The LINNEAN SOCIETY of London BOTANICAL Journal of the Linnean Society

Botanical Journal of the Linnean Society, 2016, **182**, 287–302. With 3 figures

Phylogenetic analysis of *Attalea* (Arecaceae): insights into the historical biogeography of a recently diversified Neotropical plant group

CINTIA FREITAS[1,2]*, ALAN W. MEEROW[3], JEAN-CHRISTOPHE PINTAUD[4†], ANDREW HENDERSON[1], LARRY NOBLICK[5], FLAVIA R. C. COSTA[6], CARLOS E. BARBOSA[6] and DAVID BARRINGTON[2]

[1]*The New York Botanical Garden, 2900 Southern Blvd., Bronx, NY 10458, USA*
[2]*Plant Biology Department, University of Vermont, 63 Carrigan Drive, Burlington, VT 05405, USA*
[3]*USDA-ARS, National Germplasm Repository, 13601 Old Cutler Road, Miami, FL 33158, USA*
[4]*UMR 232 DIADE/DYNADIV, Institut de Recherche pour le Développement (IRD), BP 64501, 911 avenue Agropolis, F-34394 Montpellier, France*
[5]*Montgomery Botanical Center, 11901 Old Cutler Rd, Coral Gables, FL 33156, USA*
[6]*Instituto de Pesquisas da Amazônia, Av. Ephigênio Sales 2239, CEP 69011-970, Manaus, Amazonas, Brazil*

Received 21 September 2015; revised 20 May 2016; accepted for publication 28 June 2016

We present a dated phylogenetic tree for the Neotropical genus *Attalea* (Arecaceae). We used six orthologues from the nuclear WRKY gene family across 98 accessions to address relationships among species and biogeographical hypotheses. We found that the formerly recognized groups within *Attalea* are not monophyletic and therefore there is no support for multiple genera as previously thought. Species of *Attalea*-like palms from the Atlantic forest form a well-supported clade sister to the *Attalea* species from Amazonia, the Andean valleys and Mesoamerica. Dates for the main divergence events suggest a relationship with the development of the dry forests that now span eastern South America and the now-lost Pebas Lake region in the western Amazon. *Attalea crassispatha* possibly colonized Hispaniola by a long-distance dispersal event, not via the land bridge Great Antilles Avian Ridge (GAAR), before the Panama channel closed. The common ancestor of *Attalea* appears to have been an Atlantic forest clade *c.* 30 Mya. The early split between the Atlantic-forest clade and the two Amazonian-northern Andean clades was probably the result of climatic changes that caused an increase in aridity in South America. © 2016 The Linnean Society of London, *Botanical Journal of the Linnean Society*, 2016, **182**, 287–302

KEYWORDS: Bayesian inference – molecular phylogeny – neotropics – palm phylogeny – South America.

INTRODUCTION

The incorporation of molecular phylogenetic information is now standard for finding relationships between species (Roncal *et al.*, 2010) and testing biogeographical hypotheses (Crisp, Trewick & Cook, 2011) to establish and ultimately to test hypotheses related to patterns of richness in tropical regions (e.g. Bjorholm, Svenning & Baker, 2006). However, in tropical regions, where biodiversity is highest, the lack of collections from remote areas (Hopkins, 2007) and the incomplete nature of the collections that have been made impede the correct delineation of phylogenetic relationships, the development of accurate biogeographical hypotheses and our understanding of the evolutionary history of species (Roncal *et al.*, 2010; Crisp *et al.*, 2011). Arecaceae are an ideal group for testing biogeographical hypotheses in tropical areas because palms are diverse and abundant, have a long evolutionary history in humid forests (Roncal *et al.*, 2010; Baker & Couvreur, 2013a), are known to influence the distribution of other species (Terborgh, 1992) and are phylogenetically well resolved at the subfamilial level (Asmussen *et al.*, 2006; Couvreur, Forest & Baker, 2011).

---

*Corresponding author. E-mail: cintia.freitas@ufg.edu.br
†Deceased 10 August 2015.

