

1 **TITLE PAGE**

2

3 Pezzini *et al.* Evolutionary History of Tropical Dry Forest

4

5 **Research article: Phylogeny and biogeography of *Ceiba* Mill. (Malvaceae, Bombacoideae)**

6

7 Flávia Fonseca Pezzini<sup>1,2,8</sup>, Kyle G. Dexter<sup>3</sup>, Jefferson G. de Carvalho-Sobrinho<sup>4</sup>, Catherine A. Kidner<sup>1,2</sup>,

8 James A. Nicholls<sup>5</sup>, Luciano P. de Queiroz<sup>6</sup>, R. Toby Pennington<sup>1,7</sup>

9

10 <sup>1</sup>Royal Botanic Garden Edinburgh, Edinburgh, United Kingdom

11 <sup>2</sup>School of Biological Sciences, University of Edinburgh, Edinburgh, United Kingdom

12 <sup>3</sup>School of GeoSciences, University of Edinburgh, Edinburgh, United Kingdom.

13 <sup>4</sup>Colegiado de Ciências Biológicas, Universidade Federal do Vale do São Francisco, Petrolina, Brazil

14 <sup>5</sup> Australian National Insect Collection, CSIRO, Acton, Australia

15 <sup>6</sup>Herbario, Universidade Estadual de Feira de Santana, Feira de Santana, Brazil

16 <sup>7</sup> Geography, College of Life and Environmental Sciences, University of Exeter, Exeter, United Kingdom

17 <sup>8</sup> Corresponding author: [flaviapezzini@gmail.com](mailto:flaviapezzini@gmail.com) | 20a Inverleith Row Edinburgh, EH3 5LR, UK

18

19 **ABSTRACT**

20 The Neotropics is the most species-rich area in the world and the mechanisms that generated and

21 maintain its biodiversity are still debated. This paper contributes to the debate by investigating

22 the evolutionary and biogeographic history of the genus *Ceiba* Mill. (Malvaceae:

23 Bombacoideae). *Ceiba* comprises 18 mostly neotropical species endemic to two major biomes,

24 seasonally dry tropical forests (SDTFs) and rain forests and its species are one of the most

25 characteristic elements of neotropical SDTF, one of the most threatened biomes in the tropics.  
26 Phylogenetic analyses of DNA sequence data from the nuclear ribosomal internal transcribed  
27 spacers (ITS) for 30 accessions representing 14 species of *Ceiba* recovered the genus as  
28 monophyletic and showed geographical and ecological structure in three main clades: (i) a rain  
29 forest lineage of nine accessions of *C. pentandra* sister to the remaining species; (ii) a highly  
30 supported clade composed of *C. schottii* and *C. aesculifolia* from Central American and Mexican  
31 SDTF plus two accessions of *C. samauma* from inter Andean valleys from Peru; and (iii) a  
32 highly supported South American SDTF clade including 10 species showing little sequence  
33 variation. Within this South American clade, no species represented by multiple accessions were  
34 resolved as monophyletic. We demonstrate that the patterns of species age, monophyly and  
35 geographical structure previously reported for SDTF species within the Leguminosae family are  
36 not shared by *Ceiba*, suggesting that further phylogenetic studies of unrelated groups are  
37 required.

## 38 HIGHLIGHTS

- 39
- 40 • This paper provides a well sampled phylogeny of the iconic genus *Ceiba*, one of the most  
41 characteristic tree genera of neotropical Seasonally Dry Tropical Forest (SDTF).
  - 42 • There is a clear phylogenetic signal for biome preference and geographic structure in  
43 *Ceiba*.
  - 44 • We estimate a mid Miocene origin for *Ceiba*, with the stem node age of the genus  
45 estimated at 21.1 Ma and the crown node age at 12.7 Ma.
  - 46 • *Ceiba* species have young stem ages in the SDTF clade but old stem ages in rain forest  
47 species.

- 48       • Patterns of species age, monophyly, ecological and geographical structure reported for  
49       SDTF species are only partially shared by *Ceiba*, an iconic genus of neotropical SDTF.

50

## 51 **KEY WORDS**

52

53 Seasonally Dry Tropical Forests; rain forest; radiation; neotropics; Caatinga.

54

## 55 **INTRODUCTION**

56

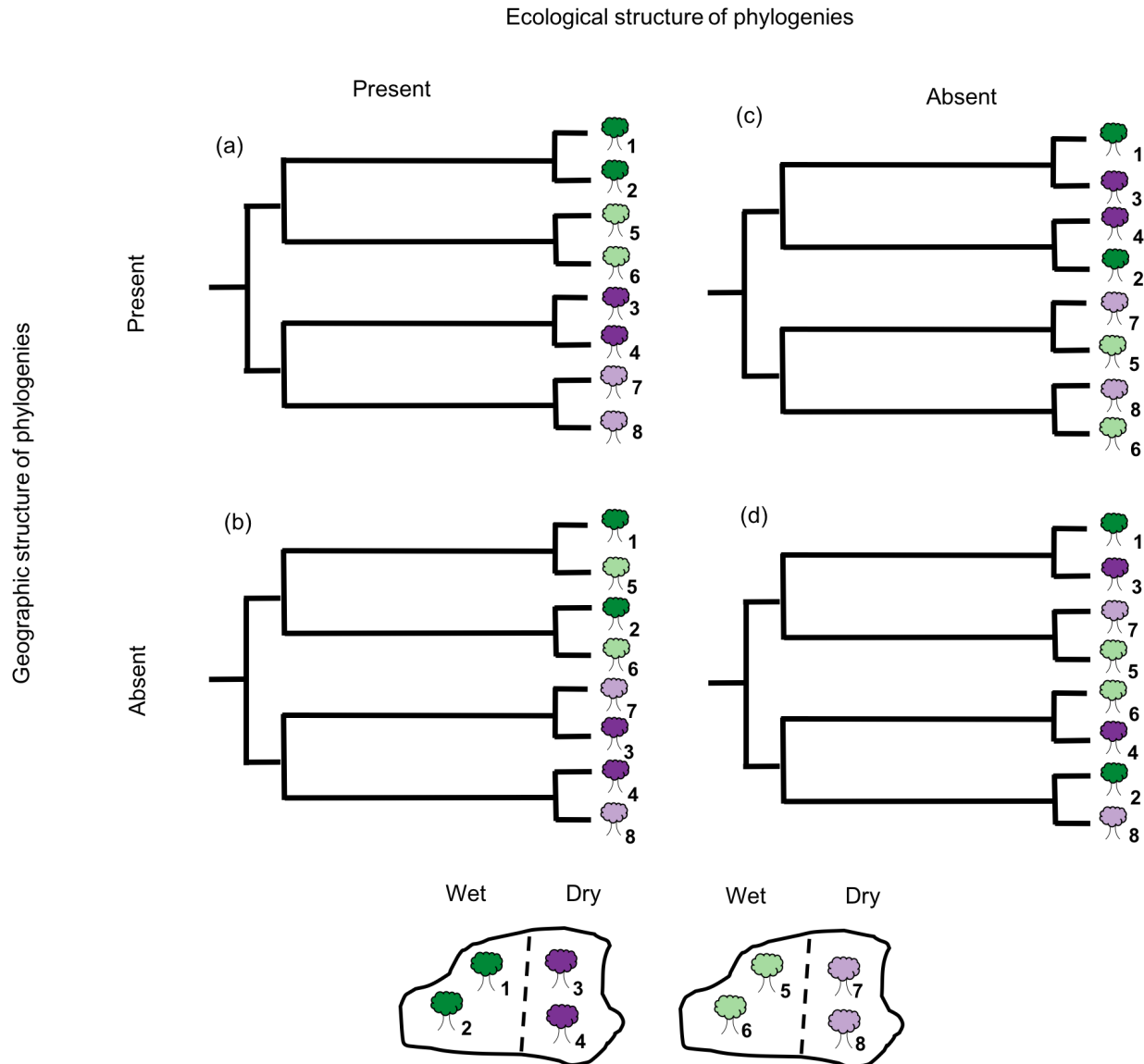
57 THE NEOTROPICS IS THE MOST SPECIES-RICH REGION IN THE WORLD AND THE MECHANISMS THAT  
58 generated and maintain its biodiversity are under constant discussion. Through evolutionary  
59 time, the Neotropics have been climatically and geologically dynamic, resulting in a diversity of  
60 biomes, from deserts to tropical rain forests (Hughes et al. 2013, Rangel et al. 2018). To  
61 understand the history and dynamics of those biomes, molecular phylogenetic and  
62 phylogeographic approaches have been used, because the relationships of taxa allow inferences  
63 to be made of the historical relationships amongst biomes and areas (Pennington et al. 2006). In  
64 recent years the dichotomy regarding the ‘cradle’ vs. ‘museum’ debate (Stebbins 1974)  
65 explaining neotropical diversity has given way to a more nuanced approach, considering that  
66 plant diversification patterns may be recent, old, slow or rapid, even within individual clades  
67 (Hughes et al. 2013, Koenen et al. 2015). As suggested in the literature more than 10 years ago  
68 (Wiens and Donoghue 2004, Pennington et al. 2006), this heterogeneity in diversification timing  
69 and rate within and among clades may be related not only to climatic and geological events, but  
70 also to the age and ecological differences of the biomes. For example, geologically old biomes

71 (e.g., rain forest) are likely to have provided lineages that colonised newer biomes (e.g.,  
72 savannas) and the relative difficulty of evolving adaptations such as drought tolerance or the  
73 ability to survive fire might determine whether a lineage can adapt to a new biome (niche  
74 evolution; Simon et al. 2009, Pennington & Lavin 2016) or remains confined to the same biome  
75 (niche conservatism) over evolutionary timescales (Crisp et al. 2009).

