environmental forces shaping the landscape of the Neotropics. Furthermore, ecological fluctuations, transitions, climatic variations, the pace of diversification within and across specific clades, and other mechanisms have ensured the biogeographical dynamism of the Neotropics across evolutionary time. Despite the wealth of knowledge accumulated about the Neotropics, the precise mechanisms underpinning the genesis and maintenance of its extraordinary diversity remain shrouded in ambiguity (Rull, 2020; Lozano-García *et al.*, 2021).

Over time, numerous attempts have been undertaken to explain the history and dynamics of Neotropical diversity using model organisms, including insects (Espeland *et al.*, 2015), birds (Derryberry *et al.*, 2011), and plants (Leal *et al.*, 2016). Insects, with their vast diversity, complex life histories, and sensitivity to environmental conditions, have served as insightful models for testing biogeographical hypotheses related to Neotropical diversification (Espeland *et al.*, 2015). Avian groups have also been instrumental, as birds are conspicuous organisms that readily disperse across geographical barriers. Molecular phylogenetic studies of Neotropical bird lineages, such as tanagers, hummingbirds, and antbirds, have revealed the importance of ecological gradients, geological dynamics, and intrinsic biological traits (e.g., dispersal ability) in promoting avian diversification (Derryberry *et al.*, 2011). However, plants have proven to be an exceptionally useful system for deciphering the evolutionary history (Soltis and Soltis, 2021). Angiosperms exhibit a remarkable variety of morphological features, particularly in their elaborate flowers, providing abundant phenotypic discontinuities between species. Floral traits such as petal shape, stamen position, and pollen morphology can be readily examined and quantified for systematic comparisons across taxa. The discreteness of these characteristics facilitates species delimitation and inferences about evolutionary relationships. Furthermore, key innovations in floral structures, such as zygomorphy and specialized pollination syndromes, directly influenced angiosperm diversification rates in Neotropical lineages (Leal *et al.*, 2016).

The genus *Ceiba* Mill. (Malvaceae) is one of the most prominent plants in the Neotropics. Other important members of the Malvaceae family include *Pseudobombax* and *Cavanillesia*. *Ceiba* is an integral part of Mesoamerican culture and is native to tropical West Africa. Although most *Ceiba* species are largely understudied, they have been posited as a model tree for conservation and restoration efforts in Africa and Asia (Gómez-Maqueo and Gamboa-deBuen, 2022). The genus *Ceiba* comprises 18 recognised species, naturally distributed in the Neotropics, including *C. aesculifolia, C. boliviana, C. chodatii, C. crispiflora, C. erianthos, C. glaziovii, C. insignis, C. jasminodora, C. lupuna, C. pentandra, C. pubiflora, C. rubriflora, C. samauma, C. schottii, C. soluta, C. speciosa, C. trischistandra,* and *C. ventricosa* (Gómez-Maqueo and Gamboa-deBuen, 2022). The species of the genus *Ceiba* present remarkable morphological diversity. The species have digitate leaves; most are deciduous and flower when leafless. Many species have sharp tips on their trunks, while others, such as *Ceiba chodatii*, *Ceiba glaziovii, Ceiba pubiflora*, and *Ceiba speciosa,* have swollen trunks. Even individuals of the same species may present variations relating to tips and trunk form. Plant heights vary widely across species, ranging from 2 m in *Ceiba jasminodora* (found in Minas Gerais, Brazil) to 50 m in *Ceiba pentandra* (from the Amazon) (Gibbs and Semir, 2002). The fruits *Ceiba spp* are typically woody capsules with a modified endocarp containing long, tubular trichomes, termed “kapok” (Gómez-Maqueo and Gamboa-deBuen, 2022). Petal colour is another source of morphological diversity in this genus. It varies from white-pink (in *Ceiba boliviano*), dense pink-lilac (in *Ceiba pubiflora*), white (in *Ceiba glaziovii* and *Ceiba erianthos*), ivory-pale yellow (in *Ceiba chodatii* and *Ceiba lupuna*), to magenta (in *Ceiba speciosa* and *Ceiba crispiflora*) (Gibbs and Semir, 2002).

Besides these morphological differences, phylogeographical differences are evident within the genus *Ceiba*. For example, while most species occur predominantly in seasonally dry Neotropical forests, *Ceiba lupuna* is limited to rainforests. Further, *Ceiba erianthos* and *Ceiba jasminodora* are nearly always found on the rocky outcrops of Brazil’s Espinhaço mountain range. *Ceiba pubiflora* prefers calcareous soils in its range, which extends from Mato Grosso Sul to the caatingas of northern Minas Gerais and Bahia. In contrast, some species, such as *Ceiba samauma* and *Ceiba speciosa*, are distributed in dry seasonal forests and humid river valleys, whilst *Ceiba pentandra* has an even higher tolerance, extending to the seasonally flooded lowland Amazon Forest and mesic subregions in Central America(Gibbs and Semir, 2002; Pezzini *et al.,* 2021).*Ceiba pentandra* is the only species that exists outside of the Neotropics, occurring in West Africa, where it is most likely native, India, Southeast Asia, and the Pacific, where man most likely introduced it (Bayer and Kubitzki, 2003).

This diversity in both geographic distribution and morphological features of the genus *Ceiba* epitomises the Neotropical diversity, making the genus a suitable model to appraise the evolutionary dynamics of arboreal flora in Neotropical biomes. Indeed, several attempts have been made to understand the taxonomy and diversity within the genus, starting as far back as 1961 (Robyns, 1961). Gibbs and colleagues proposed a taxonomic revision of thegenus *Ceiba*; however, they used only morphological characters, such as calyx, staminal tube, floral, leaves, corolla, androecium, fruit, and seed attributes (Gibbs and Semir, 2002). However, as the authors noted, the morphological characterization of *Ceiba* presents certain issues. For example, many members of the genus blossom without leaves; hence, specimens studied were either of leaves or floral parts, and seldom both tissues. Furthermore, the species present fleshy flowers with delicate colour changes that, if not dried correctly, may lose their colour and shape (Gibbs and Semir, 2002). These limitations raise concerns about the accuracy of their taxonomic revision. In another study by Carvalho-Sobrinho *et al.* (2016), Bayesian phylogenetic analyses of the sequences obtained from the nuclear ribosomal transcribed spacers and plastid markers from 13 *Ceiba* species resolved the genus as monophyletic and a close relative to *Pseudobombax*, forming a well-supported “striated bark” lineage. However, the relationships within *Ceiba* were poorly resolved, particularly as only one specimen per species was included in the consensus tree. A more recent attempt at evaluating the phylogeny of the genus *Ceiba* used only nuclear DNA characters (namely the internal transcribed spacer – *ITS* – region) for their phylogenic evaluation and forfeited the use of chloroplast informative characters (Pezzini *et al.* (2021). Thus, despite recent efforts to elucidate the evolutionary and biogeographic history of the genus *Ceiba*, these contentions remain (Gómez-Maqueo and Gamboa-deBuen, 2022).

Before the advancement of genetic technologies, comparing discontinuous morphological characters, such as the form and colour of reproductive parts, was the primary method of investigating evolutionary dynamics. Recently, phylogenetics has become increasingly applied in ecology and evolution (Pires and Dolan, 2012). Several phylogenetic and phylogeographic studies based on molecular data have proven essential to understanding evolutionary dynamics and ecological diversity, contributing to species taxonomy and genetic conservation (Espeland *et al.*, 2015; Carvalho-Sobrinho *et al.*, 2016; Garcia *et al.,* 2021; Pezzini *et al.*, 2021).Thus, the current study sought to explore the genealogical relationships among species within the genus *Ceiba* using DNA sequence data obtained from specimens within the genus range in Brazil and outside Brazil. In doing so, we addressed the following questions: (1) Do the ITS and chloroplast genes regions detain enough phylogenetic signals to resolve the genealogical relationships among congeners of Ceiba? (2) Is the genus Ceiba monophyletic, with respect to the phylogenetically closely related genus Pseudobombax? (3) What genealogical lineages make up the genus Ceiba? What are the genealogical relationships among these lineages? (4) How are these lineages distributed geographically? By addressing these questions, this study should shed light on the biogeography and evolutionary history of *Ceiba*. This will, in turn, provide insights into events that shaped and sustained the diversification in the Neotropical realm.

2. MATERIALS AND METHODS

2.1 Sampling strategy and material collection

This study included 15 species of the genus *Ceiba* out of the 18 recognised species within the genus (Table 1). Plant materials were sampled during field expeditions or collected from herbaria. A total of 23 specimens were collected, while sequence information for five *Ceiba* species (*Ceiba pentandra, Ceiba acuminata, Ceiba aesculifolia, Ceiba crispiflora* and *Ceiba schottii*) were retrieved from Duarte *et al.* (2011) (see Table 1). Sampling sites were within the geographic distribution of Ceiba in the Neotropics and included twelve locations in Brazil, two in Mexico, and one in Cuba (Figure 1). The coordinates of each specimen sampled during the field expeditions were recorded using a Global Positioning System (GPS) receiver. For herbarium specimens, the data submitted by the collector was used to estimate the corresponding coordinates. Furthermore, three congeners (*Pseudobombax longiflorum, Pseudobombax manginatum,* and *Pseudobombax croizatti*) of *Pseudobombax* (5 specimens in total) were included as outgroups (Table 1). The outgroup choice was based on a previous study that pointed out *Pseudobombax* as the phylogenetically closely related genus to *Ceiba* (Carvalho-Sobrinho *et al.,* 2016). Plant materials (leaf tissues) collected in the field were dried with silica gel and kept at room temperature in a zip-lock plastic bag for genomic DNA extraction. Herbarium specimens were stored similarly.

2.2 Geographical distribution of Ceiba

Information on the geographical distribution of the congeners of *Ceiba* was retrieved from *species*Link (*Sistema de Informação Distribuído para Coleções Biológicas*; http://splink.cria.org.br/). The *species*Link online server provides free access to information and tools to enhance conservation and sustainable use of biodiversity via research, education, and policy making. The server contains information on specimens deposited in herbariums within South America. The botanical name of each congener of *Ceiba* used in this study was imputed into the *species*Link server, and the map coordinates (longitudes and latitudes) of every specimen deposited for that species were downloaded. The coordinates were converted to degrees, minutes, and seconds (DMS) and then merged with the GTM file of South America (Morrone, 2014) within the GPS TrackMaker software (Version 13.9.608). The resulting map for each species was saved.

Table 1: Specimens of the genus *Ceiba* and *Pseudobombax* used in this study.

S/N	Species	Specimen/Herbarium No	Location
1	*Ceiba glaziovii*	1500	Cruz, Ceará, Brazil
2	*Ceiba glaziovii*	1510	Cruz, Ceará, Brazil
3	*Ceiba pubiflora*	1706	Cruz, Ceará, Brazil
4	*Ceiba pubiflora*	1743	Montes Claros, Minas Gerais, Brazil
5	*Ceiba pubiflora*	1795	Peruaçu, Minas Gerais, Brazil
6	*Ceiba rubriflora*	1881	Peruaçu, Minas Gerais, Brazil
7	*Ceiba rubriflora*	1855	Lajedo, Pernambuco, Brazil
8	*Ceiba jasminodora*	1904	Itacarambi, Minas Gerais, Brazil
9	*Ceiba jasminodora*	1908	Serra do cabral, Minas Gerias, Brazil
10	*Ceiba samauma*	1921	Serra do cabral, Minas Gerias, Brazil
11	*Ceiba speciosa*	1930	Alta Floresta, Mato Grosso, Brazil
12	*Ceiba speciosa*	1934	Alta Floresta, Mato Grosso, Brazil
13	*Ceiba samauma*	1941	Para, Brazil
14	*Ceiba crispiflora*	1958	Bahia, Brazil
15	*Ceiba erianthos*	1960	JBRM, Rio de Janeiro, Brazil
16	*Ceiba erianthos*	1965	JBRM, Rio de Janeiro, Brazil
17	*Ceiba jasminodora*	1968	JBRM, Rio de Janeiro, Brazil
18	*Ceiba lupuna*	1969	JBRM, Rio de Janeiro, Brazil
19	*Ceiba pentandra*	1970	JBRM, Rio de Janeiro, Brazil
20	*Ceiba ventricosa*	1983	JBRM, Rio de Janeiro, Brazil
21	*Ceiba erianthos*	1985	Tijuca, Rio de Janeiro, Brazil
22	*Ceiba speciosa*	3218	Sonora, Mexico
23	*Ceiba insignis*	3216	Sonora, Mexico
24	*Ceiba pentandra**	Alverson s.n.	-
25	*Ceiba acuminata**	Fairchild Botanical Gardens (FBG) X-2-206	Sonora, Mexico
26	*Ceiba aesculifolia**	FBG 83301	Soledad, Cienfuegos, Cuba
27	*Ceiba schottii**	FBG 83302	Acapulco, Mexico
28	*Ceiba crispiflora**	Pacific Tropical Garden Acc. No: 750726001	-
29	*Pseudobombax manginatum*	73501	Bodoquena, Mato Grosso do Sul, Brazil
30	*P. manginatum*	CEN 118802	Nobres, Mato Grosso, Brazil
32	*P. longiflorum*	CEN107560	Dianopolis, Tocantins, Brazil
32	*P. longiflorum*	74270	Campo Grande, Mato Grosso do Sul, Brazil
33	*P. croizatti**	Oldham s.n.	