Attalea, a prominent genus of palms throughout tropical America, belongs to subfamily Arecoideae, tribe Cocoseae; it is one of ten [previously 11, since *Lytocaryum* was included in *Syagrus* (Noblick & Meerow, 2015)] monophyletic genera of subtribe Attaleinae. *Attalea* spp. are found from Mexico as far south as Bolivia and Paraguay and from the lower slopes of the Andes (reaching 1600 m) to the north-eastern coast of Brazil (Henderson, 1995; Dransfield *et al.*, 2008), which is hypothesized to be the centre of distribution of the genus (Meerow *et al.*, 2009). *Attalea* exhibits high ecological variation, suggesting successful colonization of new areas and great adaptive potential. It occurs on dunes, in tropical rainforest and in Cerrado (Dransfield *et al.*, 2008). The Cerrado encompasses the savanna woodlands found across Brazil, i.e. campos rupestres, (rocky savannas) and Pantanal (the seasonally inundated floodplain located in the south-western most part of the cerrado biome) (Oliveira-Filho & Ratter, 2002). In the cerrado, *Attalea* is part of a characteristic vegetation associated with sandstone and quartzite soils (Ribeiro & Walter, 1998). However, the historical biogeography of *Attalea* remains elusive despite the contribution of two pioneering studies (Meerow *et al.*, 2009, 2014).

Diagnostic morphological features are related to the staminate flowers, which are asymmetrical, with three small sepals and three slightly larger petals (Dransfield *et al.*, 2008). The stamens are shorter or rarely longer than the petals and the anthers can be from straight to twisted and coiled; the flowers may have a minute pistillode or none (Dransfield *et al.*, 2008). Like many other palms, *Attalea* spp. show a variety of habits from acaulescent to erect, but the stems are always solitary (Tomlinson, 2006). According to Glassman (1999), leaf anatomy is of limited use for the classification of *Attalea*, although it seems promising in the closely allied genus *Syagrus*, in which variation was found to be congruent with molecular data (Noblick, 2013) and in the six *Allagoptera* spp. that have clearly distinct leaf anatomies (Pinedo *et al.*, 2016; this issue).

The taxonomy of *Attalea* is controversial, with treatments varying in terms of generic limits and number of species. Characters of the staminate flowers were previously used to recognize four segregate genera, but they are not currently accepted (Dransfield *et al.*, 2008). The two most recent treatments proposed different classifications: Glassman (1999) accepted 66 species in four genera (*Attalea*, *Scheelea* H.Karst., *Orbignya* Mart. ex Endl. and *Maximiliana* Mart.) and Henderson (1995) accepted a single genus *Attalea* with 27 or 31 species (A. Henderson, unpubl. data). The Kew Checklist (WCSP, 2015) accepts 66 *Attalea* spp. and four hybrids. Consensus on *Attalea* species delineation has been reached for only 20 of them (Pintaud, 2008). The source of the taxonomic disagreement may lie in the lack of appropriate botanical material, high rate of hybridization and difficulties in the interpretation of hybrids (Pintaud, 2008). Hybridization has mostly been suggested between widespread species, e.g. *Attalea speciosa* Mart. and *A. vitrivir* Zona (Henderson, 1995; Glassman, 1999). *Attalea oleifera* Barb.Rodr. and *A. funifera* Mart. are also thought to hybridize in the Atlantic forest (Bondar, 1942).

Like most palms, *Attalea* spp. have great economic and cultural importance for local communities. They are used to produce cosmetics [*Attalea insignis* (Mart.) Drude in Engl. & Prantl] and medicines against hepatitis [*A. butyracea* (Mutis ex L.f.) Wess.-Boer and *A. maripa* (Aubl.) Mart.)] (Balslev *et al.*, 2008). The leaves of some *Attalea* spp. are used in religious ceremonies and in the extraction of palm hearts, e.g. *A. microcarpa* Mart., *A. racemosa* Spruce, *A. phalerata* Mart. ex Spreng. and *A. plowmanii* (Glassman) Zona (Balslev *et al.*, 2008). *Attalea butyracea* has great commercial potential in the production of ethanol and sugar (Bernal & Galeano, 2010). Babassu (*A. speciosa*) is used to produce palm oil in Brazil (Anderson, May & Balick, 1991) and is by far the most important native palm used in that country.