76 Clades endemic to two of the major Neotropical biomes, seasonally dry tropical forests  
77 (SDTFs) and rain forests, give good examples of different and distinctive phylogenetic and  
78 biogeographic patterns, suggesting an interaction of ecology and phylogeny over evolutionary  
79 timescales (Pennington et al. 2011, Pennington and Lavin 2016, Dexter et al. 2017). SDTFs  
80 occur on fertile soils and are characterized by the absence of fire adaptation in the flora and a  
81 predominantly continuous tree canopy, which becomes more open in the drier sites, with plants  
82 shedding up to 90–95% of their leaves during the five to six month long dry season (Murphy and  
83 Lugo 1986, Pennington et al. 2009). This biome has been one of the least studied, but is one of  
84 the most threatened in the tropics (Miles et al. 2006, DRYFLOR 2016). It occurs in disjunct  
85 areas throughout the Neotropics and has high beta-diversity and plant species endemism  
86 (Pennington et al., 2009; DRYFLOR, 2016). Leguminosae and Bignoniaceae are often the most  
87 species rich and dominant families in SDTF, but species from the Bombacoideae clade  
88 (Malvaceae), the subject of this paper, are often common and distinctive.

89 SDTF-confined clades contain species that often resolve as monophyletic and with old  
90 stem ages in DNA-sequence-based phylogenies (Pennington and Lavin 2016). In addition, the  
91 geographically structured phylogenetic pattern characteristic of clades in this biome suggests  
92 dispersal-limitation caused by the stable ecological conditions of the biome maintained over long  
93 evolutionary timescales (Pennington et al. 2010, Hughes et al. 2013). By contrast, tree clades

94 confined to the Amazon rain forest, the largest expanse of tropical forest in the world, are  
 95 suggested to contain more non-monophyletic species, more species with young stem ages and  
 96 clades that lack geographical phylogenetic structure (Dexter et al. 2017). These rain forest  
 97 patterns might be explained by frequent dispersal and subsequent successful colonization over  
 98 evolutionary timescales (Pennington & Lavin, 2015) (Fig. 1).



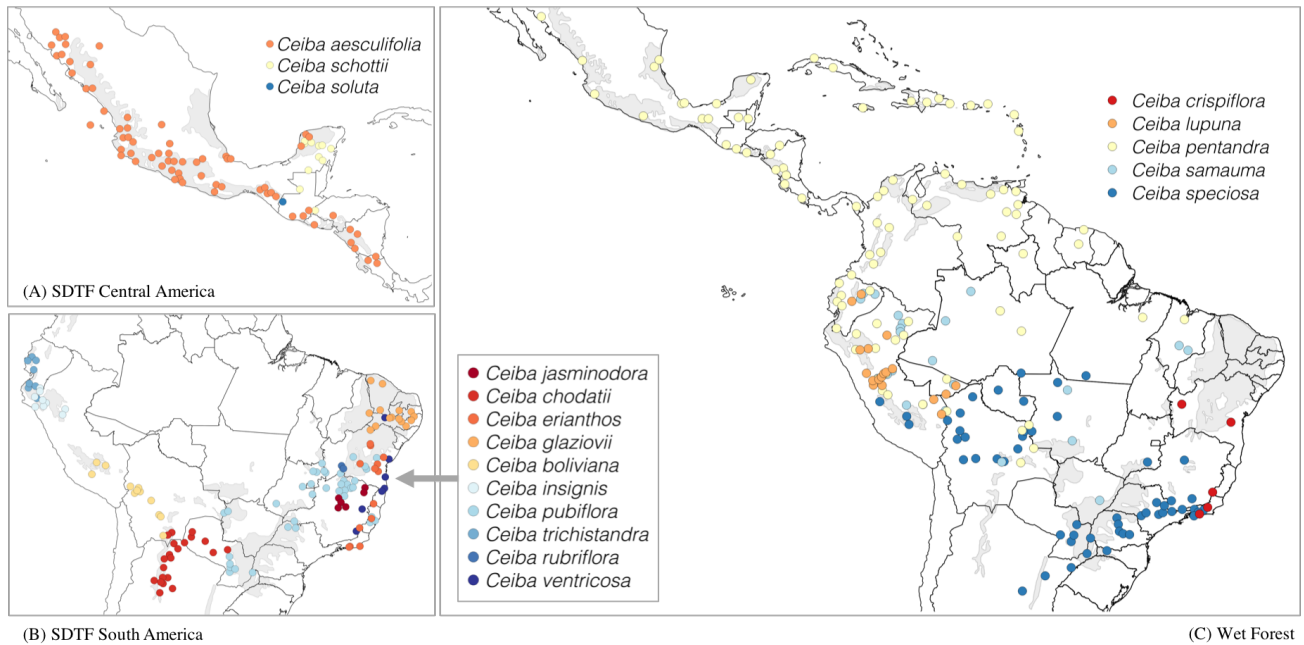
99  
 100 Figure 1. Two hypothetical islands, each with an area of seasonally dry tropical forest and rain forest. In  
 101 total eight different species occur in the two biomes and the different islands and are represented by

102 different colours (dark green species 1, dark green species 2, dark purple 3, etc.). Hypothetical  
103 phylogenies showing patterns of presence or absence of geographical and ecological structures (after  
104 Graham & Fine 2008).

105  
106 The neotropical genus *Ceiba* Mill. (Malvaceae: Bombacoideae) comprises 18 species  
107 grouped into taxonomic sections *Ceiba* and *Campylanthera* (Schott & Endl.) K. Schum. based on  
108 morphological characters of pollen and staminal appendages. It is one of the most characteristic  
109 elements of many neotropical SDTFs. However, it also contains species confined to the Amazon  
110 rain forest and is therefore a good case study to investigate biome-specific differences in the  
111 nature of species and their diversification trajectories.

112 *Ceiba* species have digitate leaves, aculeate spines on the trunk and branches and can  
113 vary from 50 m canopy emergents in seasonally flooded várzea forests in the Amazon (*C.*  
114 *pentandra*) to 2 m treelets on rocky outcrops (campos rupestres) in Minas Gerais, Brazil (*C.*  
115 *jasminodora*). In some species (*C. chodatii*, *C. pubiflora*, *C. glaziovii*, *C. speciosa*) the trunk can  
116 be ventricose (swollen), explaining its vernacular names barriguda (“swollen belly”; Brazil) and  
117 palo borracho (“drunken tree”; Peru). Most species are deciduous and flower when leafless. They  
118 occur mostly in SDTF, with the exception of the widespread *C. samauma*, *C. speciosa* and *C.*  
119 *pentandra* that also occur in more humid environments, and *C. lupuna*, which is the only species  
120 restricted to rain forests (Fig. 2). On average, each of the thirteen SDTF species have a narrower  
121 geographical distribution when compared to the five rain-forest-inhabiting species (Fig. 2).

122



123 (A) SDTF Central America (B) SDTF South America (C) Wet Forest

124 Figure 2. Distribution of 18 *Ceiba* species in three maps: (A) three species from SDTFs in Central  
125 America and North America, (B) ten species from SDTFs in South America and (C) five species from  
126 rain forests from Latin America. Grey areas represent the distribution of SDTF following DRYFLOR  
127 (2016). Occurrence records adapted from Gibbs and Semir (2003).

128

129 Previous Bayesian analyses of sequence data from the nuclear ribosomal internal and  
130 external transcribed spacers (ITS and ETS) and plastid markers (*matK*, *trnL-F*, *trnS-trnG*) for 13  
131 species recovered *Ceiba* as monophyletic and sister to *Neobuchia paullinae* (Duarte et al. 2011,  
132 Carvalho-Sobrinho et al. 2016). Together with *Spirotheca*, *Pochota fendleri* sensu Alverson and  
133 Duarte (2015), and *Pseudobombax*, these taxa form the well supported “striated bark” clade  
134 (Carvalho-Sobrinho et al. 2016). However, relationships within *Ceiba* were poorly resolved and  
135 only one individual per species was included in the phylogeny. *Ceiba* has a historically complex  
136 taxonomy with species boundaries still confused, which is aggravated by the fact that herbarium  
137 specimens are often incomplete because individuals produce flowers and fruits when leafless.

138 Therefore, a well sampled phylogeny with multiple accessions per species could be a useful tool  
139 to explore the nature of species in *Ceiba*.

140 This paper investigates the evolutionary history of *Ceiba* species. We aim to assess  
141 whether the *Ceiba* phylogeny is geographically or ecologically structured and if species confined  
142 to SDTFs are resolved differently in the phylogeny as compared with rain forest species (i.e.,  
143 monophyletic on long stem lineages).

144

## 145 **METHODS**

146

147 TAXON SAMPLING - We present the best sampled phylogeny of the genus *Ceiba* to date, covering  
148 30 accessions representing 14 of the 18 species described for the genus. Critically, this study is  
149 the first to sample multiple individuals per species for six species (Table S1). As outgroups, we  
150 included 10 accessions representing species of the closest sister clades (Carvalho-Sobrinho et al.  
151 2016): *Pseudobombax*, *Spirotheca*, *Eriotheca*, and *Pochota fendleri*. The full data set represents  
152 a combination of new sequence data from field surveys as well as from herbarium specimens,  
153 doubling the number of accessions of *Ceiba* in relation to the previous study by Carvalho-  
154 Sobrinho et al. (2016).

155

156 DNA SEQUENCE DATA - We used the ITS region to investigate species relationships in *Ceiba*. In  
157 Bombacoideae, this region has been widely explored to help elucidate relationships among  
158 genera and species (Baum et al. 1998, Duarte et al. 2011, Carvalho-Sobrinho et al. 2016), and to  
159 investigate genetic structure among populations (Dick et al. 2007). Previous systematic studies in  
160 Bombacoideae used a combination of nuclear (ITS, ETS) and plastid (matK, trnS-trnG and trnL-



161 trnF) markers and ITS had the highest number of informative characters (Duarte et al. 2011,  
162 Carvalho-Sobrinho et al. 2016). In spite of having drawbacks related to potential paralogous  
163 copies (Álvarez 2003, Buckler et al. 1997), ITS can still play an important role in the  
164 investigation of species relationships if analysed carefully, for example identifying pseudogenes  
165 and assessing orthology in the case of intra-individual polymorphism (Bailey 2003, Feliner and  
166 Rosselló 2007).