**ITS*, *intron-trnL*, and *matK* sequences for *Ceiba pentandra, Ceiba acuminata, Ceiba aesculifolia, Ceiba schottii, Ceibacrispiflora*, and *P. croizatti* were retrieved from Duarte *et al.* (2011). - = undefined location

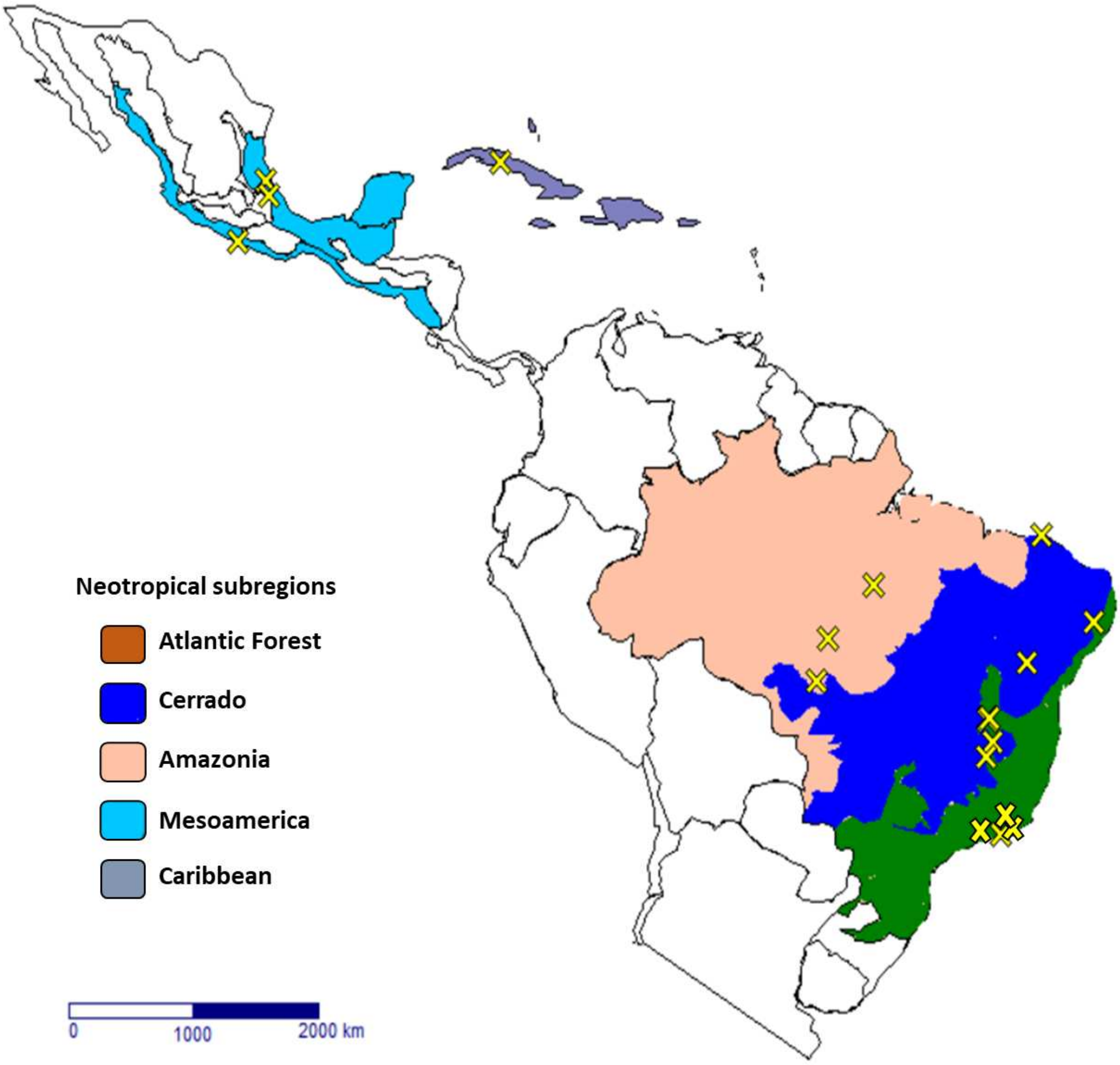


Figure 1: Geographical distribution, across the Neotropics, of the sampling sites (marked X) for the specimens of genus *Ceiba* included in this study.

2.3 Extraction of genomic DNA

Total DNA was extracted from leaf tissue using the cetyltrimethylammonium bromide (CTAB) method reported by Cota-Sánchez et al. (2006) and modified by Aboul-Maaty et al. (2019). About 50 mg of the sample was homogenized in liquid nitrogen with a pre-chilled mortar and pestle. Immediately, 800 μL of 3× CTAB extraction of buffer (containing 1.4 M NaCl, 0.1 M Tris-HCl pH 8.0, 20 mM EDTA pH 8.0, and 0.3% 2-β-mercaptoethanol; pre-heated to 65 °C) was added. The resulting mixture was incubated in a water bath at 65 °C for an hour, with gentle inversion at periods. After that, an equal volume of chloroform/isoamyl alcohol (24:1 v/v) was added and mixed by gentle inversions, followed by centrifugation at 13,000 rpm for 15 minutes (at 25 °C). This step was repeated until the upper aqueous layer was transparent (typically twice). An aliquot of 6M NaCl was added to the aqueous phase (1:2) and thoroughly mixed, followed by 3 M potassium acetate (1:10) and 500 μL of ice-cold isopropyl alcohol. DNA was allowed to precipitate at −20 °C for 30 minutes and then washed with 500 μL of ethanol (70% v/v). The DNA pellets were air-dried at room temperature and reconstituted with 30 μL of 1× Tris-EDTA buffer containing RNAse (10 mg/mL). The extracted DNA's integrity was verified using horizontal electrophoresis on a 1% agarose gel stained with ethidium bromide. A NanoDrop 2000 spectrophotometer was used to quantify the DNA concentration.

2.4 Gene amplification and sequencing

Five chloroplast genes (*trnT-trnL, intron trnL, trnS-trnG, matK*, and *trnH-psbA*) and the *ITS* region (including the 5.8S gene) were amplified using the polymerase chain reactions (PCR), with the gene-specific primer pairs (Table 2). The amplifications of the *ITS* region were carried out in a final volume of 20 μL with 1X Go Taq Green buffer (Promega), 1.75 μL of DMSO (dimethyl sulfoxide), 0.25 mM of each dNTP, 2.5 mM $MgCl_2$, 0.62 μM of each primer, 1.25 U of Taq DNA polymerase (Phoneutria) and approximately 100 ng of genomic DNA. Samples were amplified under the following conditions: 94°C for 4 minutes (initial denaturation cycle), 35 cycles of 94°C for 1 minute, 52°C (primer annealing) for 1 minute, 72°C for 45 seconds and final extension at 72°C for 5 minutes.

For each chloroplast gene region, the amplification reactions were carried out in a final volume of 25 μL with Green 1X Go Taq buffer (Promega), 17.5 μg/mL of bovine serum albumin, 0.2 mM of each dNTP, 2 mM $MgCl_2$; 0.5 μM of each primer, 1.25 U of Taq DNA polymerase (Phoneutria) and approximately 50 ng of genomic DNA. The amplification program for the chloroplast regions consisted of an initial denaturation cycle at 95ºC for 3 minutes, followed by 40 cycles of 95ºC for 1 minute, 54ºC (primer annealing) for 1 minute, 72ºC for 1 minute and a final extension of 72ºC for 10 minutes.

Following amplification, the PCR products of both *ITS* region and chloroplast genes were checked on 1% agarose gel electrophoresis. After that, ExoSAP IT (USB) was used to purify the PCR amplicons (3 µL of enzyme/9 µL of reaction). Amplicons were sequenced with the same primers used in the PCR reactions, using the DNA sequencing services of Macrogen Inc., South Korea (www.macrogen.com).

Table 2: Molecular makers for the amplification of nuclear and chloroplast regions

S/N	**Gene**	**Primers**	**Sequence (5'- 3')**	**Location**
1	*ITS*	ITS_LEU	GTCCACTGAACCTTATCATTTAG	Nuclear
		ITS_4	TCCTCCGCTTATTGATATGC	
2	*trnT-trnL*	A2	CAAATGCGATGCTCTAACCT	Chloroplast
		aC	CGTAGCGTCTACCGATTTCG	
3	*intron-trnL*	C_trnT	CGAAATCGGTAGACGCTACG	Chloroplast
		D_trnL	GGGGATAGAGGGACTTGAAC	
4	*trnS-trnG*	trn_S	GCUGCCGCTTTAGTCCACTCAGC	Chloroplast
		trn_G	GAACGAATCACACTTTTACCAC	
5	*matK*	matK_1R	ACCCAGTCCATCTGGAAATCTTGGTTC	Chloroplast
		matK_3F	CGTACAGTACTTTTGTGTTTACGAG	
6	*psbA-trnH*	psb_A	GTTATGCATGAACGTAATGCTC	Chloroplast
		trn_H	CGCGCATGGTGGATTCACAATCC	

2.5 Sequencing and datasets assembling

The sequence data were imported into the Sequencher software (version 4.8; Gene Codes Corporation) for inspection, manual editing and adjustments. The sequences were aligned using the software's default algorithm with the ends clipped and gaps introduced to compensate for the occurrence of *indels* (insertion/deletion). In sequences with double peaks of identical height, such sites were regarded as ambiguous and coded as per the International Union of Pure and Applied Chemistry (IUPAC) nucleotide code for ambiguous bases.

Given the distinctive requirements of individual analysis, several datasets were assembled in this study. For phylogeny analyses, three datasets were assembled. The *ITS* sequences publicly available in GenBank for *Ceiba* and *Pseudobombax* were downloaded and assembled with the *ITS* region sequences from this study to produce dataset A (647 bp, n = 87). The sequences of the five chloroplast genes (*trnT-trnL*: 663 bp; *intron-trnL*: 503 bp; *trnS-trnG*: 588 bp; *matK*: 787 bp; and *psbA-trnH*: 335 bp) were concatenated to give rise to dataset B (2878 bp, n = 30). Where available, publicly available GenBank sequences were included to replace missing sequences (see

Appendix). Lastly, the sequences from the *ITS* region and the chloroplast genes were concatenated to produce dataset C (3524 bp, n = 31).

For the haplotype network, the electropherograms of the gene sequences were re-examined manually to further differentiate the intraindividual polymorphism within the sequences. Ambiguous sites (presenting double peaks) or sites with more than two-character states across the dataset were discarded. These sites violate the infinite-sites model, which postulates that the mutation rate within a particular DNA sequence is so low that only a single mutation can occur at any specific homologous site (Kimura, 1969). Such sites would cause homoplasious relationships in the haplotype network. Each *indel*, independent of length, was considered a single mutation and coded as a fifth character. However, *indels* bordered by mononucleotide repeats (e.g., polyA) were excluded. Outgroup sequences were also discarded. Following these adjustments, two databases were generated. Dataset D (606 bp, n = 59) contained the *ITS* region sequences, whereas the chloroplast sequences (*intron-trnL*: 463 bp; *trnS-trnG*: 563 bp; and *matK*: 787 bp) were concatenated to generate dataset E (1813 bp, n = 27). Sequence data are in the process of being deposited in GenBank.

2.6 Data analyses

2.6.1 Phylogeny

Independent phylogenetic analyses were carried out with MRBAYES (version 3.1.2) on datasets A, B, and C, using the Bayesian inference approach (Ronquist and Huelsenbeck, 2003). Before the phylogenetic analyses, the MRMODELTEST software (version 2.3) was used to estimate the best evolution model for each genomic area using the Akaike Information Criterion (Akaike, 1973). The estimated best-fit models were GTR+G (*ITS),* GTR+I (*trnT-trnL*), F81 (*intron-trnL*), F81+I (*trnS-trnG* and *matK*), and HKY (*psbA-trnH*) out of 24 models of molecular evolution assessed. The partitioned MRBAYES analyses were done in two simultaneous runs of 10 million generations, with one cold chain and seven hot chains in each run. Every 1000 tree was sampled, with the first 250 deleted as burn-in samples. After each analysis, the average standard deviation of split frequencies was below 0.01, and the effective size values of samples (ESS) for all statistics analysed in the Tracer software (version 1.5) were higher than 200.