Recent phylogenetic studies of the palm family (Baker *et al.*, 2009; Couvreur *et al.*, 2011), subfamily Arecoideae (Asmussen *et al.*, 2006), tribe Cocoseae (Meerow *et al.*, 2014) and subtribe Attaleinae (Meerow *et al.*, 2009) have improved our understanding of the phylogenetic relationships of *Attalea*. Meerow *et al.* (2009, 2014) used WRKY genes (short for the WRKYGQK conserved amino acid sequence that gives name to the WRKY domain of these genes) to resolve phylogenetic relationships in Cocoseae and Attaleinae, but included only 16 *Attalea* spp. *Attalea* was resolved as sister to the remaining Neotropical members of the subtribe in Meerow *et al.* (2009). The analysis of Cocoseae by Meerow *et al.* (2014) resolved *Attalea* as sister to *Cocos*, albeit with some ambiguity. In a study of Arecaceae by Baker *et al.* (2009), *Attalea* was resolved as sister to a clade comprising *Allagoptera*, *Polyandrococos*, *Lytocaryum* and *Syagrus*. Although *Attalea* was retrieved as monophyletic in these studies, resolution of some of the more terminal clades in the genus was poor.

The WRKY gene superfamily has proved to be useful in resolving the phylogenetic history of palms, as demonstrated in tribe Cocoseae (Meerow *et al.*, 2009, 2014). The superfamily, mostly found in plants, has at least 55 members (Cannon *et al.*, 2004), originating during whole-genome duplication events and in the course of divergent evolution, especially in monocots (Wu *et al.*, 2005). The superfamily shows remarkably low levels of paralogy; in *Oryza sativa* subsp. *indica*, 99 of the 102 described WRKY loci are present as single copies; in *O. sativa* subsp. *japonica*, 97 of the 98 loci are present as single copies (Ross, Liu & Shen, 2007). WRKY genes are characterized by the highly conserved DNA binding sequence WRKYGQK followed by a C_2H_2- or C_2HC-type zinc finger motif that codes for transcription factors. Expanded molecular phylogenetic analyses should help to resolve some of the taxonomic problems remaining in *Attalea* and provide the basis for testing biogeographical hypothesis, as has already been shown in *Astrocaryum* (Roncal *et al.*, 2013), *Bactris* (Eiserhardt *et al.*, 2011b) and *Geonoma* (Roncal *et al.*, 2005).

The evolutionary history of *Attalea* seems to be recent (Meerow *et al.*, 2009), and there is evidence of an increase in the diversification rate in an ancestor clade of *Attalea* around 15 Mya (Baker & Couvreur, 2013b) although an overall deceleration of rate change in Arecales compared with most monocots was found in a genome-wide plastid study (Barrett *et al.*, 2015). Baker & Couvreur (2013a, b) also found evidence for a low rate of extinction in palms, which is supported by Meerow *et al.* (2009, 2014) for Cocoseae. Evidence for both a recent increase in diversification and an apparent low extinction rate indicates the influence of recent events that have led *Attalea* to reach its current diversity.

A biogeographical analysis of *Attalea* in a phylogenetic context provides the opportunity to test the role of the changing South American landscape in its diversification and to test alternative hypotheses for these landscape-level changes. The main biogeographical events that might have influenced the diversification of *Attalea* include: (1) the flow reversal of the Amazon River; (2) the formation of open dry forests across South America; (3) the marine incursions on the Amazon Region; (4) the final closure of the Panama channel; and (5) the uplift of the northernmost part of the Andes. Regarding the Amazon river flow reversal, the timing is the most contentious aspect. First, *c.* 15 Mya, a time when the Andes were approximately half their current elevation, the Amazon River was already running to the east (Hoorn *et al.*, 2010b). An opposing view dates these events as much more recent, with the uplift of the northern Andes and the flow reversal occurring between *c.* 5 and 3 Mya (Latrubesse *et al.*, 2010).

Second, we hypothesize that the distribution centre of *Attalea* was in the forest along the eastern coast of Brazil. However, the effects of the isolation of the Atlantic forest from the Amazon caused by the formation of the open dry forest across South America (Chaco, Cerrado and Caatinga) on the evolution of *Attalea* have yet to be determined. In fact, isolation of the Atlantic forest appears to be a causal factor of species endemism (Myers *et al.*, 2000). Third, the influence of the two main marine incursions in the Amazon region, with occurrences dating back to 15–10 and 5 Mya (Jordan *et al.*, 2005), can also be tested. Finally, the uplift of the northernmost Andes and the consequent closing of the Caribbean Sea and lake complex formation in the western Amazon may also have influenced the diversification of this group. These questions are central to our inquiry using a phylogenetic approach combined with a process-based biogeographical hypothesis. *Attalea*, species-rich and widespread in the region, is an excellent model lineage for testing the effects of these large-scale biogeographical events on species evolution.