167  
168 Genomic DNA extraction was performed for 36 herbarium and silica-gel dried leaf  
169 samples using Qiagen DNeasy Plant Mini Kits following the manufacturer's protocol, with the  
170 following changes: twice the volume of buffer AP1 in addition to a small amount of PVPP  
171 (polyvinyl polypyrrolidone) added at the lyse step followed by an incubation of 30 minutes;  
172 addition of 1uL of Riboshredder in the lysate solution followed by incubation at 37°C for 20  
173 minutes; addition of twice the volume of buffer P3; and final elution in 46 uL of EB buffer run  
174 through the column twice to increase yield. Each 20 uL PCR amplification reaction contained 0.5  
175 uL of template, 2 uL of dNTPs (2 mM), 2 uL of 10x reaction buffer, 1 uL of MgCl<sub>2</sub> (50 mM),  
176 0.65 uL of both forward primer and reverse primer solutions (10 uM), 0.1 uL of Taq polymerase,  
177 4 uL of CES buffer and 9.1 uL of ddH<sub>2</sub>O. Amplification followed the same procedure described  
178 in Carvalho-Sobrinho et al. (2016). Samples were submitted to the Edinburgh Genomics  
179 laboratory at the University of Edinburgh for sequencing. For low quality sequences, we tested  
180 variations of the protocol (e.g. diminishing the amount of template in the PCR reaction or  
181 varying the sequencing primer). High quality sequences were recovered for 13 out of the 36  
182 samples from which DNA was extracted for the ITS region.

183 All the inter-accession polymorphisms detected were validated visually by checking the  
184 electropherograms. Sequences were edited with Sequencher 5.4.1 (Gene Codes Corp., Ann  
185 Arbor, Michigan) and alignments were performed manually in Mesquite (Maddison and  
186 Maddison 2015). We investigated the potential presence of ITS pseudogenes by conducting  
187 phylogenetic analysis in separate partitions representing the 5.8S (conserved region) and the ITS  
188 1 and ITS 2 regions (fast evolving regions). In pseudogenes, the conserved and the fast evolving  
189 regions are expected to show similar rate of evolution whereas in functional genes the conserved  
190 region is expected to have a lower rate of evolution compared to the fast evolving regions  
191 (Bailey et al. 2003). We assigned partitions by comparison with the annotated accession of  
192 *Pseudobombax tomentosum* (GenBank KM453206), and checked for differences in rates of  
193 evolution in the partition scheme proposed with PartitionFinder2 version 2.1.1 (Lanfear et al.  
194 2017) using PhyML version 3.0 (Guindon et al. 2010) and the greedy search algorithm (Lanfear  
195 et al. 2012).

196  
197 PHYLOGENETIC ANALYSIS AND MOLECULAR DATING - We implemented maximum likelihood  
198 (ML) and Bayesian Inference (BI) analysis. To determine best fitting model of sequence  
199 evolution for each of the three partitions, we used PartitionFinder2 version 2.1.1 (Lanfear et al.  
200 2017) in the ML analysis and the reversible jump model selection (RB) implemented in BEAST2  
201 version 2.5.1 (Bouckaert et al. 2019) under the Bayesian framework. IQ-TREE version 2.0.3  
202 (Nguyen et al. 2015, Minh et al. 2020) was used to run the ML analysis with 1,000 bootstrap  
203 replicates and using the partition model option (-p) (Chernomor et al. 2016) with substitution  
204 models specified as follows: GTR+G for ITS1, K80+G for 5.8S and HKY+G for ITS2, as  
205 inferred with PartitionFinder2 (see above). BEAST2 version 2.5.1 (Bouckaert et al. 2019) was

206 used to perform BI analysis and temporally calibrate the phylogeny. Different combinations of  
207 relaxed clock models (Uncorrelated Exponential Distribution - UCED and Uncorrelated Log  
208 Normal Distribution - UCLD (Drummond et al. 2006)) and tree priors (Yule and Birth-Death)  
209 were compared. Few studies have objectively contrasted the effect of different models in the  
210 divergence time estimation, and a poorly inferred time-calibrated phylogeny can have serious  
211 consequences for our understanding of diversification history of lineages (Louca and Pennell  
212 2020). For example, different tree priors resulted in impressive differences in age estimation for  
213 cycads, with the Yule prior inferring ages three times older than the Birth-Death prior  
214 (Condamine et al. 2015). In Bayesian analysis, the most suitable model can be selected by  
215 comparing the Bayes Factor (BF). The BF is equal to the ratio of the Marginal Likelihood  
216 Estimate (MLE) of two models ( $BF = MLE1 / MLE2$ ) or to the difference of MLEs in log space  
217 ( $\log(BF) = \log(MLE1) - \log(MLE2)$ ). Positive values of BF would favour MLE1, and different  
218 values have different strengths. Values above five indicates that one model is significantly  
219 favoured over the other (Kass and Raftery 1995, Condamine et al. 2015), values above 20  
220 indicate strong support and values above 150 overwhelming support. We estimated the Marginal  
221 Likelihood using the Nested Sampling algorithm (Skilling 2006) implemented in the NS package  
222 version 1.0.4 (Russel et al. 2019) for BEAST2 version 2.5.1 with 60 particles and 10,000 chain  
223 length. The NS package also calculates the standard deviation (SD) of the estimated Marginal  
224 Likelihood, which allows us to have confidence in the BF values calculated. For each  
225 combination of priors, we ran two independent runs of 10 million generations, sampled every  
226 1,000 generations and visually inspected convergence of MCMC and ensured effective sample  
227 size  $> 200$  for all parameters of each run using Tracer v1.7.1 (Rambaut et al. 2018). Resulting  
228 trees and log files from each run were combined using LogCombiner with a burn-in of 10% and

229 the Maximum Clade Credibility Tree was summarized in TreeAnnotator with node heights as  
230 mean heights. We used r8s (Sanderson 2004) to implement the penalised likelihood method  
231 (Sanderson 2002) and calculate substitutions rates. We used the phylogram derived from the ML  
232 analysis as input and conducted a cross-validation analysis to find the best smoothing parameter.

233 We used the fossil flower of *Eriotheca prima* (Duarte 1974) from the middle to late  
234 Eocene (de Lima and Salard-Chebouldaef 1981) of Brazil as a primary calibration for our  
235 BEAST2 analysis. The flower was identified as *Eriotheca* based on its small size (*Bombacopsis*  
236 and *Pachira* have larger flowers) and androecium organisation, which is a synapomorphy for the  
237 extant species of the genus (Robyns 1963, Duarte et al. 2011, Carvalho-Sobrinho et al. 2016). In  
238 previous studies, *Eriotheca* was resolved as sister to a clade comprising *Pseudobombax*,  
239 *Spirotheca*, *Ceiba*, and *Pochota fendleri* (Duarte et al. 2011, Carvalho-Sobrinho et al. 2016).  
240 Because the dating of this fossil is imprecise (middle to late Eocene: 33-56 million years old  
241 (Ma), we assigned the offset age of 33 Ma as a minimum age to the stem node (Renner 2005,  
242 Pennington et al. 2006) of *Eriotheca* (the crown node of the clade comprising *Eriotheca*,  
243 *Spirotheca*, *Pseudobombax*, *Pochotoa fendleri* and *Ceiba*), which is equivalent to the root node  
244 of the outgroups and ingroup of this study. We assigned a log-normal distribution with a mean of  
245 1.542 and standard deviation of 1.5. This fossil calibration is conservative, with 95% of the prior  
246 distribution comprised between 33 and 47 Ma and thus the ages estimated here are considered  
247 minimum ages estimates. In order to explore the effects of using the medium and maximum ages  
248 of the *Eriotheca* fossil on phylogenetic age estimates, we also ran analyses assigning minimum  
249 ages of 47 Ma and 56 Ma to the *Eriotheca* stem (Figs. S1 and S2). We followed the dates on the  
250 Geologic Time Scale v. 5.0 (Gradstein et al. 2012).

251

252 PHYLOGENETIC SIGNAL TEST.- We tested for strength of phylogenetic signal for the binary traits  
253 related to ecology (rain vs dry forests) and geography (Central and North America vs South  
254 America) using the *D value* proposed by Fritz and Purvis (2010), and implemented using the  
255 Caper package (v. 1.0.1) (Orme 2013) in R, with 5,000 permutations. Under a null model of  
256 Brownian motion evolution of a binary trait, *D* has an expected value of 0. A negative *D* value  
257 indicates a strongly clustered phylogenetic pattern for a given binary trait (perhaps due to some  
258 process of evolutionary constraint), a value of one indicates a completely random pattern with  
259 respect to the phylogeny (i.e. no correlation between phylogeny and the trait at all) and values  
260 above one indicate an overdispersed phylogenetic pattern (perhaps due to divergent selection).  
261 We assigned species to ecological and geographical categories following Gibbs and Semir  
262 (2003) (Table S1). Despite occurring mainly in rain forests, *Ceiba speciosa* is also recorded in  
263 dry semi-deciduous woodland (Figure 2). To check for possible bias in the ecological affinity of  
264 *C. speciosa*, we also conducted a phylogenetic signal analysis assigning this species to dry forest,  
265 observing no difference when comparing to the analysis run assigning the species to rain forest.

266

## 267 **RESULTS**

268

269 The total length of the aligned sequences was 814 nucleotides, of which 283 were variable and  
270 178 (22%) were parsimony-informative characters. The ML and BI trees showed congruent  
271 topologies (Figs. 3 and 4). *Ceiba* was strongly supported as monophyletic, with posterior  
272 probability (pp) = 1 and bootstrap value = 100 and was recovered as sister to *Pseudobombax*  
273 (Fig. 3). The UCLD and UCED clock models and Yule and Birth-Death tree models inferred  
274 similar crown and stem ages with overlapping credibility intervals (95% Highest Posterior

275 Density (HDP), Table 1). For the UCLD clock prior, the Bayes Factor value of 6.15 support the  
276 Yule tree prior as the most suitable tree model. For the UCED clock prior, the Bayes Factor of  
277 1.15 indicates that the neither tree model is favoured (Table 1). Therefore, results shown  
278 onwards for Bayesian analysis are those inferred using the UCLD clock model and the Yule tree  
279 model. Using the 33 Ma fossil calibration, the stem node age of *Ceiba* is 21.1 (14.7-27.1 [95%  
280 HPD]) Ma, the crown node age is 12.7 (8.2-17.6 [95% HPD]) Ma (Table 1, Fig. 4), and  
281 substitution rates estimated in  $1.592 \times 10^{-9}$  substitutions per site per year (s/s/y). *Ceiba* shows  
282 slow substitution rates for ITS when compared to other tropical tree species. For example, for the  
283 rain forest tree genus *Inga* (Leguminosae), substitution rates have been estimated in  $2.34 \times 10^{-9}$   
284 s/s/y (Richardson et al. 2001) and in  $7.1-7.9 \times 10^{-9}$  s/s/y (Lavin 2006), and for the dry forest  
285 genus *Coursetia* (Leguminosae) in  $5.0-8.2 \times 10^{-9}$  (Lavin 2006).