2.6.2 Haplotype Network

Gene genealogies for *ITS* and chloroplast regions sequences were inferred separately using datasets D and E, respectively. The Median Joining network approach (Bandelt *et al.* 1999) was used to infer genealogical relationships among lineages of genus *Ceiba*, as implemented in the Network application (version 5.0.0.3; Fluxus Technology Ltd.). with default parameters. The DnaSP v5 software was used to estimate the gene diversity and nucleotide diversity of the datasets used for haplotype network analyses (Librado and Rozas, 2009).

3. RESULTS

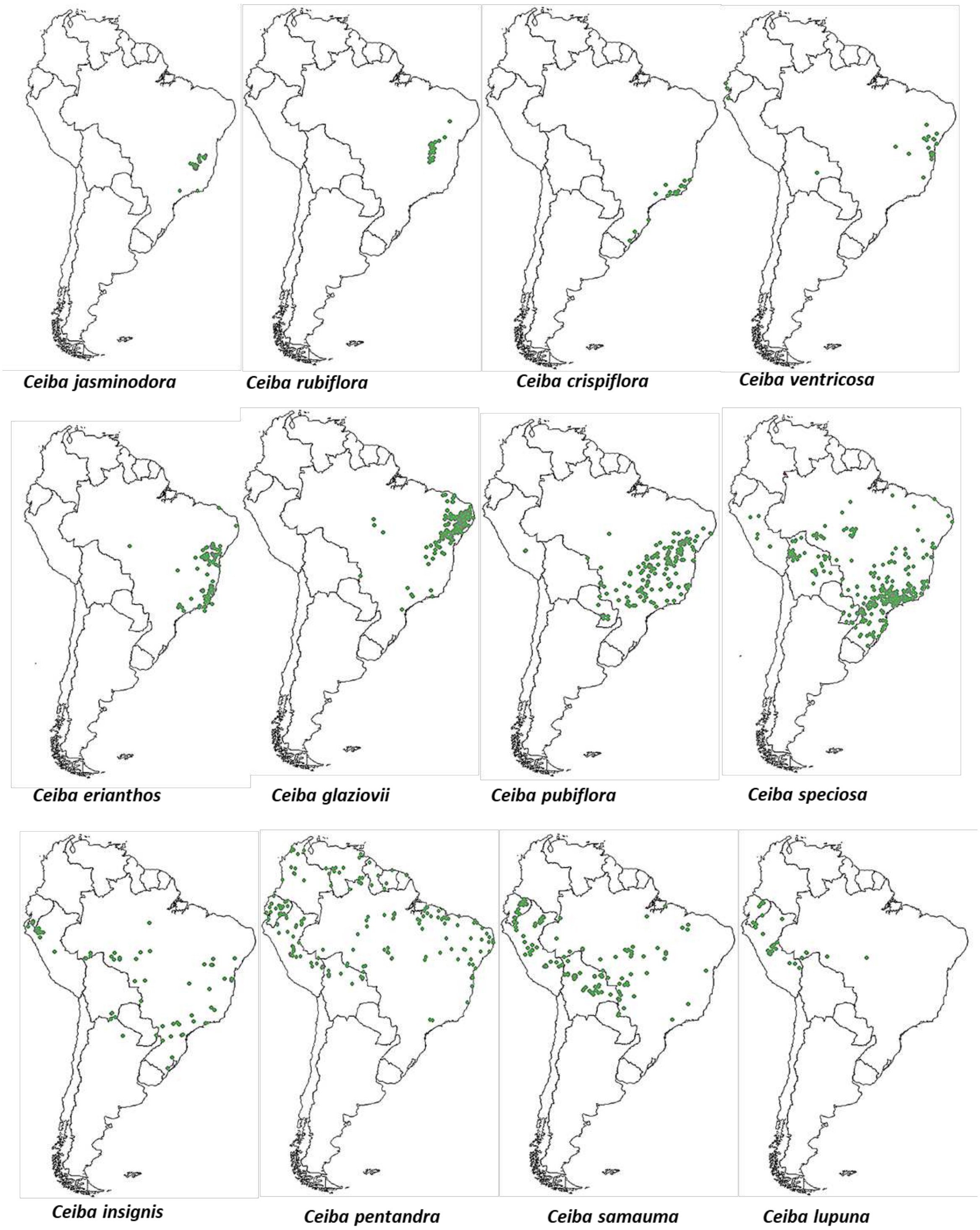
3.1 Geographical distribution of *Ceiba*

The collection coordinates of specimens deposited for the congeners of *Ceiba* were retrieved from *species*Link. Out of the 15 species of *Ceiba* used in this study, two had unspecified locations (*Ceiba schottii* and *Ceiba aesculifolia*), while no records were found for *Ceiba acuminata*. The coordinates of the remaining 12 species were transformed into a map depicting the geographical distribution of each species within South America (Figure 2). Herbarium records of *Ceiba jasminodora, Ceiba rubriflora, Ceiba crispiflora, Ceiba ventricosa,* and *Ceiba erianthos* were distributed across the Atlantic Forest, which stretches from Rio Grande do Norte (Northeast Brazil) to Rio Grande do Sul (Southern Brazil) and inland as far as Paraguay and Argentina's Misiones Province. Records of *Ceiba glaziovii* and *Ceiba pubiflora* were distributed across the Atlantic Forest and Cerrado. Unlike the Atlantic Forest, the Cerrado is a tropical savanna ecoregion (including forest, wooded, park, and gramineous-woody savannas), occupying the eastern states of Brazil: Goiás, Tocantins, Minas Gerais, Mato Grosso do Sul, Mato Grosso, and the Federal District. The spatial distribution of *Ceiba speciosa* and *Ceiba insignis* extended from the Atlantic Forest and Cerrado into the moist broadleaf tropical rainforest of the Amazonia; they were distributed across Brazil, Bolivia, and Peru. Similarly, *Ceiba pentandra* were widely distributed across the South American region, extending towards the Mesoamerica subregion. Specimens of *Ceiba samauma* were sparsely distributed in the Atlantic Forest but densely concentrated within the Amazon. Specimens of *Ceiba lupuna* were confined to the extreme of the Amazon (Figure 2).

3.2 Phylogenetic analyses

The Bayesian analyses generated three phylogenetic trees. There was one tree each for the *ITS* region (Figure 3), chloroplast genes (Figure 4), and combined datasets (Figure 5), respectively. Molecular data from those datasets allowed for the recovery of three consensus trees in which most of the internal nodes had high posterior probability (PP) support (85-100%) for the main clades. The three consensus trees depicted the phylogenetic relationships among the 15 species of *Ceiba*, having the phylogenetically closely-related genus *Pseudobombax* as the outgroup.

The sequences of the *ITS* region of *Ceiba* (n = 63) and *Pseudobombax* (n = 24) were aligned to form dataset A and used to generate a Bayesian phylogenetic tree (Figure 3). Specimens of either *Ceiba* or *Pseudobombax* formed two distinct genus-specific lineages, with only two exceptions – *Ceiba crispiflora* (specimen no.1958) and *Ceiba ventricosa* (specimen no.1983). They instead clustered together with specimens of the outgroups (PP = 100%). Therefore, considering the molecular data obtained from the *ITS* region, *Ceiba* was not monophyletic.



Ceiba jasminodora *Ceiba rubiflora* *Ceiba crispiflora* *Ceiba ventricosa*

Ceiba erianthos *Ceiba glaziovii* *Ceiba pubiflora* *Ceiba speciosa*

Ceiba insignis *Ceiba pentandra* *Ceiba samauma* *Ceiba lupuna*

Figure 2: Geographical distribution of the specimens of *Ceiba* spp obtained from herbarium records that are available in *species*Link.

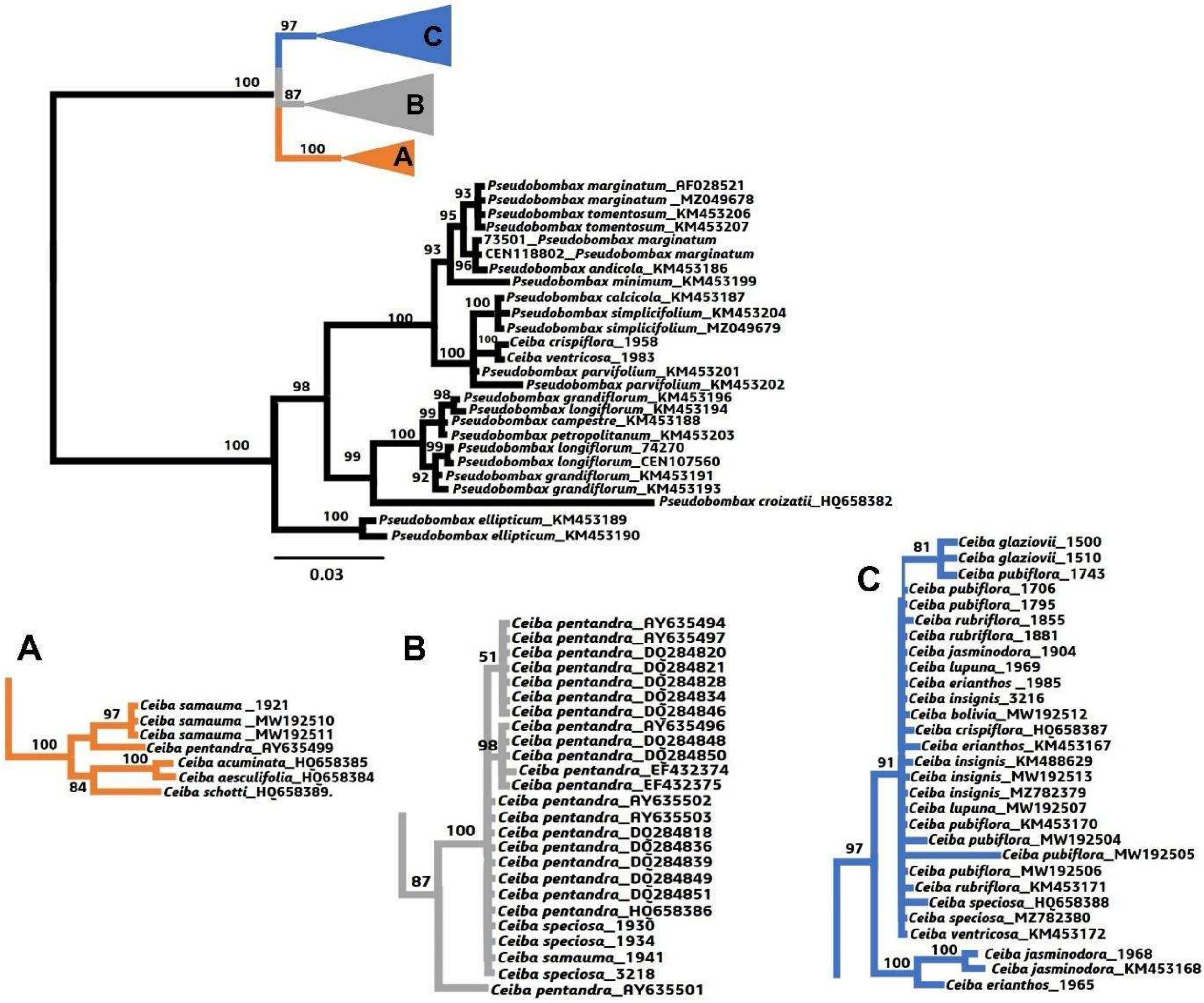


Figure 3: Bayesian analysis (consensus tree) depicting the phylogenetic relationships between the *Ceiba* spp., with *Pseudobombax* spp. as the outgroup. The dataset contained the sequences of the *ITS* region (647 bp, n = 87). The full phylogeny with the three clades (A-C) collapsed is shown on top. (A) Expanded clade A. (B) Expanded clade B. (C) Expanded clade Ceiba The branch lengths are drawn to scale and above the branches is the node's reliability, expressed in posterior probability (%). Tips indicate species and their specimen or GenBank accession numbers, while the scale bar indicates the estimated number of substitutions per site. *The *ITS* sequences of these specimens were retrieved from Duarte *et al.* (2011). See Appendix for additional information on the accession numbers.

The *ITS*-based phylogenetic tree showed that the remaining 61 specimens of *Ceiba* were grouped into three main clades (hereafter referred to as clades A, B, and C, respectively), with high PP support (> 85%). Clade A comprised two subclades. Subclade A1 included *Ceiba acuminata, Ceiba aesculifolia,* and *Ceiba schottii*, all sampled from sites outside Brazil (two from the Mesoamerican subregion and one from the Caribbean subregion; Table 1). The sister subclade A2 included three specimens of *Ceiba samauma* and a specimen of *Ceiba pentandra.* Subclade A2 shared species with the large Clade B.