In this study we sought to add to the rapidly expanding understanding of infrageneric tropical palm species relationships (Roncal *et al.*, 2005; Eiserhardt *et al.*, 2011b), with a phylogenetic and biogeographical analysis of *Attalea*. The main objectives of this study were to build a robust dated phylogenetic tree for *Attalea* based on an extensive collection and to use this tree to: (1) increase the level of taxonomic consensus in the group; and (2) test phylogenetic and biogeographical hypotheses. More specifically, our research questions were as follows. (1) Are there monophyletic lineages within *Attalea*? (2) Was the most recent common ancestor of *Attalea* originally from the Brazilian Atlantic forest? (3) How do Amazon-forest biogeographical scenarios relate to the divergence history of *Attalea*? (4) What is the role of the Cerrado biome in divergence of *Attalea*?

MATERIAL AND METHODS

Taxon sampling and molecular markers

Most samples were collected fresh in the wild at several sites across the Brazilian Amazon, Cerrado and Atlantic forests. Additional samples were collected at Fairchild Tropical Botanic Garden and Montgomery Botanical Center or donated by specialists. For widespread species, multiple accessions from different regions of the distribution area were obtained when possible. In total, we assembled 79 accessions representing 43 putative species (Supporting Information, Table S1). To increase the species sampling and make it comparable with Meerow *et al.* (2009, 2014), we downloaded from GenBank accessions of 19 additional *Attalea* samples and five other members of the subtribe for the six included WRKY markers. Overall, the sample included 100% of the 31 species recognized by A. Henderson (unpubl. data) and 57% of the 66 species recognized by Glassman (1999). Due to amplification difficulties, some sequences were not available for all taxa (Supporting Information, Table S2). The data were then concatenated into a supermatrix combining the six WRKY datasets.

Based on previous studies with the WRKY loci (Meerow *et al.*, 2009, 2014), we chose *Allagoptera arenaria* (Gomes) Kuntze, *Beccariophoenix madagascariensis* Jum. & H.Perrier, *Cocos nucifera* L., *Jubea chilensis* (Molina) Baill. and *Syagrus sancona* H.Karst. as outgroups. Where previous work included multiple species (in *Allagoptera* and *Syagrus*), we chose the earliest-divergent species for inclusion in the outgroup. Voucher information and GenBank accession numbers are reported in Supporting Information (Table S3). Six loci of the transcription factor WRKY family (WRKY6, WRKY7, WRKY12, WRKY16, WRKY19 and WRKY21) were sampled for this study.

DNA EXTRACTION, AMPLIFICATION AND SEQUENCING

Total DNA was extracted from a 0.5-cm piece of silica-dried leaf material stored at −80 °C using an adaptation of the CTAB protocol (Doyle & Doyle, 1987). The WRKY samples were amplified using the primers reported in Meerow *et al.* (2009) and one of two different protocols. First, 2.5 μL buffer, 0.5 μL dNTPs, 0.5 μL bovine serum albumin (BSA), 17.875 μL ddH$_2$O, 0.125 μL Taq, 0.5 μL forward primer, 0.5 μL reverse primer and 1 μL template DNA, adding up to 23.5 μL. Second, 10.5 μL ddH$_2$O, 12.5 μL Taq Bullseye, 0.5 μL forward primer F, 0.5 μL reverse primer and 1 μL template DNA, adding up to 25 μL (see Supporting Information, Table S4 for primer sequences and annealing temperatures).

We amplified the samples in TC3000 and TC3000x thermal cyclers (Techne; Bibby Scientific) using the following conditions: initial denaturation of 95 °C for 2 min, [95 °C, 30 s, 58–64 °C, 30 s, 72 °C, 60 s] × 35 cycles, final extension at 72 °C for 10 min, and final hold at 4 °C. The success of the amplifications was tested by electrophoresis in a solution of TAE buffer and ethidium bromide and visualized in a 1.5% agarose gel under UV light (Bio-Rad). We sent the samples for sequencing to Beckman Coulter Genomics.