286

287

288

289

290

291

292

293

294

295

296

297

298 Table 1. Absolute ages estimate for ten nodes under different tree (Yule and Birth-Death) and clock priors  
 299 (UCLD – Uncorrelated Lognormal Distribution and UCED – Uncorrelated Exponential  
 300 Distribution). Ages are reported in million years as mean ages followed by the 95% Highest Posterior  
 301 Density (HDP) as a result of the combined independent runs for each tree and clock priors. MLE (SD):  
 302 marginal likelihood estimated followed by standard deviation; BF: Bayes factor calculated as the  
 303 difference between the MLE of the Yule and the Birth-Death prior for each clock prior UCLD and  
 304 UCED. Values above five indicates that one model is significantly favoured over the other.

Clock prior	UCLD		UCED	
	Yule	Birth-death	Yule	Birth-death
<b>MLE (SD)</b>	-3844.31 (1.85)	-3838.15 (1.79)	-3838.74 (1.88)	-3837.62 (1.76)
<b>BF</b>	6.15		1.15	
<b><i>Ceiba</i> crown</b>	12.7 (8.2-17.6)	10.9 (6.5-15.7)	13.5 (8.6-18.8)	11.5 (6.1-17.4)
<b><i>Ceiba</i> stem</b>	21.1 (14.7-27.1)	20.3 (13.5-26.9)	21.0 (14.4-27.5)	19.1 (11.3-26.9)
<b>SDTF SA crown</b>	8.6 (5.0-12.5)	7.0 (3.6-10.7)	9.6 (5.6-14.1)	7.8 (3.8-12.5)
<b>SDTF SA stem</b>	11.2 (7.2-15.6)	9.5 (5.6-13.9)	12.0 (7.5-16.7)	10.0 (5.2-15.4)
<b><i>C. aesculifolia</i> crown</b>	1.5 (0.2-3.2)	1.2 (0.2-2.3)	1.7 (0.3-3.9)	1.3 (0.2-2.9)
<b><i>C. aesculifolia</i> stem</b>	6.3 (3.1-9.8)	5.1 (2.1-8.3)	7.6 (3.7-12.0)	4.7 (1.7-8.3)
<b><i>C. samauma</i> crown</b>	0.4 (0.0001-1.4)	0.3 (0.0001-1.0)	0.5 (0.0001-1.7)	0.3 (0.0001-1.2)
<b><i>C. samauma</i> stem</b>	6.3 (3.1-9.8)	5.1 (2.1-8.3)	6.3 (2.6-10.2)	6.1 (2.4-10.3)
<b><i>C. pentandra</i> crown</b>	3.9 (1.6-6.5)	3.0 (1.1-5.2)	5.5 (2.3-9.2)	4.2 (1.5-7.5)
<b><i>C. pentandra</i> stem</b>	12.7 (8.2-17.6)	10.9 (6.5-15.7)	13.5 (8.6-18.9)	11.5 (6.1-17.4)

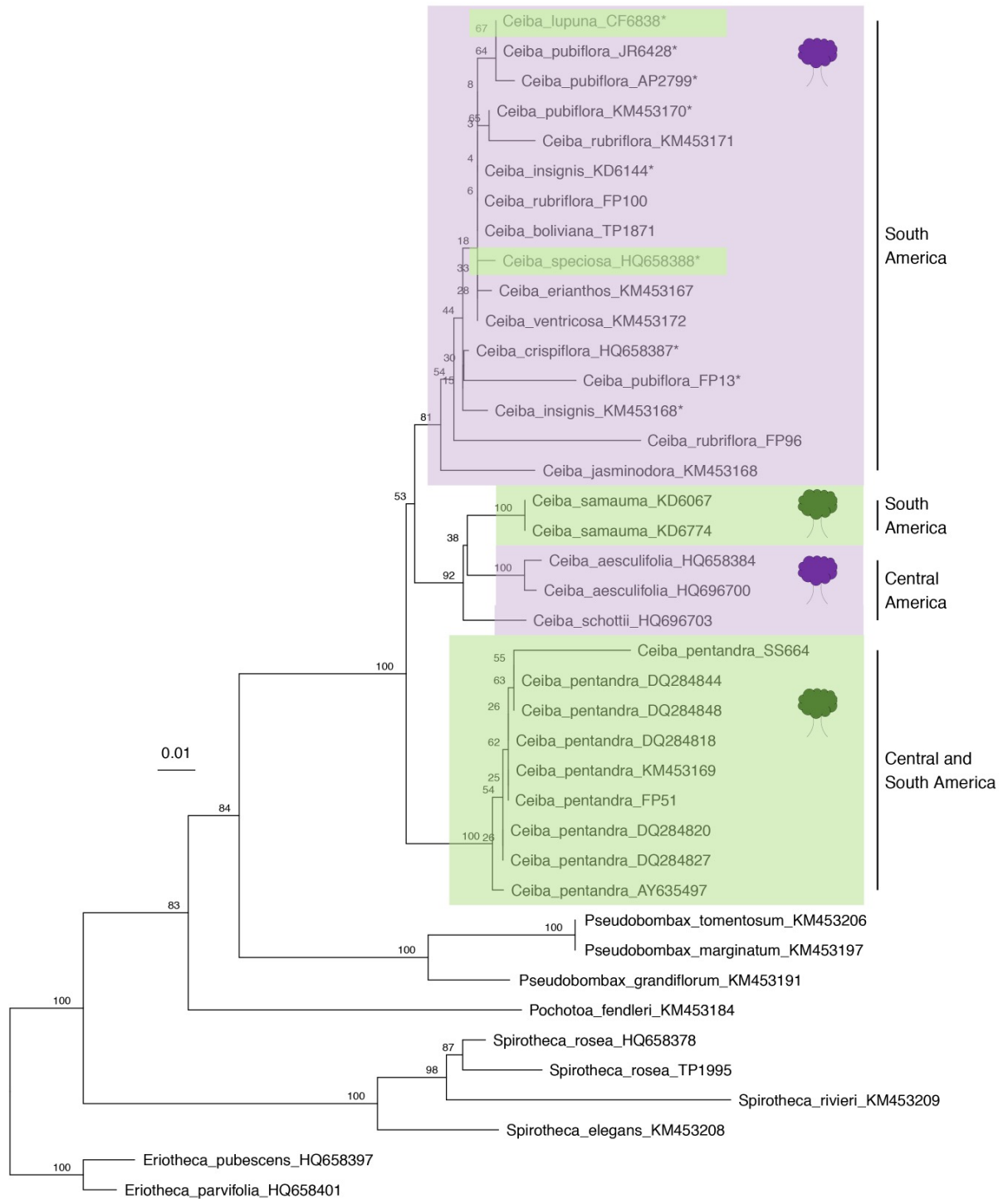
305  
 306 *Ceiba* comprises three main clades: (i) a rain forest lineage of the three accessions of *C.*  
 307 *pentandra*, which are strongly supported as monophyletic [posterior probability (pp) = 1 and

308 bootstrap value = 100] and sister to the remaining species and with stem node age of 12,7 (8,2-  
309 17,6 [95% HPD]) Ma and crown node age of 3,9 (1.6–6,5 [95% HPD]) Ma.; (ii) a highly  
310 supported clade [posterior probability (pp) = 1 and bootstrap value = 92] composed of *C. schottii*  
311 and *C. aesculifolia* from Central American and Mexican SDTF plus two accessions of *C.*  
312 *samauma* from inter-Andean valleys in Peru, with stem node age of 11,2 (7,2-15,6 [95% HPD])  
313 Ma and crown age of 7,3 (3,7-11,0 [95% HPD]) Ma; and (iii) a highly supported [posterior  
314 probability (pp) = 0,99 and bootstrap value = 81] South American SDTF clade including 10  
315 species showing little sequence variation, with stem node age of 11,2 (7,2–15,6 [95% HPD]) Ma  
316 and crown node of 8,6 (5,0–12,5 [95% HPD]) Ma. Within this South American clade, *C.*  
317 *rubriflora*, *C. pubiflora* and *C. insignis*, which were represented by multiple accessions, were  
318 resolved as monophyletic. The South American clade represents SDTF species, except for *C.*  
319 *lupuna*, a species with a distribution restricted to rain forest (Figs. 3 and 4). *Ceiba aesculifolia*  
320 was the only SDTF species recovered as monophyletic with stem node age of 6.3 (3.1–9.8 [95%  
321 HPD]) Ma and crown node of 1.5 (0.2–3.2 [95% HPD]) Ma.

322 The *D* test shows significant phylogenetic signal for both ecological preference ( $D =$   
323 0.1247542,  $P(D=1) = 0.001$ ,  $P(D=0) = 0.3422$ ) and geographical occurrence ( $D = 0.2204991$ ,  $P$   
324 ( $D=1) = 0.0176$ ,  $P(D=0) = 0.3152$ ). Both *D* values are statistically indistinguishable from zero,  
325 which indicates that closely related species are more likely to show the same ecological  
326 preference or geographical occurrence, as expected under a Brownian model of evolution,  
327 whereby there would be a constant rate of state switching over time and any given lineage is  
328 more likely to stay within the same biome and geographic region per unit time than to switch to  
329 the alternative biome or geographic region.



330           The phylogeny supports the monophyly of the two sections of the genus, *Ceiba* and  
331 *Campylanthera*, which are based on pollen and staminal appendages characters. However, it  
332 does not support monophyly of the ‘insignis’ species complex (Gibbs and Semir 2003). This  
333 species aggregate includes seven species (*C. pubiflora*, *C. chodatii*, *C. insignis*, *C. ventricosa*, *C.*  
334 *lupuna*, *C. speciosa* and *C. crispiflora*, indicated with an asterisk in Figs. 3 and 4) characterized  
335 by their entire staminal tube terminating in a collar of anthers, with the exception of *C. pubiflora*  
336 which has free stamens.