Clade B comprised specimens of *Ceiba pentandra, Ceiba samauma,* and *Ceiba speciosa.* Two *Ceiba speciosa* specimens (i.e., specimen no. 1930 and 1934) were collected from the Amazonia, while the third (3218) originated from the Mesoamerican subregion. These three specimens were clustered into Clade B, independent of their location. Ceiba pentandra was the most prominent member of this clade. Clade C consisted of two strongly supported inner subclades (PP > 90%). The smaller subclade C1 included two specimens of *Ceiba jasminodora* (specimen no. 1965 and 1968, collected from the Atlantic Forest) and one of *Ceiba erianthos*. The larger subclade C2 consisted of specimens of *Ceiba ventricosa* (n = 1), *Ceiba speciosa* (n = 2), *Ceiba rubriflora* (n = 3), *Ceiba pubiflora* (n = 7), *Ceiba insignis* (n = 4), *Ceiba lupuna* (n = 2), *Ceiba erianthos* (n = 2), *Ceiba jasminodora* (n = 1), *Ceiba bolivia* (n = 1), *Ceiba crispiflora* (n = 1), and *Ceiba glaziovii* (n = 2). A putative subclade consisting of *Ceiba glaziovii* and *Ceiba pubiflora* exhibited low support (PP = 81%) and was not considered further. The specimens present in subclade C2 were collected from the Cerrado (n = 9) and Atlantic Forest (n = 6) (Figure 3).

DNA sequences were obtained from the five chloroplast genes of 31 specimens. The sequences were concatenated to form dataset B (2878 bp long), which was subsequently used to generate a Bayesian phylogenetic tree (Figure 4). The consensus tree obtained from the concatenated chloroplast genes indicated that *Ceiba* constitutes a monophyletic group. However, the phylogenetic relationships among congeners within *Ceiba* were unclear as the tree exhibited relatively low support (PP < 90%) for all major nodes. Therefore, the Bayesian phylogenetic analysis of the chloroplast gene data could not fully resolve the species relationships within the genus *Ceiba.* Nonetheless, the species clustered in a manner reminiscent of the clades observed in the *ITS*-based phylogenetic tree (Figure 3). For example, *Ceiba acuminata* and *Ceiba aesculifolia* grouped in a subclade (PP = 90%) that had identical composition to subclade A1 (Figure 3), while two specimens of *Ceiba speciosa* (specimen no. 1930 and 1934), *Ceiba samauma,* and *Ceiba pentandra* clustered in a subclade (PP = 96%), which was identical in composition to Clade B (Figure 3). Species composition in the topmost part of the tree (as shown in Figure 4) is similar to subclade C (Figure 3). Interestingly, the two specimens – *Ceiba crispiflora* (1958) and *Ceiba ventricosa* (1983) – that had been included together with the outgroup clade in the *ITS* consensus tree (Figure 3) were part of the *Ceiba* clade in the consensus tree obtained from the chloroplast genes (Figure 4).

Figure 4: Bayesian analysis (consensus tree) depicting the phylogenetic relationships between the *Ceiba* spp., with *Pseudobombax* spp. as the outgroup, using the dataset (2878 bp, n = 30) obtained from the catenation of sequences from five regions of the chloroplast genome (*trnT-trnL*, *intron-trnL*, *trnS-trnG*, *matK*, and *psbA-trnH*). The branch lengths are drawn to scale and above the branches is the node's reliability, expressed in posterior probability (%). Tips indicates species and their specimen numbers, whereas the scale bar indicates the estimated number of substitutions per site. *The chloroplast genes sequences of these specimens were retrieved from Duarte *et al.* (2011). See Appendix for additional information on the accession numbers.

Figure 5: Bayesian analysis (consensus tree) depicting the phylogenetic relationships between the *Ceiba* spp., with *Pseudobombax* spp. as the outgroup, using the dataset (3524 bp, n = 31) obtained from the catenation of sequences from regions of the chloroplast genes (*trnT-trnL*, *intron-trnL*, *trnS-trnG*, *matK*, and *psbA-trnH*) and *ITS*. The branch lengths are drawn to scale and above the branches is the node's reliability, expressed in posterior probability (%). Tips indicates species and their specimen numbers, whereas the scale bar indicates the estimated number of substitutions per site. The colour of the branches indicates the clades: I (grey), II (orange), and III (blue). Tips indicates species with their specimen or GenBank accession numbers. *Sequences of these specimens were retrieved from Duarte *et al.* (2011). See Appendix for additional information on the accession numbers.

Lastly, the sequences from the *ITS* region and the chloroplast genes were concatenated (dataset C) and used to generate a final consensus tree, rooted with *Pseudobombax* spp. as the outgroup (Figure 5). This tree yielded the best resolution of the phylogenetic relationships within the genus *Ceiba*, which was somewhat similar to that obtained with the *ITS* region alone (Figure 3). In the tree, the genus *Ceiba* occupied three main clades (hereafter referred to as clades I, II, and III). Clade I had high PP support of 100% and comprised *Ceiba speciosa*, *Ceiba samauma,* and *Ceiba pentandra* (corresponding to clade B of Figure 3). Similarly, clade II (PP = 100), consisting of *Ceiba acuminata*, *Ceiba aesculifolia, Ceiba samauma*, and *Ceiba schottii*, was identical to clade A, as shown in Figure 3.

In the combined consensus tree, clade III (PP = 100%) was grouped into two major subclades, similar in composition to clade C, which was recovered in the *ITS*-based phylogenetic tree (Figure 3). Although few differences were observed, species composition in subclades IIIA and IIIB were similar to those of subclades C1 or C2, respectively. However, *Ceiba jasminodora* (specimen no. 1904) was included in the subclade IIIA and excluded from subclade C2 in the combined tree in contrast to the *ITS*-based tree, where the reverse was observed. Also, in the tree obtained from the combined dataset, all specimens of *Ceiba erianthos* were resolved into subclade IIIB (Figure 5), whereas, in the *ITS*-based tree, *Ceiba erianthos* specimens were included in both subclade C1 and C2 (Figure 3).

3.3 Network analyses

To further explore genealogical relationships within the genus Ceiba, haplotype networks were constructed separately from the sequences of either chloroplast genes or *ITS* region. Characteristics of the respective datasets used for the haplotype network analyses and the parsimony results were summarized in Table 3. The *ITS* region provided 130 informative characters (21.5% of the total 606 characters), of which 122 were base substitutions (20.1%) and 8 were indels (1.4%). The chloroplast genes were less variable and provided 60 informative characters (3.3% of 1813 bp), consisting of 54 base substitutions (3.0%) and six indels (0.3%). Nucleotide diversity was higher in the *ITS* dataset compared to the chloroplast dataset (0.026 versus 0.007). In contrast, haplotype diversity was higher in the chloroplast dataset than in the *ITS* dataset (Table 3).

Table 3: Characteristics of DNA sequences and overview of parsimony analyses of the molecular data sets used for haplotype network.

Characters	**Nucleus**	**Chloroplast**
Gene region(s) included (bp)	*ITS*	*intron-trnL*, *trnS-trnG*, *matK*
No. of specimens included	59	27
Length of aligned matrices (bp)	606	1813
No. of parsimony informative characters	130	60
Base substitutions	122	54
Indels	8	6
Nucleotide diversity (per site)	0.026	0.007
Haplotype (gene) diversity	0.897	0.979

The haplotype network constructed from the *ITS* region recovered 23 haplotypes from 59 specimens (Figure 6). Overall, the network showed four sets of haplotypes (hereafter referred to as haplogroups), and the general network topology was highly consistent with the consensus tree obtained from the phylogenetic analysis. Thus, the haplogroups were coloured to match the phylogenetic clades displayed in the *ITS*-based tree (Figure 3): orange (corresponded to clade A), grey (clade B), blue (clade C), and black (specimens that clustered together with the outgroups).

The black haplogroup contained sequences from specimens of *Ceiba* included in the *Pseudobombax* clade, i.e., *Ceiba crispiflora* (specimen no. 1958) and *Ceiba ventricosa* (specimen no. 1983). This haplogroup was 70 mutational steps away from the nearest haplogroup and was the most derived haplogroup. The orange haplogroup consisted of five highly divergent haplotypes (3, 18, 10, 11, and 22) that were recovered from species of the clade A of Figure 3: *Ceiba samauma, Ceiba pentandra, Ceiba acuminata, Ceiba aesculifolia,* and *Ceiba schottii,* respectively. Except for *Ceiba samauma*, these specimens were collected outside Brazil (Table 1).

The grey haplogroup matched those of contained sequences obtained from specimens that belonged to the clade B of Figure 3. Haplotypes 19, 15 and 4 descended from the internal haplotype 15. Specimens of this haplogroup were collected from the Amazon and the Mesoamerican subregion and belong to either *Ceiba samauma, Ceiba speciosa,* or *Ceiba pentandra*. Haplotype 4 was the most frequent in the grey haplotype. It contained 13 *Ceiba pentandra and* the specimens of *Ceiba speciosa* and *Ceiba samauma*.

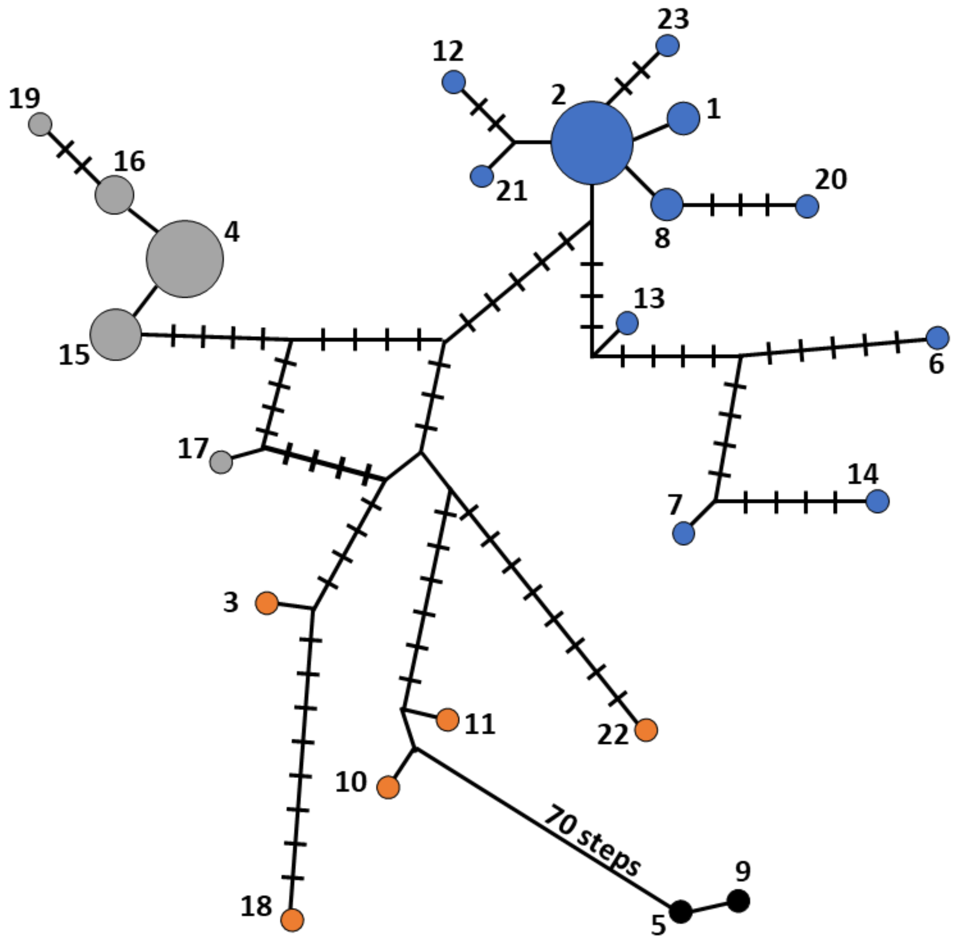


Figure 6: Median-joining haplotype network for 59 specimens of *Ceiba* spp. The dataset consisted of the *ITS* region sequences (606 bp). Circles within the network depict a specific haplotype (identified with numbers) and the size of each circle is proportional to the frequency of that haplotype. Bars represent number of mutational steps when there is more than one (unless the number of steps is given). The colour codes of the circles correspond with the clades identified in the *ITS*-based consensus tree (Figure 3): black corresponds to specimens clustered with the outgroups, orange corresponds to clade A, grey corresponds to clade B, and blue corresponds to clade C. See Appendix for additional information on the haplotypes.

The blue haplogroup consisted of two components. The first component was arranged around its most frequent member (haplotype 2, N=13), which occurred in specimens of either *Ceiba rubriflora*, *Ceiba pubiflora, Ceiba erianthos*, *Ceiba jasminodora, Ceiba insignis*, or *Ceiba crispiflora*. Its tip haplotypes 1, 12, 20, 21, and 23 included specimens of either *Ceiba glaziovii, Ceiba pubiflora, Ceiba erianthos*, *Ceiba pubiflora, Ceiba speciosa*, or *Ceiba lupuna*. The second component of the blue haplogroup contained only tip haplotypes (6, 7, 13, and 14), all highly derived haplotypes. Specimens in the blue haplogroup were collected within Brazil (the Cerrado or the Atlantic Forest). The distances between the nearest haplotypes varied, usually requiring many mutational steps to connect to the frequent haplotype 2 (from 5 to 20 steps).