ALIGNMENT AND PHYLOGENETIC ANALYSES

Sequences were aligned using the L-INS-I algorithm in MAFFT v7.017 (Katoh *et al.*, 2002) as implemented in Geneious (Biomatters Development Team, 2005–2013; Biomatters Ltd) and edited manually. Gaps were coded as characters using the simple coding method of Simmons & Ochoterena (2000), as implemented in SeqState 1.0 (Müller, 2005). Optimal evolutionary models were tested for each region (partition) (Supporting Information, Table S4) using jModeltest 2 (Darriba *et al.*, 2012) using the Akaike information criterion (AIC) to select the best model. We assumed that each of the WRKY regions had evolved independently as they were spread across different groups of the WRKY gene family and four of them are intron-only regions. The concatenated nuclear sequences were analysed using Bayesian inference (BI) with MrBayes v3.2 (Ronquist *et al.*, 2012). BI ran for 200 million generations in two independent analyses of four Monte Carlo Markov chains (MCMCs), one ‘cold’ and three heated, and with trees sampled every 20 000 generations. Stationarity was determined using the log-likelihood scores for each run in the program Tracer v1.6 (Drummond *et al.*, 2012). Twenty-five per cent of the trees were discarded as the burn-in phase. A 50% majority rule consensus tree was calculated for the remaining trees and visualized together with posterior probabilities from MrBayes on FigTree v1.4.1 (Drummond *et al.*, 2012). All analyses were carried out on the CIPRES Portal (Miller, Pfeiffer & Schwartz, 2010).

MOLECULAR DATING AND FOSSIL CALIBRATION

Beast v1.8.1 (Drummond *et al.*, 2012) was used to estimate the divergence times assuming a log-normal uncorrelated relaxed molecular clock under the Yule tree prior. We used fossils described in Futey *et al.* (2012) to set 62 Mya as a calibration point for Attaleinae and 23.6 Mya, the age reported by Meerow *et al.* (2014), for the stem age of *Attalea*. The fossil described in Futey *et al.* (2012) is a specimen of *Tripylocarpa*, a fossil genus assigned to Cocoseae which consists of one-seeded ovoid drupes with three longitudinal grooves that might indicate the presence of three valves, and an apical germination pore and basal hilum. We used a uniform distribution of priors for the Attaleinae calibration point with 62 Mya as the lower bound and initial value and 100 Mya as the upper bound, following Eiserhardt *et al.* (2011a). For the secondary calibration point we used a normal distribution with 23.6 Mya as the mean based on Meerow *et al.* (2009), 1.0 as initial value and a standard deviation of 3. We ran the Markov chain Monte Carlo (MCMC) analysis for 200 million generations to reach ESS values > 200 for all parameters evaluated in Tracer v1.6. We sampled the MCMC every 20 000 trees and 25% were then excluded as burn-in. The remaining trees were used to obtain the posterior distribution of node ages in a maximum clade credibility (MCC) chronogram, visualized in Figtree v1.4.1 (Drummond *et al.*, 2012).

Recombination detection

We used the RDP4 (Martin & Rybicki, 2000), MaxChi (Maynard Smith, 1992) and GENECONV (Padidam, Sawyer & Fauquet, 1999) methods implemented in the RDP4 program (Martin *et al.*, 2010) to detect recombination events between samples. These methods implement a search for possible breakpoints that could indicate recombination events in triplets of sequences in the alignment and then attempt to identify the parent and the recombinant genotypes. We defined a criterion to accept the breakpoint positions and recombinant identities (Supporting Information, Table S5). First, we noted how many methods support a specific event as recommended by the authors. We excluded events that were supported by only one method. Second, we visually inspected the original alignments looking for major changes in the triplets of sequences matching with the breakpoints identified by the methods. We also compared the breakpoint positions between the different methods for congruence assessment. Third, we checked how the supposed recombinant changed position compared with the potential parents using UPGMA and maximum-likelihood trees provided by the program. A recombination event was accepted if it was supported by more than one of the three methods.