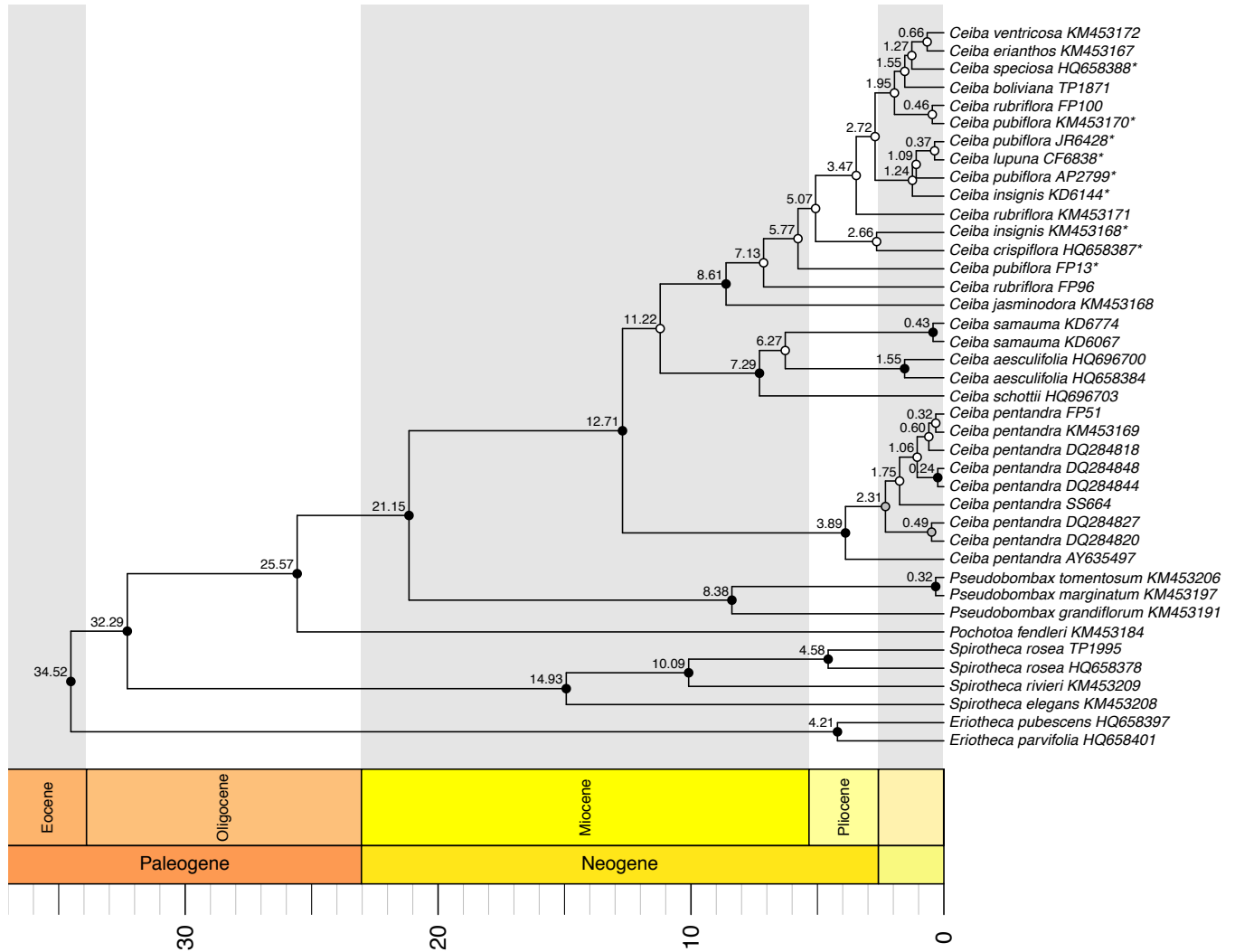


337

338 Figure 3. Maximum likelihood phylogram derived from analysis of nuclear ribosomal ITS sequence data

339 sets for 14 species of *Ceiba*. Species with asterisk belong to the *Ceiba insignis* species aggregate. Values

340 above branches represent bootstrap values for internal nodes. Tree symbols in front of accessions  
 341 represent species occurring in SDTF (purple) and rain forests (green).  
 342



343  
 344 Figure 4. Maximum clade credibility tree resulting from BEAST2 analysis of nuclear ribosomal ITS  
 345 sequence data sets for 14 species of *Ceiba*, 33 Ma offset calibration, using Yule tree prior and the  
 346 Uncorrelated Lognormal Distribution clock model. Values above branches represent nodes ages reported  
 347 in million years as mean ages. Circles represent posterior probabilities for internal nodes: black  $\geq 0.95$ ;  
 348 grey  $< 0.95$  and  $\geq 0.75$ , and white  $< 0.75$ .  
 349

350 **DISCUSSION**

351

352 GEOGRAPHIC AND ECOLOGICAL STRUCTURE - The D test shows clear phylogenetic signal for  
353 ecological preference and geographic phylogenetic structure (i.e., clear Central and South  
354 American clades) in *Ceiba*. However, our data suggest multiple shifts from dry to rain forests  
355 within *Ceiba* (Figs. 3 and 4) because rain forest species are nested within the two dry forest  
356 clades. For example, the two accessions representing *C. samauma*, occurring in rain and riverine  
357 forest in South America, are nested within the Central American and Mexican clade and the rain  
358 forest species *C. lupuna* and *C. speciosa*, are nested within the South American SDTF clade.

359

360 BIOME-SPECIFIC DIFFERENCES IN THE NATURE OF SPECIES AND THEIR DIVERSIFICATION

361 TRAJECTORIES - Our results show young crown and stem ages for species in the SDTF clade.  
362 Rain forest species such as *Ceiba pentandra* and *C. samauma* show patterns of long stems with  
363 shallow crown groups (Table 1). These patterns contrast to previous studies of individual SDTF  
364 species that showed them to be older, with stem ages of 5-10my (e.g., Pennington et al. 2010, de  
365 Queiroz & Lavin 2011), and runs contrary to the prediction of Pennington and Lavin (2016) that  
366 rain forest species might, on average, tend to have more recent origins.

367         The stem age of *C. pentandra* is estimated at 12.7 Ma (Table 1). The long stem and  
368 shallow crown suggest this is an old rain forest lineage with low levels of extant genetic diversity  
369 populations. Likewise, *C. samauma* was recovered as monophyletic, and has a crown node age  
370 estimated as 0.4 Ma and a stem node age of 6,3 Ma. Both species therefore contrast with the  
371 suggested predominant patterns for rain forest species. Our result, recovering *C. pentandra* as  
372 monophyletic, with low sequence divergence amongst accessions, is consistent with that of Dick

373 et al. (2007) who showed *C. pentandra* to have extremely weak phylogeographical structure  
374 based on ITS and chloroplast *psbB-psbF* for 51 individuals. In addition to that, the disjunct  
375 distribution of this species in Africa was demonstrated to be due to relatively recent long  
376 distance dispersal because of low genetic divergence of the African populations.

377         Within the two predominantly SDTF clades, there is little evidence for old lineages  
378 representing morphologically recognized species with long stems and monophyletic crown  
379 groups, as predicted by Pennington and Lavin (2016). The crown age of the South American  
380 SDTF clade, containing 10 species, is estimated at 8.6 Ma, and the Mexican SDTF clade,  
381 containing 2 species, is estimated at 7.9 Ma with a stem age for both estimated at 11.2 Ma. Only  
382 one species from SDTF were recovered as monophyletic, *Ceiba aesculifolia* with a crown age  
383 estimated at 1.5 Ma and stem age at 6.3 Ma. Even when assigning a minimum age of 56 Ma to  
384 the *Eriotheca* stem, the same pattern is observed (Fig. S2). The crown age of the South American  
385 SDTF clade is estimated at 14.9 Ma and the Mexican SDTF clade, containing two species, is  
386 estimated at 12.6 Ma with a stem age for both estimated at 19.5 Ma. The crown age of *Ceiba*  
387 *aesculifolia* was estimated at 2.7 Ma and stem age at 10.79 Ma (Fig. S2).

388         The lack of resolution among the dry forest accessions, with most species being recovered  
389 as non-monophyletic, indicates absence of intraspecific coalescence for the ITS locus.  
390 Explanations for this include incomplete lineage sorting after speciation events, paralogous gene  
391 copies, inaccurate species delimitation and/or hybridization followed by introgression (Naciri  
392 and Linder 2015, Pennington and Lavin 2016). We eliminated sequences with possible  
393 paralogues by visual inspection of the electropherograms and by comparing substitution rates  
394 along branch lengths following Bailey et al. (2003). Some species of *Ceiba* are hypothesised to  
395 be interfertile and hybridise (Gibbs and Semir 2003), especially within the *insignis* species

396 aggregate. However, it is also suggested that those species diverge in time of anthesis and  
397 pollinator type as well, and we have seen no evidence of putative hybrids in the field (Pezzini,  
398 pers. obs.). Eight out of the ten species within the South American SDTF clade are from Brazil  
399 and of these, four are distributed in the *Caatinga*, the largest area of SDTF in the Neotropics  
400 (700,000km<sup>2</sup>) (Silva de Miranda et al. 2018). *Ceiba* species such as *C. pubiflora* are often  
401 widespread (Fig. 2) and abundant (Lima et al. 2010). Taken together, this evidence suggests that  
402 the non-monophyly of *Ceiba* species found in SDTF such as *C. pubiflora* may be a reflection of  
403 large effective population sizes and hence a longer time to coalescence (Naciri and Linder 2015,  
404 Pennington and Lavin 2016), rather than due to hybridisation or ITS paralogy.

405 Our study illustrates that the general patterns of species age, monophyly and geographical  
406 structure reported for species belonging to the Leguminosae family and endemic to SDTF  
407 (Pennington and Lavin 2016) are not shared by one of the most characteristic SDTF tree genera  
408 and suggests that further phylogenetic studies of unrelated groups are required.