The second haplotype network was constructed from the chloroplast sequences from 27 specimens of *Ceiba* (Figure 7). Sequences from the *trnT-trnL* and *psbA-trnH* regions contained several missing data and were thus excluded from subsequent analyses. Therefore, the network analysis included only sequence information from *trnT-trnL*, *intron-trnL*, *trnS-trnG*, and *matK* genes. As presented in Figure 7, the network recovered 21 haplotypes. Despite the high haplotype diversity, low nucleotide diversity estimates suggest relatively low levels of polymorphisms across the dataset (Table 3). The haplotype network showed that most haplotypes were singletons, i.e., that particular sequence appeared only once in the dataset. Short distances separated the haplotypes, and thus, they showed small levels of differentiation. A similar trend was observed in the chloroplast gene-based consensus tree, where the partitioning of distinct lineages was inefficient. Haplotypes 6 and 3 had the highest number of sequences (both contained three specimens), followed closely by haplotype 12 (2 specimens), while the remaining haplotypes were singletons. Conspicuously, several haplotypes were organised around the most frequent haplotype, particularly around haplotype 6, with the less common haplotypes organised at the tips of the network. Nonetheless, this haplotype network had two components. In the first component, some haplotypes were organised around the most frequent haplotype, particularly haplotype 6, with the less common haplotypes organised at the tips. The other part is mainly comprised of highly derived tip haplotypes.

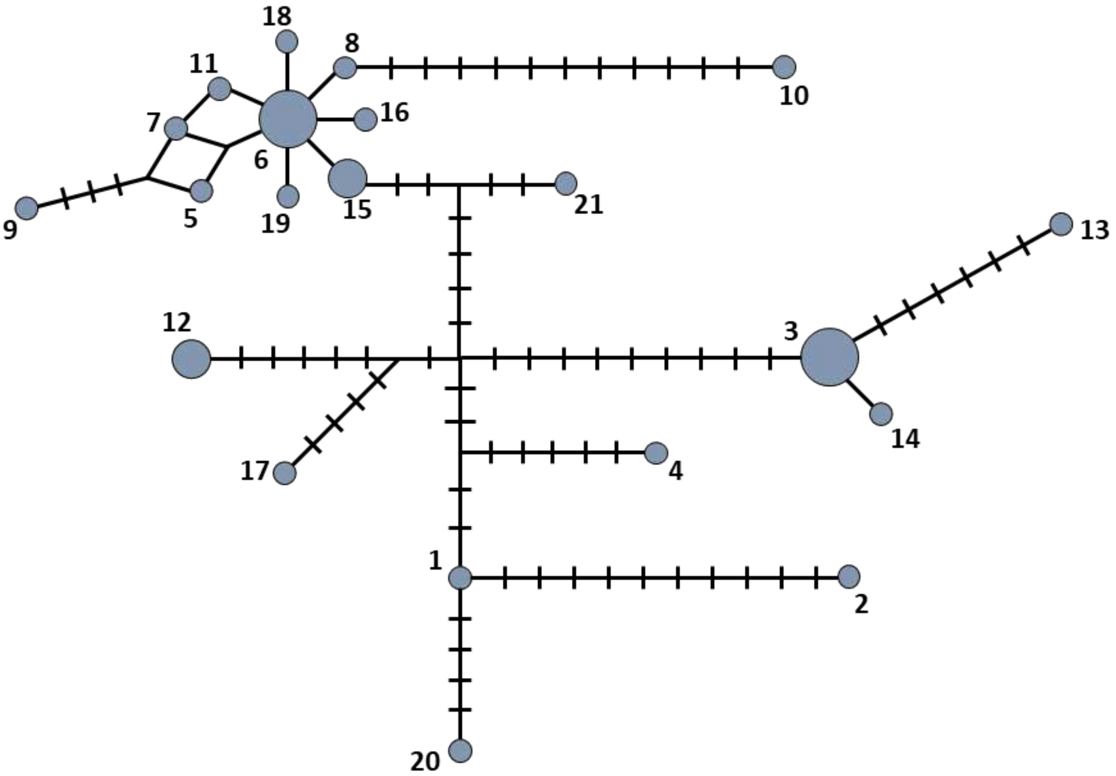


Figure 7: Median-joining haplotype network for 27 specimens of *Ceiba* spp. The dataset consisted of sequences from three chloroplast region (*trnT-trnL*, *intron-trnL*, *trnS-trnG*, and *matK*; 1813 bp). Circles within the network depicts a specific haplotype (identified with numbers) and the size of each circle is proportional to the frequency of that haplotype. Bars represent number of mutational steps when there is more than one (unless the number of steps is given). See Appendix for additional information on the haplotypes.

4. DISCUSSION

4.1 Combined dataset produced a better resolution of the genus *Ceiba*

Given the difficulty in distinguishing different processes that define evolutionary dynamics, the use of multiple genes for general population phylogenies is necessary to increase the reliability of the phylogeny (Futuyma and Kirkpatrick, 2013). In this study, we carried out three separate phylogenetic analyses based on molecular characters from the *ITS* region, chloroplast genes, and combined datasets, generating a consensus tree for each (Figure 4-5). Our *ITS*-based phylogenetic tree was highly congruent with the maximum likelihood phylogram reported previously by Pezzini *et al.* (2021). However, some specimens and species were not well resolved, especially factoring in their geographical distribution. For example, our *ITS*-based phylogenetic tree did not distinguish between *Ceiba speciosa* specimens from the Amazon and Mesoamerican subregions, whereas *Ceiba erianthos* (all from the Atlantic Forest) and *Ceiba jasminodora* (all from the Cerrado) specimens were separated into two subclades. On the other hand, the chloroplast sequence dataset, on its own, failed to resolve the phylogenetic relationships among the species of *Ceiba* and exhibited relatively low support (PP < 85%). However, the distinction between the *Ceiba* and *Pseudobombax* was strongly supported (PP = 90%).

However, analysis of the combined dataset improved the resolution of the species of *Ceiba*. The resulting consensus tree presented a pattern of phylogeographic congruence, in which the phylogenetic placement of the members of distinct clades and subclades paralleled the geographic distribution of the specimens. Furthermore, although the *Ceiba speciosa* specimens were included in the same clade, a strongly supported distinction (PP = 100%) was observed between the specimens from the Amazon (specimen no. 1930 and 1934) and Mesoamerican (specimen no. 3218) regions. Also, the specimens of *Ceiba erianthos* and *Ceiba jasminodora* were resolved into the subclades IIIA and IIIB, respectively.

Overall, the phylogenetic relationships within species of the genus *Ceiba* were better revealed using the combined sequence data from the *ITS* and chloroplast genes regions. Although the sequences from the *ITS* region present far more parsimony informative characters, the previous reports of Oliveira *et al.* (2007), Tan *et al.* (2015), and Wang *et al.* (2018) support the superiority of combinations of nuclear and plastid markers to individual datasets.

4.2 Main clades within the genus *Ceiba*

Results obtained from this study indicate that the genera of *Ceiba* and *Pseudobombax* were largely separated into two distinct lineages. However, based on *ITS* data alone, the genus *Ceiba* is not monophyletic, which is demonstrated by the inclusion of specimen no. 1958 (*Ceiba crispiflora*) and 1983 (*Ceiba ventricosa*) within the *Pseudobombax* clade in the *ITS*-based and combined dataset-based phylogenetic trees. Nonetheless, the consensus trees were well supported (PP > 85%) and were concordant with previous molecular phylogenies of the genus *Ceiba* (Duarte *et al.,* 2011, Carvalho-Sobrinho *et al.*, 2016).

The high levels of genetic differentiation and the topology of the combined dataset-based tree present strong evidence that three genealogical lineages (clades) make up the genus *Ceiba*, which is consistent with a ML phylogram of Pezzini *et al.* (2021). The first clade brought together specimens of *Ceiba speciosa, Ceiba samauma,* and *Ceiba pentandra* (the most frequent member). As per the description by Gibbs and Semir (2003), members of this clade are morphologically similar, with trees 10 to 20 m high, oblong-spathulate petals, long (10 to 15 mm) and stout pedicels, and fruits shaped into ellipsoidal/pyriform capsule. The specimens collected are distributed between the seasonally dry forests of the Mesoamerican region (Mexico) and the Amazon rainforest of Brazil. The *Ceiba speciosa* specimen from the Mesoamerican region (specimen no. 3218) was distinguished from the *Ceiba speciosa* specimens from the Amazon in the combined and chloroplast data analyses. The *ITS*-based tree and haplotype network made no distinction. In contrast, in the chloroplast genes-derived haplotype network, there was a mutational step between the specimens from the Amazon, while the specimen from the Mesoamerican region was 17 mutational steps away. Cytoplasmic gene flow frequently occurs among related species of plants and can be detected even when nuclear gene flow is not evident (Rieseberg and Soltis, 1991). Thus, incomplete lineage sorting or hybridization of the chloroplast region as the species transitioned to another region may explain this incongruence. Nonetheless, future studies with a more robust sampling strategy may seek to clarify this incongruence.

Members of the second clade have relatively shorter trunks than the first clade (~8 m) which were aculeate. They have 3 to 7 foliate leaves, with axillary flowers born singly. The pedicles are stout, and the fruits (ellipsoidal to pyriform capsule) enclose non-winged seeds (Gibbs and Semir, 2002; Carvalho-Sobrinho *et al.,* 2016). Specimens in this clade, comprising *Ceiba acuminata, Ceiba aesculifolia,* and *Ceiba schottii*, were collected from outside Brazil, which suggests a range-dependent polymorphism for this clade. The only exception was the *Ceiba samauma* collected from the Amazon.

The third clade is the largest and most diversified clade, comprising specimens collected from Mesoamerican, Cerrado, and Atlantic Forest. The subclades are also well-supported (PP > 85%; Figure 6). The morphological features of the species within this clade vary widely as well. For example, height can range from as low as 1-2 m (in *Ceiba jasminodora*), 10-15 m in (*Ceiba erianthos, Ceiba crispiflora, Ceiba insignis,* and *Ceiba glaziovii*), to as high as 50 m in (*Ceiba pentandra* and *Ceiba lupuna*) (Gibbs and Semir, 2002). The trunks are typically swollen (ventricose) and aculeate. Leaflets may appear coriaceous, entire, or denticulate, with various forms of oblong leaves. Fruits are ellipsoidal or pyriform capsules with varied dimensions (Gibbs and Semir, 2002).

The *ITS*-based haplotype network provided further insights into the genealogical relationships among the members of this clade (which coincides with the blue haplogroup; Figure 6). This haplogroup presents two coexisting scenarios. In the first scenario, several tip haplotypes were descended fromhaplotype2 (the most prominent haplotype), suggesting a recent population expansion among these congeners of *Ceiba*, with not enough time to fully differentiate (Garcia *et al.*, 2021; Nehra *et al.*, 2022). In the second scenario, the large number of mutational steps among the highly derived haplotypes, also descended from haplotype 2, suggest an ancient expansion, at least long enough to allow for genetic differentiation (Garcia *et al.*, 2021; Nehra *et al.*, 2022). Furthermore, specimens of the ancestor haplotype 2 were mainly distributed across the Atlantic Forest, suggesting that the members of the third clade from the Mesoamerican and Cerrado evolved under transition from this region.

4.3 Incongruent placement of *Ceiba crispiflora* and *Ceiba ventricosa*

The placement of *Ceiba crispiflora* and *Ceiba ventricosa* within the outgroups in the phylogenic trees independently generated from the *ITS* and combined dataset contradicted the apparent correspondence in the phylogeny and was unexpected. Besides the possible problems of gene duplication, the incongruency can be explained by two non-exclusive evolutionary mechanisms that influence individual gene flow: introgressive hybridization and incomplete lineage sorting (ILS). On the one hand, ILS sustains ancestral genetic polymorphisms, invoking incongruences. However, ILS events are most likely to occur when new lineages rapidly descend from ancestors (i.e., during speciation) with large effective population sizes (Feng *et al.,* 2022). The specimens of *Ceiba crispiflora* and *Ceiba ventricosa* within the outgroups were 13 and 14 mutational steps, respectively, away from the nearest specimen of *Pseudobombax* (data not shown). This suggests a longer divergent time (enough for speciation) and reduces the likelihood of ILS.