Reconstructions of ancestral areas

We used Lagrange build 20130526 (Ree & Smith, 2008) to perform an ancestral area reconstruction (hereafter AAR) for *Attalea*. This program uses a maximum-likelihood inference of geographical range evolution based on the dispersal, extinction and cladogenesis (DEC) model. Based on palaeogeography and current distribution areas for *Attalea* we chose seven areas (see Fig. 3): Central America and Hispaniola (A); Chocó including Colombian and Venezuelan valleys (B); Andes (C); Amazon forest (D); Cerrado (E); Atlantic forest (F); and Madagascar (G), to which *Beccariophoenix madagascariensis* is endemic. We decided not to subdivide the Amazon forest because of the low resolution of some Amazonian clades and because our question does not require such a level of detail. We also decided to keep Hispaniola together with Central America to avoid further subdivision in justification of a single species, *A. crassispatha* (Mart.) Burret, that is endemic to the island even though we know that those two regions have different biogeographical histories. We built a presence–absence matrix by distribution area as the input file for Lagrange. We used all the nodes of the MCC tree from Beast and we allowed the ancestral lineage to occur in four areas (maximum number of ancestral areas – max. range size parameter at the configuration file). To reconstruct the ancestral areas we implemented four time slices (62–42, 42–25, 25–11 and 11 Mya–present) with spatial and temporal constraints reflecting landscape changes in the whole distribution area through time.

First, we established no constraints for adjacent areas such as Amazon forest, Cerrado and Atlantic forest. Overall we established a 1.0 rate for adjacent areas, 0.5 for non-adjacent areas with no other barrier or for adjacent areas with a barrier and a low, but not impossible, rate of 0.01 between distant areas. We established a 0.01 rate between Central America and Chocó, increasing to 0.5 in the third time slice to reflect the first pulse of biotic interchange (Bacon *et al.*, 2015) and to 1.0 in the fourth time slice to represent the final and most significant closure. An initial rate of 0.5 was assigned between the Andes and Chocó, lowered to 0.01 in the fourth time slice indicating the full rise of the Andes. The dispersal rate involving Cerrado was kept as 0.0 until the third time slice because this region is hypothesized to have not existed before then. For the third time slice onward, it is hypothesized that there was a decrease in temperatures and rainfall causing open vegetation to start dominating the central area of Brazil, isolating the coast from the Amazon region after the closure of the Drake Passage *c.* 40 Mya (Scher & Martin, 2006; Le Roux, 2012). We thus increased the dispersal rate according to our adjacency criteria. We hypothesize that the dry belt of South America including the Cerrado, Caatinga and Chaco (Chaco being the seasonally dry forest comprising Paraguay, Bolivia and Argentina; Zak & Cabido, 2002) appeared *c.* 10 Mya. Dispersal between Madagascar and all other areas was kept at 0.01 throughout all periods of time for simplification. The dispersal cost matrices are available in Supporting Information (Table S6).

RESULTS

Phylogenetic analysis

Gene trees for single markers showed low resolution overall with the exception of WRKY16 and 21 (Supporting Information, Figs S1–S6). For the 79 samples, amplification success was relatively consistent across the markers, 68 (86%) for WRKY6, 63 (79%) for WRKY7, 67 (85%) for WRKY12, 57 (72%) for WRKY16, 53 (67%) for WRKY19 and 64 (81%) for WRKY21. The Bayesian 50% majority-rule tree (Fig. 1, see node numbering to follow description) recovered *Attalea s.l.* as monophyletic with strong support [node 2; posterior probability (PP) = 100] and *Allagoptera* was recovered as sister to the genus with strong support (node 1; PP = 100). In *Attalea*,