409  
410 TAXONOMIC IMPLICATIONS.- Our data support: (i) the circumscription of *Chorisia* within *Ceiba*,  
411 as proposed by Gibbs et al. (1988), Ravenna (1998) and Gibbs and Semir (2003) and confirmed  
412 by recent molecular studies (Carvalho-Sobrinho et al. 2016); (ii) the non-monophyly of the *C.*  
413 *insignis* aggregate species proposed by Gibbs & Semir (2003). Our data suggest that *C.*  
414 *boliviana*, *C. erianthos* and *C. rubriflora*, not included by Gibbs and Semir (2003) are also part  
415 of the *insignis* clade (Fig. 3). It is suggested that those species are interfertile but also diverge in  
416 time of anthesis and pollinator type (Gibbs and Semir 2003). Five of the seven species within  
417 this complex are restricted to the SDTF patches of South America, while *C. speciosa* is  
418 widespread and *C. lupuna* occurs in riverine rain forests in the Peruvian and Brazilian Amazon

419 (Fig. 2); and (iii) the monophyly of the section *Campylanthera* (Gibbs and Semir 2003) that  
420 includes the Central American species *C. aesculifolia*, *C. schottii* and the widespread *C.*  
421 *samauma*.

422

## 423 **ACKNOWLEDGMENTS**

424

425 This paper is part of the PhD thesis of F.P. F.P. acknowledges the support of the Brazilian  
426 National Research Council (CNPq) for funding her PhD at the University of Edinburgh and the  
427 Royal Botanic Garden Edinburgh under the Science without Borders programme (process  
428 206954/2014-0 - GDE). Fieldwork was partially supported by Fundação de Amparo à Pesquisa  
429 do Estado da Bahia (FAPESB - process APP0006/2011). F.P. thanks F.C. Diniz for fieldwork  
430 collaboration, A. Griffiths for help with the phylogenetic signal analysis and E. Koenen for  
431 insightful discussions. The authors are thankful to the reviewers who helped improve the  
432 manuscript.

433

434 Author contributions: F.F.P. and R.T.P conceived the original idea; F.F.P. executed fieldwork,  
435 data collection, laboratory work and data analysis with input from K.G.D; F.F.P. led the writing  
436 of the manuscript with input, comments and review from K.G.D., J.G.C.-S., C.A.K., J.A.N.  
437 L.P.Q. and R.T.P.

438

## 439 **DATA AVAILABILITY STATEMENT**

440

441 **Data Availability:** The data used in this study are archived at Genbank (accession numbers XXX

442 – numbers will be inserted on final accepted version). Scripts for analysis conducted here and for  
443 making figures are available in F.F.P.'s GitHub page (<https://github.com/fpezzini>).

444

445 Supplementary Materials

446 The following materials are available as part of the online article: Supplementary Table 1.

447 Collection details of each accession and ecological preference for the species of *Ceiba*; Figure S1.

448 Maximum clade credibility tree resulting from BEAST2 analysis of nuclear ribosomal ITS

449 sequence data sets for 14 species of *Ceiba* for 47 Ma offset calibration, using Yule tree prior and

450 the Uncorrelated Lognormal Distribution clock model; Figure S2. Maximum clade credibility tree

451 resulting from BEAST2 analysis of nuclear ribosomal ITS sequence data sets for 14 species of

452 *Ceiba* for 56 Ma offset calibration, using Yule tree prior and the Uncorrelated Lognormal

453 Distribution clock model.

454

## 455 LITERATURE CITED

456

457 Álvarez, I. (2003) Ribosomal ITS sequences and plant phylogenetic inference. *Molecular*  
458 *Phylogenetics and Evolution* 29, 417–434.

459 Alverson, W.S. & Duarte, M.C. (2015) Hello Again Pochota , Farewell Bombacopsis  
460 (Malvaceae). *Novon: A Journal for Botanical Nomenclature* 24, 115–119.

461 Bailey, C. (2003) Characterization of angiosperm nrDNA polymorphism, paralogy, and  
462 pseudogenes. *Molecular Phylogenetics and Evolution* 29, 435–455.

463 Baum, D.A., Small, R.L. & Wendel, J.F. (1998) Biogeography and Floral Evolution of Baobabs  
464 *Adansonia*, Bombacaceae as Inferred From Multiple Data Sets D. Cannatella (ed.).



- 465            Systematic Biology 47, 181–207.
- 466    Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., et al. (2019) BEAST 2.5: An advanced  
467            software platform for Bayesian evolutionary analysis M. Pertea (ed.). PLOS Computational  
468            Biology 15, e1006650.
- 469    Buckler, E.S., Ippolito, A. & Holtsford, T.P. (1997) The evolution of ribosomal DNA: divergent  
470            paralogues and phylogenetic implications. Genetics 145, 821–32.
- 471    Carvalho-Sobrinho, J.G., Alverson, W.S., Alcantara, S., Queiroz, L.P., Mota, A.C. & Baum,  
472            D.A. (2016) Revisiting the phylogeny of Bombacoideae (Malvaceae): Novel relationships,  
473            morphologically cohesive clades, and a new tribal classification based on multilocus  
474            phylogenetic analyses. Molecular Phylogenetics and Evolution 101, 56–74.
- 475    Chernomor, O., von Haeseler, A. & Minh, B.Q. (2016) Terrace Aware Data Structure for  
476            Phylogenomic Inference from Supermatrices. Systematic Biology 65, 997–1008.
- 477    Condamine, F.L., Nagalingum, N.S., Marshall, C.R. & Morlon, H. (2015) Origin and  
478            diversification of living cycads: A cautionary tale on the impact of the branching process  
479            prior in Bayesian molecular dating. BMC Evolutionary Biology 15, 1–18.
- 480    Crisp, M.D., Arroyo, M.T.K., Cook, L.G., Gandolfo, M.A., Jordan, G.J., McGlone, M.S.,  
481            Weston, P.H., Westoby, M., Wilf, P. & Linder, H.P. (2009) Phylogenetic biome  
482            conservatism on a global scale. Nature 458, 754–756.
- 483    Dexter, K.G., Lavin, M., Torke, B.M., Twyford, A.D., Kursar, T.A., Coley, P.D., Drake, C.,  
484            Hollands, R. & Pennington, R.T. (2017) Dispersal assembly of rain forest tree communities  
485            across the Amazon basin. Proceedings of the National Academy of Sciences 114, 2645–  
486            2650.
- 487    Dick, C.W., Bermingham, E., Lemes, M.R. & Gribel, R. (2007) Extreme long-distance dispersal

- 488 of the lowland tropical rainforest tree *Ceiba pentandra* L. (Malvaceae) in Africa and the  
489 Neotropics. *Molecular ecology* 16, 3039–49.
- 490 Drummond, A.J., Ho, S.Y.W., Phillips, M.J. & Rambaut, A. (2006) Relaxed Phylogenetics and  
491 Dating with Confidence D. Penny (ed.). *PLoS Biology* 4, e88.
- 492 DRYFLOR (2016) Plant diversity patterns in neotropical dry forests and their conservation  
493 implications. *Science* 353, 1383–1387.
- 494 Duarte, L. (1974) Sobre uma flor de Bombacaceae da Bacia Terciária de Fonseca, MG. *Anais da*  
495 *Academia Brasileira de Ciências* 46, 407–411.
- 496 Duarte, M.C., Esteves, G.L., Salatino, M.L.F., Walsh, K.C. & Baum, D. a. (2011) Phylogenetic  
497 Analyses of *Eriotheca* and Related Genera (Bombacoideae, Malvaceae). *Systematic Botany*  
498 36, 690–701.
- 499 Feliner, G.N. & Rosselló, J.A. (2007) Better the devil you know? Guidelines for insightful  
500 utilization of nrDNA ITS in species-level evolutionary studies in plants. *Molecular*  
501 *Phylogenetics and Evolution* 44, 911–919.
- 502 Fritz, S.A. & Purvis, A. (2010) Selectivity in Mammalian Extinction Risk and Threat Types: a  
503 New Measure of Phylogenetic Signal Strength in Binary Traits. *Conservation Biology* 24,  
504 1042–1051.
- 505 Gibbs, P. & Semir, J. (2003) A taxonomic revision of the genus *Ceiba* Mill. (Bombacaceae).  
506 *Anales del Jardín Botánico de Madrid* 60, 259–300.
- 507 Gibbs, P.E., Semir, J. & Da Cruz, N.D. (1988) A proposal to unite the genera *Chorisia* Knuth  
508 with *Ceiba* Miller (Bombacaceae). *Notes from the Royal Botanic Garden Edinburgh* 45,  
509 125–136.
- 510 Gradstein, F.M., Ogg, J.G., Schmitz, M.D. & Ogg, G.M. (2012) *The Geologic Time Scale*,

- 511 Elsevier.
- 512 Graham, C.H. & Fine, P.V.A. (2008) Phylogenetic beta diversity: linking ecological and  
513 evolutionary processes across space in time. *Ecology Letters* 11, 1265–1277.
- 514 Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010) New  
515 Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the  
516 Performance of PhyML 3.0. *Systematic Biology* 59, 307–321.
- 517 Hughes, C.E., Pennington, R.T. & Antonelli, A. (2013) Neotropical Plant Evolution: Assembling  
518 the Big Picture. *Botanical Journal of the Linnean Society* 171, 1–18.
- 519 Kass, R.E. & Raftery, A.E. (1995) Bayes Factors. *Journal of the American Statistical Association*  
520 90, 773–795.
- 521 Koenen, E.J.M., Clarkson, J.J., Pennington, T.D. & Chatrou, L.W. (2015) Recently evolved  
522 diversity and convergent radiations of rainforest mahoganies (Meliaceae) shed new light on  
523 the origins of rainforest hyperdiversity. *New Phytologist* 207, 327–339.
- 524 Lanfear, R., Calcott, B., Ho, S.Y.W. & Guindon, S. (2012) PartitionFinder: Combined Selection  
525 of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. *Molecular*  
526 *Biology and Evolution* 29, 1695–1701.
- 527 Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T. & Calcott, B. (2017) Partitionfinder 2:  
528 New methods for selecting partitioned models of evolution for molecular and  
529 morphological phylogenetic analyses. *Molecular Biology and Evolution* 34, 772–773.
- 530 Lavin, M. (2006) Floristic and Geographical Stability of Discontinuous Seasonally Dry Tropical  
531 Forests Explains Patterns of Plant Phylogeny and Endemism. In: R.T. Pennington, G.P.  
532 Lewis, & J.A. Ratter (eds.) *Neotropical Savannas and Seasonally Dry Forests: Plant*  
533 *Diversity, Biogeography, and Conservation*. Boca Raton: CRC Press, pp. 433–447.