On the other hand, the specimens in contention could have acquired the *ITS* region from a congener of *Pseudobombax,* in a hybridization event. Natural selection could have favoured the integration of the gene into the nuclear genome (Stevison, 2008). Although several studies have reported chloroplast genes hybridization, resulting in incongruent phylogeny (Brunsfeld, 1990; Oliveira *et al.,* 2007), nuclear hybridization has also been reported in plants, with the absence of chloroplast hybridization. Critchfield (1985), Forrest (1980), and Wheeler and Guries (1987), for example, detected nuclear character hybridization in jack pine and lodgepole populations. Despite extensive sampling of sympatric and a1lopatric populations, significant chloroplast gene hybridization was undetected in these populations (Wagner *et al.,* 1987). While the rate and ease of chloroplast gene transfer across species barriers seem far higher than nuclear genes, this trend is far from constant (Rieseberg and Soltis, 1991). Besides, hybridization events occur after speciation (Feng *et al.,* 2022), and may be the case in this study. Additionally, these specimens of *Ceiba crispiflora* and *Ceiba ventricosa* were collected from the Atlantic Forest, which has a multifaceted phylogeographic structure that indicates the long-term survival of populations within the region (Oliveira *et al.,* 2010). Thus, intraregional variation is another plausible explanation for the observed incongruence. Nevertheless, distinguishing the evolutionary events that cause phylogenetic congruence remains problematic, and more data are required to make firm inferences regarding the source of the incongruency.

4.4 Biogeography of *Ceiba* spp

The geographic distribution of the specimens of *Ceiba* spp (Figure 2). is highly congruent with the clades identified in the combined dataset-based tree (Figure 5). Members of clade III were mainly distributed across the dry seasonal Atlantic Forest (*Ceiba jasminodora, Ceiba glaziovii, Ceiba rubriflora, Ceiba pubiflora, Ceiba crispiflora, Ceiba erianthos*), with a few extending into the Cerrado savannas (*Ceiba pubiflora* and *Ceiba erianthos*). *Ceiba speciosa* and *Ceiba pentandra,* which appear to originate from the Atlantic Forest and then spread into the Amazon, were the most frequent member of clade I. Specimens of *Ceiba samauma,* the only member of clade II whose geographical distribution could be retrieved from *species*Link, were densely distributed across the Amazon, with very few specimens from the Atlantic Forest. Interestingly, *Ceiba lupuna*, the only species of *Ceiba* without any specimen from the Atlantic Forest, was included in clade III, which is dominated by specimens from the Atlantic Forest.

These data posit the Atlantic Forest as the probable origin of the genus *Ceiba*, which is reinforced by the observation that the grey, orange, and black haplogroups descended from the blue haplogroup (Figure 6). The congeners of the genus then diversified and dispersed across the Cerrado into the Amazon, with species such as *Ceiba pentandra* extending into the Mesoamerican region. Indeed, the topography of the *ITS*-based haplotype network shows that these descendants were mostly tip haplotypes, indicating that they were of more recent origin.

5. CONCLUSIONS

The results of this study provided molecular insights into the genealogical relationships among *Ceiba* species, including evidence for range-dependent polymorphism, the extent of genetic diversity in *Ceiba* in relation to geography and vegetation type, and the non-monophyly of the genus *Ceiba*. With *Ceiba* being a model genus for research on Neotropical diversity, these insights contribute to the existing knowledge on the evolutionary dynamics and mechanisms maintaining the diversity of arboreal flora within the Neotropical biomes.

REFERENCES

Aboul-Maaty, N.A.F. and Oraby, H.A.S., 2019. Extraction of high-quality genomic DNA from different plant orders applying a modified CTAB-based method. Bull. Natl. Res. Cent. 43(1),1-10.

Akaike, H., 1973. Information theory and an extension of maximum likelihood principle,in: Petrov, BN, Csaki, F. (Eds.), Second International Symposium on Information Theory. AkademiaiKiado, Budapest, pp. 267–281.

Antonelli, A., Zizka, A., Carvalho, F.A., Scharn, R., Bacon, C.D., Silvestro, D. and Condamine, F.L., 2018. Amazonia is the primary source of Neotropical biodiversity. Proc. Nat. Acad. Sci. 115(23), 6034-6039.

Bandelt, HJ, Forster, P., Rohl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol.Evol.16, 37–48.

Bayer, C. and Kubitzki, K., 2003. Malvaceae, in: Kubitzki, K., Bayer, C. (Eds.), Flowering plants· dicotyledons: The Families and Genera of Vascular Plants. Springer, Berlin, Heidelberg, pp. 225-311.

Brunsfeld, S.J., 1990. *Systematics and evolution in Salix section Longifoliae*. Washington State University.

Carvalho-Sobrinho, J.G., Alverson, W.S., Alcantara, S., Queiroz, L.P., Mota, A.Ceiba and Baum, D.A., 2016. Revisiting the phylogeny of Bombacoideae (Malvaceae): Novel relationships, morphologically cohesive clades, and a new tribal classification based on multilocus phylogenetic analyses. Mol. Phylogenet. Evol. 101, 56-74.

Cota-Sánchez, JH, Remarchuk, K., Ubayasena, K., 2006. Ready-to-use DNA extracted with a CTAB method adapted for herbarium samples and mucilaginous plant tissue. Plant Mol. Biol. Rep.24(2), 161.

Critchfield, W.B., 1985. The late Quaternary history of lodgepole and jack pines. Can. J. For. Res.15(5), 749-772.

Derryberry, E.P., Claramunt, S., Derryberry, G., Chesser, R.T., Cracraft, J., Aleixo, A., Pérez-Emán, J., Remsen, Jr, J.V. and Brumfield, R.T., 2011. Lineage diversification and morphological evolution in a large-scale continental radiation: the Neotropical ovenbirds and woodcreepers (Aves: Furnariidae). Evol.: Int. J. Org. Evol. 65(10), 2973-2986.

Drummond, AJ, Suchard, MA, Xie, D., Rambaut, A., 2012 Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol.Evol.29,1969–1973.

Duarte, M.Ceiba, Esteves, G.L., Salatino, M.L.F., Walsh, K.Ceiba and Baum, D.A., 2011. Phylogenetic analyses of Eriotheca and related genera (Bombacoideae, Malvaceae). Syst. Bot. 36(3), 690-701.

Espeland, M., Hall, J.P., DeVries, P.J., Lees, D.Ceiba, Cornwall, M., Hsu, Y.F., Wu, L.W., Campbell, D.L., Talavera, G., Vila, R. and Salzman, S., 2015. Ancient Neotropical origin and recent recolonisation: Phylogeny, biogeography and diversification of the Riodinidae (Lepidoptera: Papilionoidea). Mol. Phylogenet. Evol. 93, 296-306.

Feng, S., Bai, M., Rivas-González, I., Li, C., Liu, S., Tong, Y., Yang, H., Chen, G., Xie, D., Sears, K.E., Franco, L.M., 2022. Incomplete lineage sorting and phenotypic evolution in marsupials. Cell. 185(10), 1646-1660.

Forrest, G.I., 1980. Geographical variation in the monoterpenes of Pinus contorta oleoresin. Biochem. Syst. Ecol. 8(4), 343-359.

Futuyma, D.J. and Kirkpatrick, M., 2013. Evolution, third ed.). Sunderland, Massachusetts.

Garcia, E., Wright, D., Gatins, R., Roberts, M.B., Pinheiro, H.T., Salas, E., Chen, J.Y., Winnikoff, J.R., Bernardi, G., 2021. Haplotype network branch diversity, a new metric combining genetic and topological diversity to compare the complexity of haplotype networks. PloS one. 16(6), e0251878.

Gibbs, P. and Semir, J., 2002. A taxonomic revision of the genus Ceiba Mill.(Bombacaceae). An. Jard. Bot. Madr. 60(2), 259-300.

Gómez-Maqueo, X. and Gamboa-deBuen, A., 2022. The Biology of the Genus Ceiba, a Potential Source for Sustainable Production of Natural Fiber. Plants. 11(4), 521.

Kimura, M., 1969. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. Genetics. 61(4), 893.

Leal, B.S.S., Palma da Silva, Ceiba and Pinheiro, F., 2016. Phylogeographic studies depict the role of space and time scales of plant speciation in a highly diverse Neotropical region. Crit. Rev. Plant Sci. 35(4), 215-230.

Librado, P. and Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 25(11), 1451-1452.

Lozano-García, S., Figueroa-Rangel, B., Sosa-Nájera, S., Caballero, M., Noren, A.J., Metcalfe, S.E., Tellez-Valdés, O. and Ortega-Guerrero, B., 2021. Climatic and anthropogenic influences on vegetation changes during the last 5000 years in a seasonal dry tropical forest at the northern limits of the Neotropics. Holocene. 31(5), 802-813.

Nehra,A.K., Kumari, A., Kundave, V.R., Vohra, S., Ram, H., 2022. Molecular insights into the population structure and haplotype network of Theileriaannulata based on the small-subunit ribosomal RNA (18S rRNA) gene. Infect. Genet. Evol. 99, 105252.

Oliveira, L.O., Huck, R.B., Gitzendanner, M.A., Judd, W.S., Soltis, D.E. and Soltis, P.S., 2007. Molecular phylogeny, biogeography, and systematics of Dicerandra (Lamiaceae), a genus endemic to the southeastern United States. Am. J. Bot. 94(6), 1017-1027.

Oliveira, L.O.D., Rossi, A.A.B., Martins, E.R., Batista, F.R.D.Ceiba and Silva, R.S., 2010. Molecular phylogeography of Carapichea ipecacuanha, an amphitropical shrub that occurs in the understory of both semideciduous and evergreen forests. Mol. Ecol. 19(7), 1410-1422.

Pezzini, F.F., Dexter, K.G., de Carvalho-Sobrinho, J.G., Kidner, CeibaA., Nicholls, J.A., De Queiroz, L.P. and Pennington, R.T., 2021. Phylogeny and biogeography of *Ceiba*Mill.(Malvaceae, Bombacoideae). Front. Biogeogr.13, 1–18.

Pires, N.D. and Dolan, L., 2012. Morphological evolution in land plants: new designs with old genes. Philos. Trans. R. Soc. B Biol. Sci. 367(1588), 508-518.

Rieseberg, L.H. and Soltis, D.E., 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. Evol. Trends Plants. 5, 65-84

Robyns, A., 1961. Contribution a l'etudemonographique du genre Bombax sl II. La Typification de B. Ceiba L. Taxon. 10(6), 156-160.

Ronquist, F., Huelsenbeck, JP, 2003. MrBayes 3: bayesian phylogenetic inference under mixed models. Bioinformatics. 19, 1572–1574.

Rull, V., 2020. Neotropical diversification: historical overview and conceptual insights, in: Rull, V., Carnaval, A. (Eds.), Neotropical diversification: Patterns and processes, Fascinating Life Sciences. Springer, Cham, pp. 13-49.

Soltis, P.S. and Soltis, D.E., 2021. Plant genomes: Markers of evolutionary history and drivers of evolutionary change. Plants people planet. 3(1), 74-82.

Stevison, L., 2008. Hybridization and gene flow. Nat. Edu. 1(1), 111

Tan, G., Muffato, M., Ledergerber, C., Herrero, J., Goldman, N., Gil, M., Dessimoz,C., 2015. Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. Syst. Biol. 64(5), 778-791.

Urtubey, E., Stuessy, T.F., Tremetsberger, K. and Morrone, J.J., 2010. The South American biogeographic transition zone: an analysis from Asteraceae. Taxon. 59(2), 505-509.

Wang, K., Lenstra, J.A., Liu, L., Hu, Q., Ma, T., Qiu, Q., Liu, J., 2018. Incomplete lineage sorting rather than hybridization explains the inconsistent phylogeny of the wisent. Commun. Biol. 1(1), 1-9.

Wheeler, N.Ceiba and Guries, R.P., 1987. A quantitative measure of introgression between lodgepole and jack pines. Can. J. Bot. 65(9), 1876-1885.