- 534 de Lima, M.R. & Salard-Cheboldaeff, M. (1981) Palynologie des bassins de Gandarela et  
535 Fonseca (Eocene de l'état de Minas Gerais, Bresil). *Boletim IG* 12, 33–54.
- 536 Lima, M.S. de, Damasceno-Júnior, G.A. & Tanaka, M.O. (2010) Aspectos estruturais da  
537 comunidade arbórea em remanescentes de floresta estacional decidual, em Corumbá, MS,  
538 Brasil. *Revista Brasileira de Botânica* 33, 437–453.
- 539 Louca, S. & Pennell, M.W. (2020) Extant timetrees are consistent with a myriad of  
540 diversification histories. *Nature* 11.
- 541 Maddison, W.P. & Maddison, D.R. (2015) Mesquite: A Modular System for Evolutionary  
542 Analysis.
- 543 Miles, L., Newton, A.C., DeFries, R.S., Ravilious, C., May, I., Blyth, S., Kapos, V. & Gordon,  
544 J.E. (2006) A global overview of the conservation status of tropical dry forests. *Journal of*  
545 *Biogeography* 33, 491–505.
- 546 Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A.  
547 & Lanfear, R. (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic  
548 Inference in the Genomic Era E. Teeling (ed.). *Molecular Biology and Evolution* 37, 1530–  
549 1534.
- 550 Murphy, P.G. & Lugo, A.E. (1986) Ecology of Tropical Dry Forest. *Annual Review of Ecology*  
551 *and Systematics* 17, 67–88.
- 552 Naciri, Y. & Linder, H.P. (2015) Species delimitation and relationships: The dance of the seven  
553 veils. *Taxon* 64, 3–16.
- 554 Nguyen, L.-T., Schmidt, H.A., von Haeseler, A. & Minh, B.Q. (2015) IQ-TREE: A Fast and  
555 Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies.  
556 *Molecular Biology and Evolution* 32, 268–274.

- 557 Orme, C.D.L. (2013) The caper package: comparative analyses in phylogenetics and evolution in  
558 R. [Http://Caper.R-Forge.R-Project.Org/](http://Caper.R-Forge.R-Project.Org/), 1–36.
- 559 Pennington, R.T., Daza, A., Reynel, C. & Lavin, M. (2011) *Poissonia eriantha* (Leguminosae)  
560 From Cuzco, Peru: An Overlooked Species Underscores a Pattern of Narrow Endemism  
561 Common to Seasonally Dry Neotropical Vegetation. *Systematic Botany* 36, 59–68.
- 562 Pennington, R.T. & Lavin, M. (2016) The contrasting nature of woody plant species in different  
563 neotropical forest biomes reflects differences in ecological stability. *New Phytologist* 210,  
564 25–37.
- 565 Pennington, R.T., Lavin, M. & Oliveira-Filho, A. (2009) Woody Plant Diversity, Evolution, and  
566 Ecology in the Tropics: Perspectives from Seasonally Dry Tropical Forests. *Annual Review*  
567 *of Ecology, Evolution, and Systematics* 40, 437–457.
- 568 Pennington, R.T., Lavin, M., Sarkinen, T., Lewis, G.P., Klitgaard, B.B. & Hughes, C.E. (2010)  
569 Contrasting plant diversification histories within the Andean biodiversity hotspot.  
570 *Proceedings of the National Academy of Sciences* 107, 13783–13787.
- 571 Pennington, R.T., Richardson, J.E. & Lavin, M. (2006) Insights into the historical construction of  
572 species-rich biomes from dated plant phylogenies, neutral ecological theory and  
573 phylogenetic community structure. *New Phytologist* 172, 605–616.
- 574 de Queiroz, L.P. & Lavin, M. (2011) *Coursetia* (Leguminosae) From Eastern Brazil: Nuclear  
575 Ribosomal and Chloroplast DNA Sequence Analysis reveal the Monophyly of Three  
576 Caatinga-inhabiting Species. *Systematic Botany* 36, 69–79.
- 577 Rambaut, A., Drummond, A.J., Xie, D., Baele, G. & Suchard, M.A. (2018) Posterior  
578 Summarization in Bayesian Phylogenetics Using Tracer 1.7 E. Susko (ed.). *Systematic*  
579 *Biology* 67, 901–904.

- 580 Rangel, T.F., Edwards, N.R., Holden, P.B., Diniz-Filho, J.A.F., Gosling, W.D., Coelho, M.T.P.,  
581 Cassemiro, F.A.S., Rahbek, C. & Colwell, R.K. (2018) Modeling the ecology and evolution  
582 of biodiversity: Biogeographical cradles, museums, and graves. *Science* 361, eaar5452.
- 583 Ravenna, P. (1998) On the identity, validity, and actual placement in *Ceiba* of several *Chorisia*  
584 species (Bombacaceae), and description of two new south american species. *Onira* 3, 42–51.
- 585 Renner, S.S. (2005) Relaxed molecular clocks for dating historical plant dispersal events. *Trends*  
586 in Plant Science 10, 550–558.
- 587 Richardson, J.E., Pennington, R.T., Pennington, T.D. & Hollingsworth, P.M. (2001) Rapid  
588 Diversification of a Species-Rich Genus of Neotropical Rain Forest Trees. *Science* 293,  
589 2242–2245.
- 590 Robyns, A. (1963) Essai de monographie du genre *Bombax* s.l. (Bombacaceae) (Suite). *Bulletin*  
591 du Jardin botanique de l'État a Bruxelles 33, 145.
- 592 Russel, P.M., Brewer, B.J., Klaere, S. & Bouckaert, R.R. (2019) Model Selection and Parameter  
593 Inference in Phylogenetics Using Nested Sampling T. Stadler (ed.). *Systematic Biology* 68,  
594 219–233.
- 595 Sanderson, M.J. (2002) Estimating Absolute Rates of Molecular Evolution and Divergence  
596 Times: A Penalized Likelihood Approach. *Molecular Biology and Evolution* 19, 101–109.
- 597 Sanderson, M.J. (2004) r8s, version 1.70, User's Manual. , 1–42.
- 598 Silva de Miranda, P.L., Oliveira-Filho, A.T., Pennington, R.T., Neves, D.M., Baker, T.R. &  
599 Dexter, K.G. (2018) Using tree species inventories to map biomes and assess their climatic  
600 overlaps in lowland tropical South America. *Global Ecology and Biogeography* 27, 899–  
601 912.
- 602 Simon, M.F., Grether, R., de Queiroz, L.P., Skema, C., Pennington, R.T. & Hughes, C.E. (2009)

603           Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of  
604           adaptations to fire. *Proceedings of the National Academy of Sciences* 106, 20359–20364.

605   Skilling, J. (2006) Nested sampling for general Bayesian computation. *Bayesian Analysis* 1,  
606           833–859.

607   Stebbins, G.L. (1974) *Flowering plants: evolution above the species level*. T.B.P. of H.U. Press  
608           (ed.), Cambridge, MA.

609   Wiens, J.J. & Donoghue, M.J. (2004) *Historical biogeography, ecology and species richness*.  
610           *Trends in Ecology & Evolution* 19, 639–644.

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627 **Supplementary Material**

628 Supplementary Table 1. Collection details of each accession and ecological preference for the

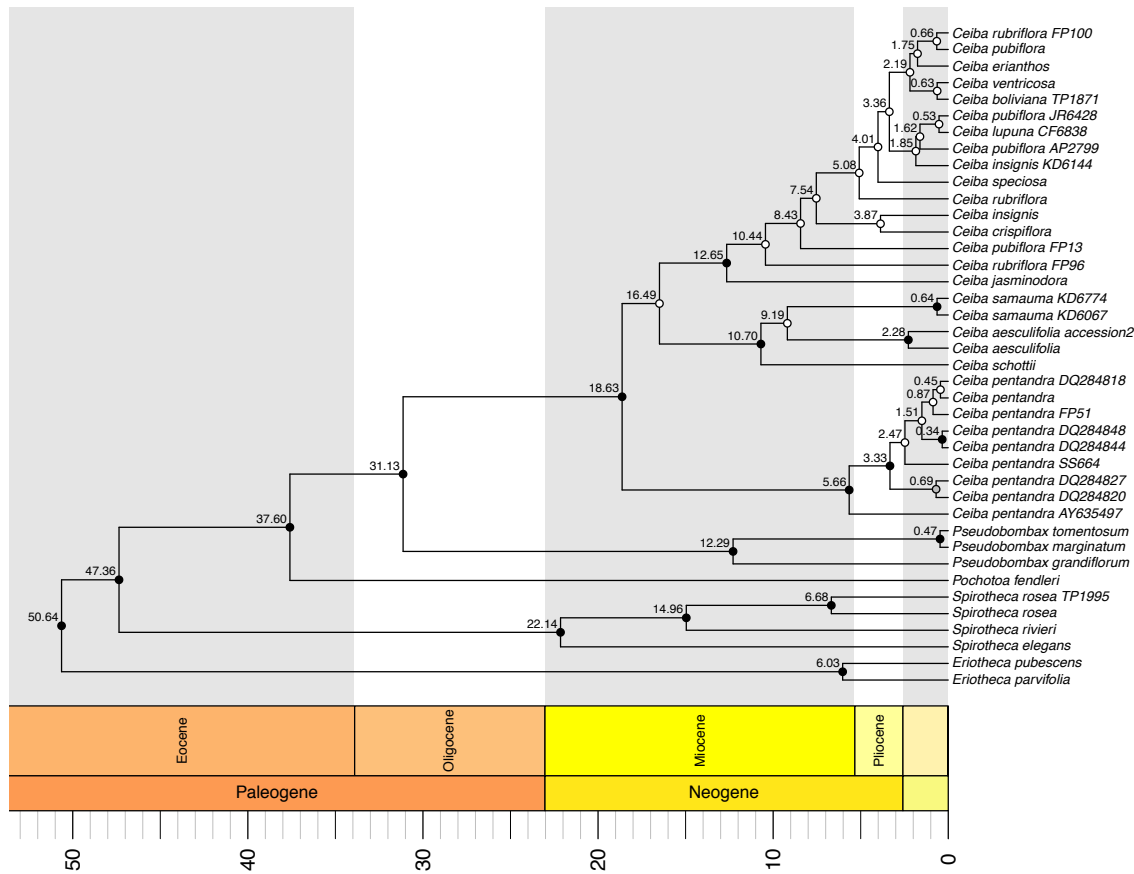
629 species of *Ceiba*.