APPENDIX

Table S1: List of sequences included in the *ITS* dataset (dataset A)

S/N	Species	Specimen/ GenBank No.	S/N	Species	Specimen/ GenBank No.
1	*Ceiba acuminata**	HQ658385	45	*Ceiba pubiflora*	KM453170
2	*Ceiba aesculifolia**	HQ658384	46	*Ceiba pubiflora*	MW192504
3	*Ceiba boliviana*	MW192512	47	*Ceiba pubiflora*	MW192505
4	*Ceiba crispiflora**	1958	48	*Ceiba pubiflora*	MW192506
5	*Ceiba crispiflora**	HQ658387	49	*Ceiba rubriflora**	1855
6	*Ceiba erianthos**	1965	50	*Ceiba rubriflora**	1881
7	*Ceiba erianthos**	1985	51	*Ceiba rubriflora*	KM453171
8	*Ceiba erianthos**	KM453167	52	*Ceiba samauma**	1921
9	*Ceiba glaziovii**	1500	53	*Ceiba samauma**	1941
10	*Ceiba glaziovii**	1510	54	*Ceiba samauma*	MW192510
11	*Ceiba insignis**	3216	55	*Ceiba samauma*	MW192511
12	*Ceiba insignis*	KM488629	56	*Ceiba schottii**	HQ658389
13	*Ceiba insignis*	MW192513	57	*Ceiba speciosa**	1930
14	*Ceiba insignis*	MZ782379	58	*Ceiba speciosa**	1934
15	*Ceiba jasminodora**	1904	59	*Ceiba speciosa**	3218
16	*Ceiba jasminodora**	1968	60	*Ceiba speciosa*	HQ658388
17	*Ceiba jasminodora*	KM453168	61	*Ceiba speciosa*	MZ782380
18	*Ceiba lupuna**	1969	62	*Ceiba ventricosa*	1983
19	*Ceiba lupuna*	MW192507	63	*Ceiba ventricosa*	KM453172
20	*Ceiba pentandra*	AY635494	64	*Pseudobombax andicola*	KM453186
21	*Ceiba pentandra*	AY635496	65	*Pseudobombax calcicola*	KM453187
22	*Ceiba pentandra*	AY635497	66	*Pseudobombax campestre*	KM453188
23	*Ceiba pentandra*	AY635499	67	*Pseudobombax croizatti**	HQ658382
24	*Ceiba pentandra*	AY635501	68	*Pseudobombax ellipticum*	KM453189
25	*Ceiba pentandra*	AY635502	69	*Pseudobombax ellipticum*	KM453190
26	*Ceiba pentandra*	AY635503	70	*Pseudobombax grandiflorum*	KM453191
27	*Ceiba pentandra*	DQ284818	71	*Pseudobombax grandiflorum*	KM453193
28	*Ceiba pentandra*	DQ284820	72	*Pseudobombax grandiflorum*	KM453196
29	*Ceiba pentandra*	DQ284821	73	*Pseudobombax longiflorum**	74270
30	*Ceiba pentandra*	DQ284828	74	*Pseudobombax longiflorum*	CEN107560
31	*Ceiba pentandra*	DQ284834	75	*Pseudobombax longiflorum*	KM453194
32	*Ceiba pentandra*	DQ284836	76	*Pseudobombax manginatum*	73501
33	*Ceiba pentandra*	DQ284839	77	*Pseudobombax manginatum*	AF028521
34	*Ceiba pentandra*	DQ284846	78	*Pseudobombax manginatum**	CEN118802
35	*Ceiba pentandra*	DQ284848	79	*Pseudobombax manginatum*	MZ049678
36	*Ceiba pentandra*	DQ284849	80	*Pseudobombax minimum*	KM453199
37	*Ceiba pentandra*	DQ284850	81	*Pseudobombax parvifolium*	KM453201
38	*Ceiba pentandra*	DQ284851	82	*Pseudobombax parvifolium*	KM453202
39	*Ceiba pentandra*	EF432374	83	*Pseudobombax petropolitanum*	KM453203
40	*Ceiba pentandra*	EF432375	84	*Pseudobombax simplicifolium*	KM453204
41	*Ceiba pentandra**	HQ658386	85	*Pseudobombax simplicifolium*	MZ049679
42	*Ceiba pubiflora**	1706	86	*Pseudobombax tomentosum*	KM453206
43	*Ceiba pubiflora**	1743	87	*Pseudobombax tomentosum*	KM453207

44	*Ceiba pubiflora**	1795

**ITS* sequences included in the combined dataset.

Table S2: Information on the sequences included in the chloroplast-based dataset (dataset B).

S/N	Species	*trnT-trnL*	*intron-trnL*	*trnS-trnG*	*matK*	*psbA-trnH*
1	*Ceiba glaziovii*_1500	✓	✓	✓	✓	✓
2	*Ceiba glaziovii*_1510	-	✓	KM453117	-	-
3	*Ceiba pubiflora*_1706	-	✓	✓	✓	✓
4	*Ceiba pubiflora*_1743	✓	✓	✓	✓	✓
5	*Ceiba pubiflora*_1795	✓	✓	✓	✓	✓
6	*Ceiba rubriflora*_1881	✓	✓	✓	✓	✓
7	*Ceiba rubriflora*_1855	✓	✓	✓	✓	✓
8	*Ceiba jasminodora*_1904	✓	✓	✓	✓	-
9	*Ceiba jasminodora*_1908	-	✓	✓	✓	✓
10	*Ceiba samauma*_1921	-	-	✓	-	-
11	*Ceiba speciosa*_1930	✓	✓	✓	✓	✓
12	*Ceiba speciosa*_1934	✓	✓	✓	✓	✓
13	*Ceiba samauma*_1941	✓	✓	✓	✓	✓
14	*Ceiba crispiflora*_1958	✓	✓	✓	✓	✓
15	*Ceiba erianthos*_1960	✓	✓	✓	✓	✓
16	*Ceiba erianthos*_1965	-	-	KM453116	✓	✓
17	*Ceiba jasminodora*_1968	✓	✓	✓	✓	✓
18	*Ceiba lupuna*_1969	-	✓	-	-	-
19	*Ceiba pentandra*_1970	✓	✓	✓	✓	✓
20	*Ceiba erianthos*_1985	✓	-	✓	✓	✓
21	*Ceiba speciosa*_3218	-	✓	✓	✓	✓
22	*Ceiba insignis*_3216	-	✓	✓	-	-
23	*Ceiba acuminata*	-	HQ696752	-	HQ696700	-
24	*Ceiba aesculifolia*	-	HQ696751	-	HQ696699	-
25	*Ceiba pentandra*	-	HQ696753	KM453118	HQ696701	-
26	*Ceiba schotti*	-	HQ696756	-	HQ696703	-
27	*Ceiba crispiflora*	-	HQ696754	-	AY321169	-
28	*Pseudobombax manginatum*	✓	✓	✓	✓	-
29	*Pseudobombax longiflorum*	✓	✓	✓	✓	-
30	*Pseudobombax croizatti*	-	HQ696749	KM453147	HQ696697	

✓= samples collected in this study; - = missing data. Numbers preceded by KM, HQ or AY are sequences retrieved from GenBanK.

Table S3: Information on the sequences included in the combined dataset (dataset C).

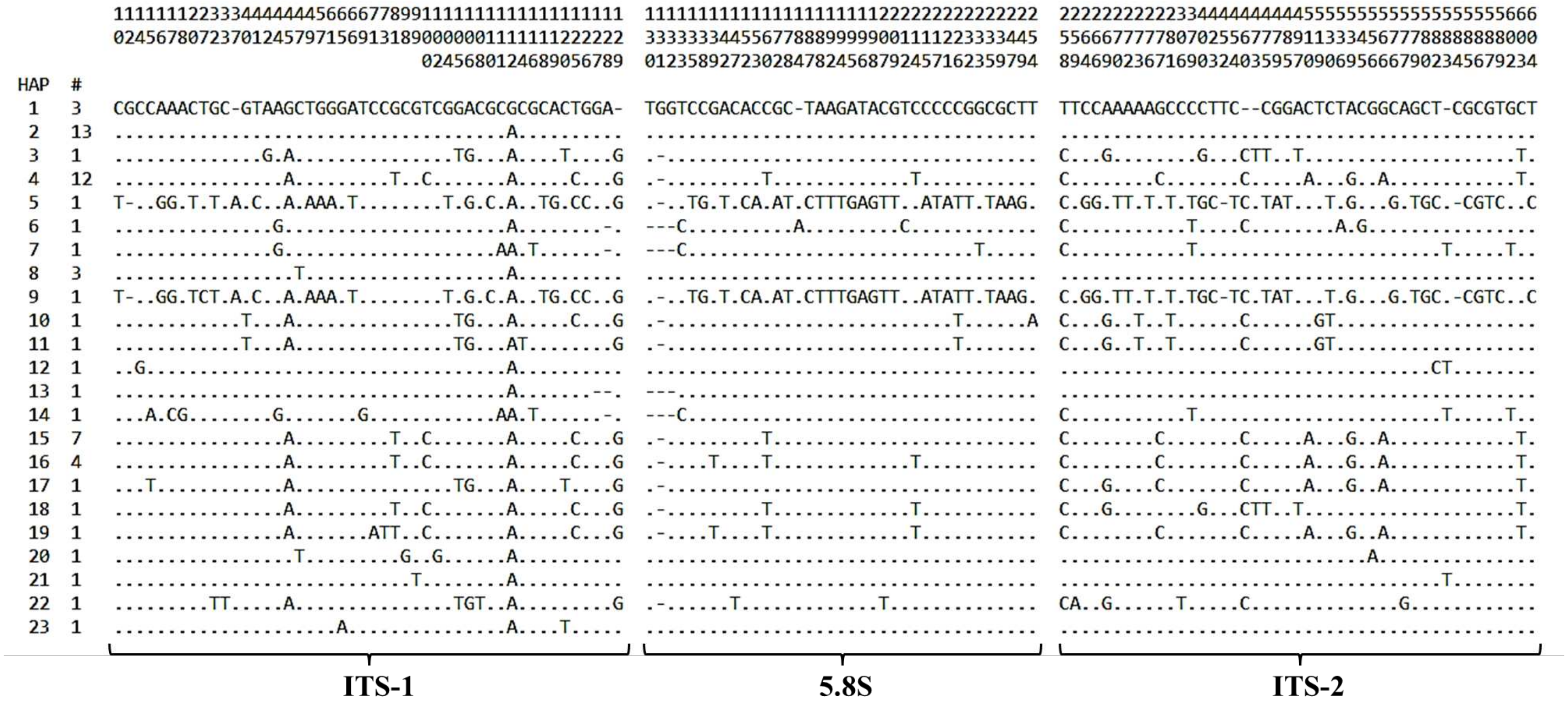
S/N	Species	ITS	*trnT-trnL*	*intron-trnL*	*trnS-trnG*	*matK*	*psbA-trnH*
1	*Ceiba glaziovii*_1500	✓	✓	✓	✓	✓	✓
2	*Ceiba glaziovii*_1510	✓	-	✓	KM453117	-	-
3	*Ceiba pubiflora*_1706	✓	-	✓	✓	✓	✓
4	*Ceiba pubiflora*_1743	✓	✓	✓	✓	✓	✓
5	*Ceiba pubiflora*_1795	✓	✓	✓	✓	✓	✓
6	*Ceiba rubriflora*_1881	✓	✓	✓	✓	✓	✓
7	*Ceiba rubriflora*_1855	✓	✓	✓	✓	✓	✓
8	*Ceiba jasminodora*_1904	✓	✓	✓	✓	✓	-
9	*Ceiba jasminodora*_1908	-	-	✓	✓	✓	✓
10	*Ceiba samauma*_1921	✓	-	-	✓	-	-
11	*Ceiba speciosa*_1930	✓	✓	✓	✓	✓	✓
12	*Ceiba speciosa*_1934	✓	✓	✓	✓	✓	✓
13	*Ceiba samauma*_1941	✓	✓	✓	✓	✓	✓
14	*Ceiba crispiflora*_1958	✓	✓	✓	✓	✓	✓
15	*Ceiba erianthos*_1960	✓	✓	✓	✓	✓	✓
16	*Ceiba erianthos*_1965	KM453167	-	-	KM453116	✓	✓
17	*Ceiba jasminodora*_1968	✓	✓	✓	✓	✓	✓
18	*Ceiba lupuna*_1969	✓	-	✓	-	-	-
19	*Ceiba pentandra*_1970	-	✓	✓	✓	✓	✓
20	*Ceiba ventricosa*_1983	✓	-	-	-	-	-
21	*Ceiba erianthos*_1985	✓	✓	-	✓	✓	✓
22	*Ceiba speciosa*_3218	✓	-	✓	✓	✓	✓
23	*Ceiba insignis*_3216	✓	-	✓	✓	-	-
24	*Ceiba acuminata*	HQ658385	-	HQ696752	-	HQ696700	-
25	*Ceiba aesculifolia*	HQ658384	-	HQ696751	-	HQ696699	-

26	*Ceiba pentandra*	HQ658386	-	HQ696753	KM453118	HQ696701	-
27	*Ceiba schotti*	HQ658389	-	HQ696756	-	HQ696703	-
28	*Ceiba crispiflora*	HQ658387	-	HQ696754	-	AY321169	-
29	*Pseudobombax manginatum*	✓	✓	✓	✓	✓	-
30	*Pseudobombax longiflorum*	✓	✓	✓	✓	✓	-
31	*Pseudobombax croizatti*	HQ658382	-	HQ696749	KM453147	HQ696697	

✓= samples collected in this study; - = missing data. Numbers preceded by KM, HQ or AY are sequences retrieved from GenBanK.