n	Accession	Species	Locality	Country	GenBank accession	Collector name + number (Herbarium)	Ecological preference
1	HQ658384	<i>Ceiba aesculifolia</i>	-	-	HQ658384	Fairchild Botanical Gardens acc. no. 83301	SDTF
2	HQ696700	<i>Ceiba aesculifolia</i> accession2	-	-	HQ696700	Fairchild Botanical Gardens acc. no. X-2-206	SDTF
3	TP1871	<i>Ceiba boliviana</i>	Huancavelica	Peru	in process of submission	Pennington, R. Toby (E)	SDTF
4	HQ658387	<i>Ceiba crispiflora</i>	-	-	HQ658387	Pacific Tropical Garden acc. no. 750726001	rain forests
5	KM453167	<i>Ceiba erianthos</i>	Bahia	Brazil	KM453167	E.R.Souza 710 (HUEFS)	SDTF
6	KD6144	<i>Ceiba insignis</i>	Pucacaca	Peru	in process of submission	Dexter, K.G. (E)	SDTF
7	KM488629	<i>Ceiba insignis</i>	Cajamarca	Peru	KM488629	J. Campos & P. López 4953 (US)	SDTF
8	KM453168	<i>Ceiba jasminodora</i>	Rio de Janeiro	Brazil	KM453168	Carvalho-Sobrinho 3070 (HUEFS)	SDTF
9	CF6838	<i>Ceiba lupuna</i>	Acre	Brazil	in process of submission	Daly, D.C. 6838 (E)	rain forests
10	FP51	<i>Ceiba pentandra</i>	Piauí	Brazil	in process of submission	Pezzini, F.F. 51 (HUEFS)	rain forests
11	SS664	<i>Ceiba pentandra</i>	Minas	Brazil	in process of submission	Sant'ana, S.C. (E)	rain forests
12	KM453169	<i>Ceiba pentandra</i>	-	-	KM453169	Carvalho-Sobrinho s.n. (HUEFS)	rain forests
13	DQ284818	<i>Ceiba pentandra</i>	Camuy	Puerto Rico	DQ284818	-	rain forests
14	DQ284848	<i>Ceiba pentandra</i>	Kourou	French Guiana	DQ284848	-	rain forests
15	DQ284844	<i>Ceiba pentandra</i>	Castanhal Veado (Rio Trombetas)	Brazil	DQ284844	-	rain forests
16	DQ284827	<i>Ceiba pentandra</i>	Borbon	Ecuador	DQ284827	-	rain forests
17	DQ284820	<i>Ceiba pentandra</i>	Morelos	Mexico	DQ284820	-	rain forests
18	AY635497	<i>Ceiba pentandra</i>	Jobero	Panama	AY635497	-	rain forests
19	FP13	<i>Ceiba pubiflora</i>	Bahia	Brazil	in process of submission	Pezzini, F.F. 13 (HUEFS)	SDTF
20	JR6428	<i>Ceiba pubiflora</i>	Minas	Brazil	in process of submission	Ratter, J. (E)	SDTF
21	AP2799	<i>Ceiba pubiflora</i>	Minas	Brazil	in process of submission	A. Pott (E)	SDTF
22	KM453170	<i>Ceiba pubiflora</i>	Bahia	Brazil	KM453170	Carvalho-Sobrinho 3066 (HUEFS)	SDTF
23	FP96	<i>Ceiba rubriflora</i>	Minas	Brazil	in process of submission	Pezzini, F.F. 96 (HUEFS)	SDTF
24	FP100	<i>Ceiba rubriflora</i>	Minas	Brazil	in process of submission	Pezzini, F.F. 100 (HUEFS)	SDTF
25	KM453171	<i>Ceiba rubriflora</i>	Bahia	Brazil	KM453171	Carvalho-Sobrinho 574 (HUEFS)	SDTF
26	KD6067	<i>Ceiba samauma</i>	Buenos Aires	Peru	in process of submission	Dexter, K.G. (E)	rain forests
27	KD6774	<i>Ceiba samauma</i>	Echarate	Peru	in process of submission	Dexter, K.G. (E)	rain forests
28	HQ696703	<i>Ceiba schottii</i>	-	-	HQ696703	Fairchild Botanical Gardens acc. no. 83302	SDTF
29	HQ658388	<i>Ceiba speciosa</i>	-	Brazil	HQ658388	W.S. Alverson s.n. (WIS)	rain forests
30	KM453172	<i>Ceiba ventricosa</i>	Bahia	Brazil	KM453172	Carvalho-Sobrinho 3124 (HUEFS)	SDTF
31	HQ658401	<i>Eriotheca parvifolia</i>	-	-	HQ658401	M.C. Duarte 109 (SP)	-
32	HQ658397	<i>Eriotheca pubescens</i>	-	-	HQ658397	M.C. Duarte 115 (SP)	-
33	KM453184	<i>Pochota fenderi</i>	-	-	KM453184	P.E. Kaminski s.n. (HUEFS)	-
34	KM453191	<i>Pseudobombax grandiflorum</i>	-	-	KM453191	Carvalho-Sobrinho 2946 HUEFS	-
35	KM453197	<i>Pseudobombax marginatum</i>	-	-	KM453197	L.P. Queiroz 14753 (HUEFS)	-
36	KM453206	<i>Pseudobombax tomentosum</i>	-	-	KM453206	Carvalho-Sobrinho 2874 HUEFS	-
37	KM453208	<i>Spirotheca elegans</i>	-	-	KM453208	Carvalho-Sobrinho 2964 HUEFS	-
38	KM453209	<i>Spirotheca rivieri</i>	-	-	KM453209	Carvalho-Sobrinho s.n. HUEFS	-
39	TP1995	<i>Spirotheca rosea</i>	Junin	Peru	in process of submission	Pennington, R. Toby (E)	-
40	HQ658378	<i>Spirotheca rosea</i>	-	-	HQ658378	W.S. Alverson 2185 (WIS)	-

630

631



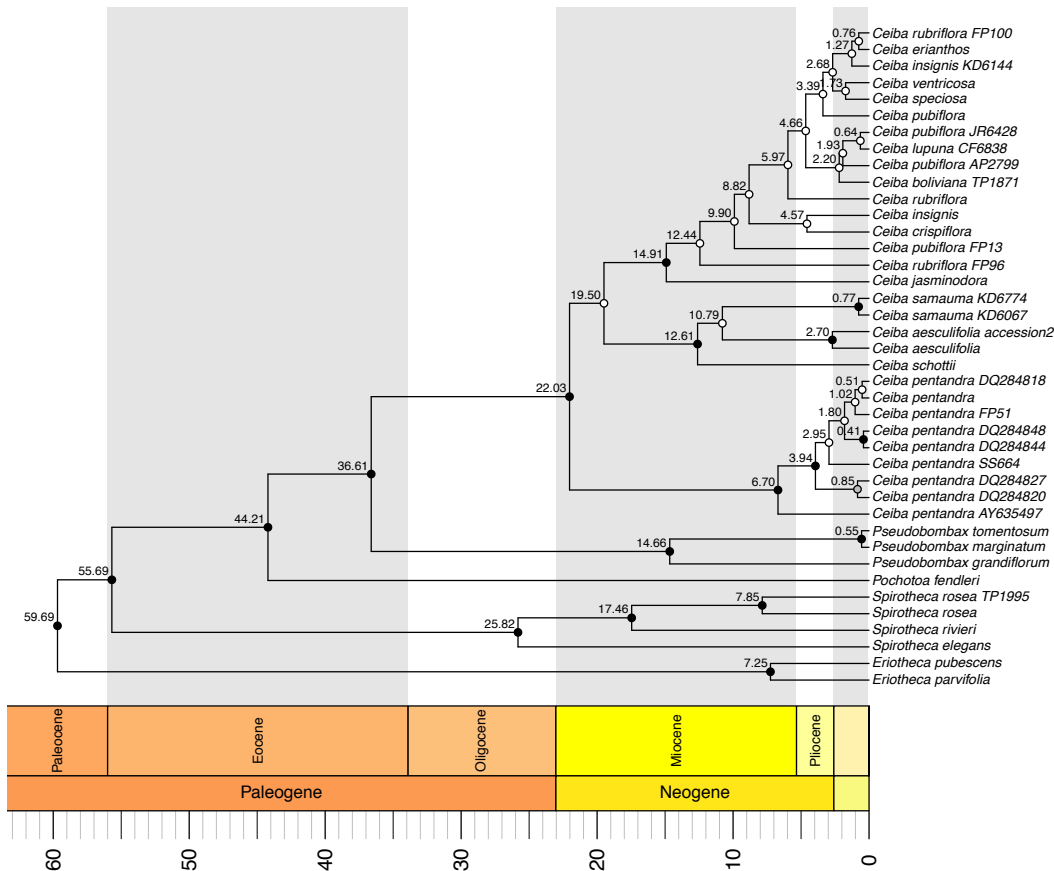


632

633 Supplementary Figure 1. Maximum clade credibility tree resulting from BEAST2 analysis of nuclear  
 634 ribosomal ITS sequence data sets for 14 species of *Ceiba* and 47 Ma offset calibration, using Yule tree  
 635 prior and the Uncorrelated Lognormal Distribution clock model. Values above branches represent nodes  
 636 ages reported in million years as mean ages. Circles represent posterior probabilities for internal nodes:  
 637 black  $\geq 0.95$ ; grey  $< 0.95$  and  $\geq 0.75$ , and white  $< 0.75$ .

638

639



Supplementary Figure 2. Maximum clade credibility tree resulting from BEAST2 analysis of nuclear ribosomal ITS sequence data sets for 14 species of *Ceiba* and 56 Ma offset calibration, using Yule tree prior and the Uncorrelated Lognormal Distribution clock model. Values above branches represent nodes ages reported in million years as mean ages. Circles represent posterior probabilities for internal nodes: black  $\geq 0.95$ ; grey  $< 0.95$  and  $\geq 0.75$ , and white  $< 0.75$ .