Table S4: Components of the ITS region sequences-derived haplotype network

Haplotype	Specimens
1	*Ceiba glaziovii*_1500, *Ceiba glaziovii*_1510, *Ceiba lupuna*_1743
2	*Ceiba pubiflora*_1706, *Ceiba pubiflora*_1795, *Ceiba rubriflora*_1855, *Ceiba rubriflora*_1881, *Ceiba jasminodora*_1904, *Ceiba erianthos*_1985, Cins_2476, *Ceiba crispliflora*_HQ658387, *Ceiba insignis*_MZ782379, *Ceiba pubiflora*_KM453170, *Ceiba pubiflora*_MW192506, *Ceiba speciosa*_MZ782380, *Ceiba ventricosa*_KM453172
3	*Ceiba samuama*_1921
4	*Ceiba speciosa*_1930, *Ceiba speciosa*_1934, *Ceiba samuama*_1941, *Ceiba speciosa*_3218, *Ceiba pentandra*_AY635502, *Ceiba pentandra*_AY635503, *Ceiba pentandra*_DQ284818, *Ceiba pentandra*_DQ284836, *Ceiba pentandra*_DQ284839, *Ceiba pentandra*_DQ284849, *Ceiba pentandra*_DQ284851, *Ceiba pentandra*_HQ658386
5	*Ceiba crispliflora*_1958
6	*Ceiba erianthos*_1965
7	*Ceiba jasminodora*_1968
8	*Ceiba lupuna*_1969, *Ceiba insignis*_KM488629, *Ceiba lupuna*u_MW192507
9	*Ceiba ventricosa*_1983
10	*Ceiba acuminata*_HQ658385
11	*Ceiba aessclufolia*_HQ658384
12	*Ceiba erianthos*a_KM453167
13	*Ceiba insignis*_MW192513
14	*Ceiba jasminodora*_KM453168
15	*Ceiba pentandra*_AY635494, *Ceiba pentandra*_AY635497, *Ceiba pentandra*_DQ284820, *Ceiba pentandra*_DQ284821, *Ceiba pentandra*_DQ284828, *Ceiba pentandra*_DQ284834, *Ceiba pentandra*_DQ284846
16	*Ceiba pentandra*_AY635496, *Ceiba pentandra*_DQ284848, *Ceiba pentandra*_DQ284850, *Ceiba pentandra*_EF432375
17	*Ceiba pentandra*_AY635499
18	*Ceiba pentandra*_AY635501
19	*Ceiba pentandra*_EF432374
20	*Ceiba pubiflora*_MW192504
21	*Ceiba rubriflora*_KM453171
22	*Ceiba schottii*_HQ658389
23	*Ceiba speciosa*_HQ658388

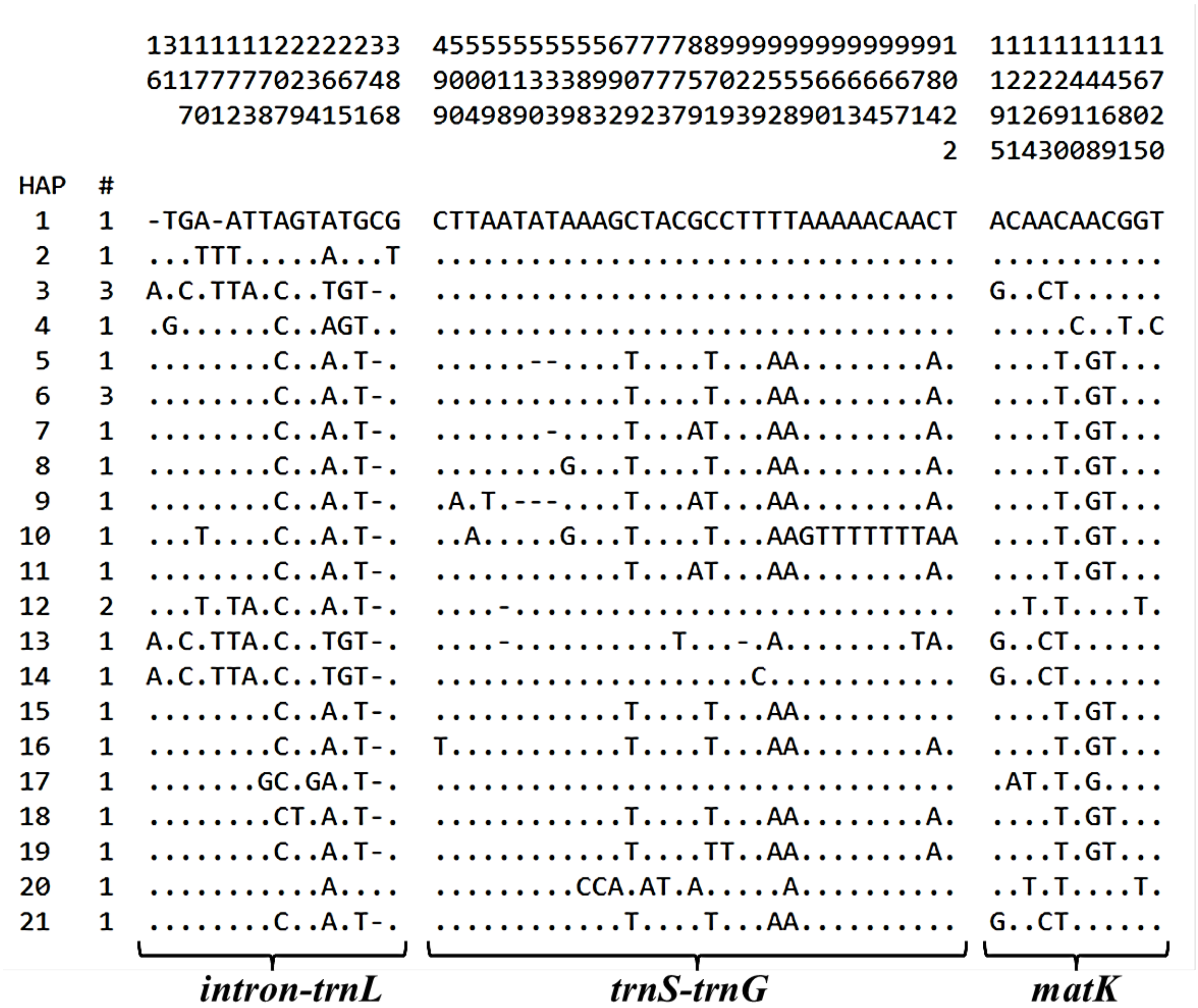


```
          111111122333444444456666778991111111111111111111  1111111111111111111111222222222222222  222222222223344444444445555555555555555555666
          024567807237012457971569131890000001111111222222  3333333445567788899999001111223333445  556667777780702556777891133345677788888888000
                                       0245680124689056789  0123589272302847824568792457162359794  894690236716903240359570906956667902345679234
HAP  #
  1  3    CGCCAAACTGC-GTAAGCTGGGATCCGCGTCGGACGCGCGCACTGGA-  TGGTCCGACACCGC-TAAGATACGTCCCCCGGCGCTT  TTCCAAAAAGCCCCTTC--CGGACTCTACGGCAGCT-CGCGTGCT
  2  13   .....................................A..........  .....................................  .............................................
  3  1    ..............G.A...............TG...A....T....G  .-...................................  C...G........G...CTT..T....................T.
  4  12   ................A.........T..C.......A.....C...G  .-.........T.............T...........  C........C.......C.....A...G..A............T.
  5  1    T-..GG.T.T.A.C..A.AAA.T........T.G.C.A..TG.CC..G  .-..TG.T.CA.AT.CTTTGAGTT..ATATT.TAAG.  C.GG.TT.T.T.TGC-TC.TAT...T.G...G.TGC.-CGTC..C
  6  1    ...............G.....................A........-.  ---C..........A.........C............  C...........T....C........A.G................
  7  1    ...............G....................AA.T......-.  ---C...........................T.....  C...........T.......................T.....T..
  8  3    .................T...................A..........  .....................................  .............................................
  9  1    T-..GG.TCT.A.C..A.AAA.T........T.G.C.A..TG.CC..G  .-..TG.T.CA.AT.CTTTGAGTT..ATATT.TAAG.  C.GG.TT.T.T.TGC-TC.TAT...T.G...G.TGC.-CGTC..C
 10  1    ............T...A...............TG...A.....C...G  .-...........................T......A  C...G..T..T......C......GT...................
 11  1    ............T...A...............TG...AT........G  .-...........................T.......  C...G..T..T......C......GT...................
 12  1    ..G..................................A..........  .....................................  ...................................CT........
 13  1    .....................................A.......--.  ---..................................  .............................................
 14  1    ...A.CG........G.......G............AA.T......-.  ---C.................................  C...........T.......................T.....T..
 15  7    ................A.........T..C.......A.....C...G  .-.........T.........................  C........C.......C.....A...G..A............T.
 16  4    ................A.........T..C.......A.....C...G  .-....T....T.............T...........  C........C.......C.....A...G..A............T.
 17  1    ...T............A...............TG...A....T....G  .-...................................  C...G....C.......C.....A...G..A............T.
 18  1    ................A.........T..C.......A.....C...G  .-.........T.............T...........  C...G........G...CTT..T....................T.
 19  1    ................A.......ATT..C.......A.....C...G  .-....T....T.............T...........  C........C.......C.....A...G..A............T.
 20  1    .................T.........G..G......A..........  .....................................  .............................A...............
 21  1    ............................T........A..........  .....................................  ....................................T........
 22  1    .........TT.....A...............TGT..A.........G  .-......T.............T..............  CA..G......T.....C..............G............
 23  1    .....................A...............A....T.....  .....................................  .............................................
```

Figure S1: Sequence alignment of the variable sites in the ITS region that define the 23 haplotypes of *Ceiba*. Each fragment spans 606 bases. Dots indicates similarity to haplotype A and hyphens indicate gaps. Numbers on top indicate the nucleotide position with haplotype A acting as the reference sequence during alignment. The numbers of occurrences of each haplotype (#) were as indicated.

Table S5: Components of the chloroplast gene sequences-derived haplotype network

Haplotype	Specimens
1	*Ceiba acuminata*
2	*Ceiba aesclufolia*
3	*Ceiba pentandra*, *Ceiba speciosa*_1930, *Ceiba samuama*_1941
4	*Ceiba schottii*
5	*Ceiba glaziovii*_1500
6	*Ceiba glaziovii*_1510, *Ceiba erianthos*_1965, *Ceiba erianthos*_1985
7	*Ceiba pubiflora*_1706
8	*Ceiba pubiflora*_1743
9	*Ceiba pubiflora*_1795
10	*Ceiba rubriflora*_1855
11	*Ceiba rubriflora*_1881
12	*Ceiba jasminidora*_1904, *Ceiba jasminidora*_1908
13	*Ceiba samuama*_1921
14	*Ceiba speciosa*_1934
15	*Ceiba crispliflora*_1958, *Ceiba crispliflora*
16	*Ceiba erianthos*_1960
17	*Ceiba jasminidora*_1968
18	*Ceiba lupuna*_1969
19	*Ceiba pentandra*_1970
20	*Ceiba insignis*_3216
21	*Ceiba speciosa*_3218



```
        1311111122222233  455555555556777788999999999999991  11111111111
        6117777702366748  900011333899077757022555666666780  12222444567
          70123879415168  904989039832923791939289013457142  91269116802
                                                          2  51430089150
HAP  #
 1   1  -TGA-ATTAGTATGCG  CTTAATATAAAGCTACGCCTTTTAAAAACAACT  ACAACAACGGT
 2   1  ...TTT.....A...T  .................................  ...........
 3   3  A.C.TTA.C..TGT-.  .................................  G..CT......
 4   1  .G......C..AGT..  .................................  .....C..T.C
 5   1  ........C..A.T-.  ......--....T....T...AA........A.  ....T.GT...
 6   3  ........C..A.T-.  ............T....T...AA........A.  ....T.GT...
 7   1  ........C..A.T-.  .......-....T...AT...AA........A.  ....T.GT...
 8   1  ........C..A.T-.  ........G...T....T...AA........A.  ....T.GT...
 9   1  ........C..A.T-.  .A.T.---....T...AT...AA........A.  ....T.GT...
10   1  ...T....C..A.T-.  ..A.....G...T....T...AAGTTTTTTTAA  ....T.GT...
11   1  ........C..A.T-.  ............T...AT...AA........A.  ....T.GT...
12   2  ...T.TA.C..A.T-.  ....-............................  ..T.T....T.
13   1  A.C.TTA.C..TGT-.  ....-..........T...-.A........TA.  G..CT......
14   1  A.C.TTA.C..TGT-.  ....................C............  G..CT......
15   1  ........C..A.T-.  ............T....T...AA..........  ....T.GT...
16   1  ........C..A.T-.  T...........T....T...AA........A.  ....T.GT...
17   1  .......GC.GA.T-.  .................................  .AT.T.G....
18   1  ........CT.A.T-.  ............T....T...AA........A.  ....T.GT...
19   1  ........C..A.T-.  ............T....TT..AA........A.  ....T.GT...
20   1  ...........A....  .........CCA.AT.A.....A..........  ..T.T....T.
21   1  ........C..A.T-.  ............T....T...AA..........  G..CT......
```

Figure S2: Sequence alignment of the variable sites in the chloroplast genes that define the 21 haplotypes of *Ceiba*. Each fragment spans 1813 bases. Dots indicates similarity to haplotype A and hyphens indicate gaps. Numbers on top indicate the nucleotide position with haplotype A acting as the reference sequence during alignment. The numbers of occurrences of each haplotype (#) were as indicated.