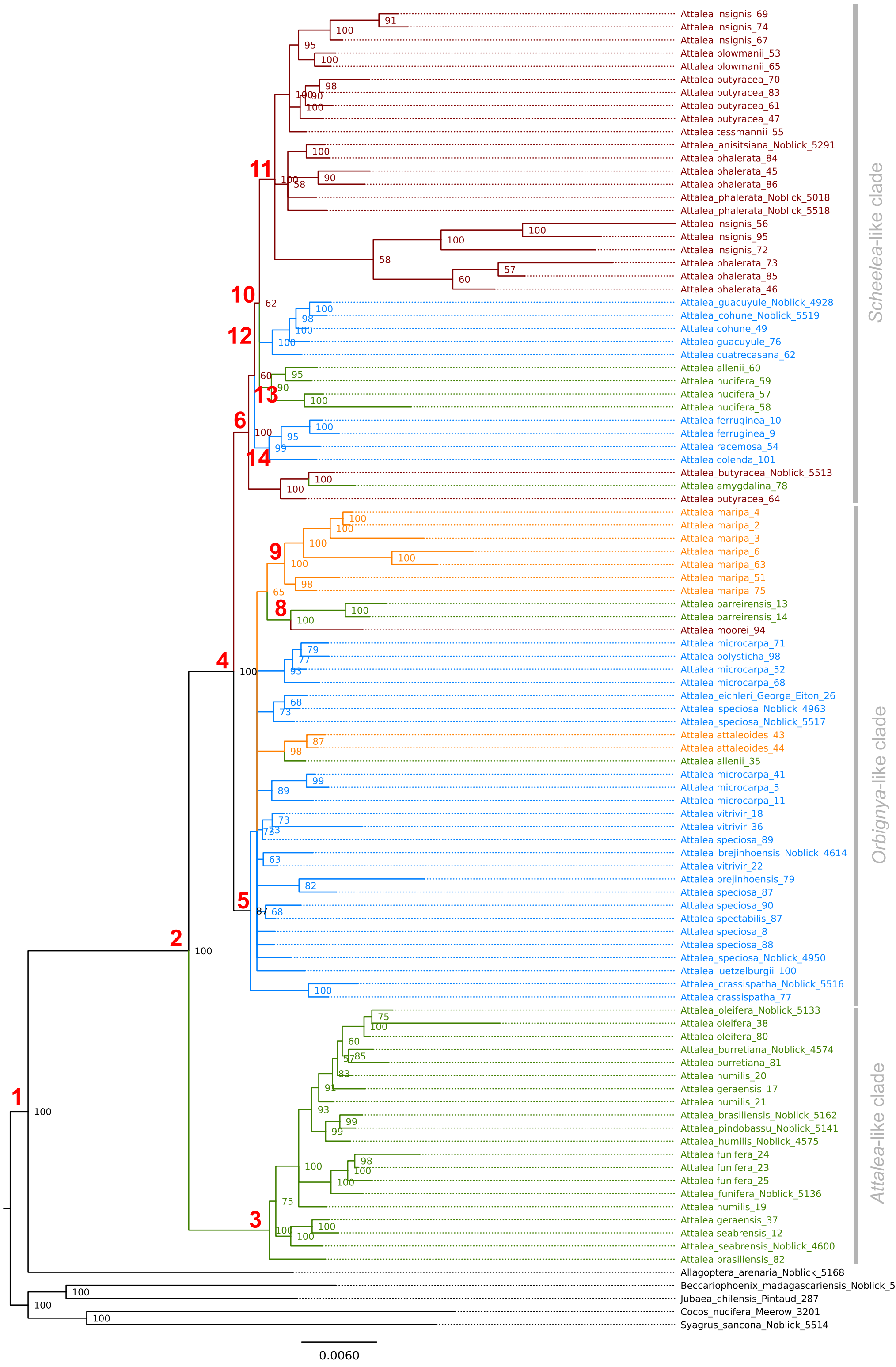

Figure 1. Bayesian 50% majority rule consensus tree of *Attalea* spp. and outgroup as a result of the concatenation of six WRKY nuclear markers in a supermatrix approach. Values above the branches are posterior probabilities. Numbers in bold red represent clades mentioned in the main text to help readers to follow results. Colours represent the groups in *Attalea* previously recognized as independent genera: *Orbignya* (blue), *Maximiliana* (orange), *Scheelea* (red) and *Attalea s.s.* (green).

we recovered three main clades, the first comprising Atlantic Coastal forest *Attalea* spp. recovered with strong support (node 3; PP = 100), the second (node 5; PP = 87) including mostly species once recognized as *Orbignya*, primarily from Amazon and Cerrado (hereafter called *Orbignya*-like), and the third (node 6; PP = 100) comprising mostly species that were once recognized as *Scheelea* (*Scheelea*-like), and also mostly *Attalea*-like *Orbignya*-like from Chocó-Amazonia (Fig. 1).

The *Orbignya*-like clade resolved with *A. crassispatha*, an endemic species from the southern peninsula of Hispaniola in the Greater Antilles, as sister to the remaining species (node 5; PP = 87). Within the clade, resolution is limited, but small well-supported clades included one uniting *A. allenii* Moore from Panama and two accessions of *A. attaleoides* (Barb.Rodr.) Wess.Boer from the central Amazon (node 7; PP = 98). Also resolved is a clade comprising *A. barreirensis* Glassman from the Brazilian Cerrado and *A. moorei* (Glassman) Zona from the Peruvian Amazon with high support (node 8; PP = 100) as sister to a *Maximiliana*-like clade (node 9; PP = 100) mostly from the Amazon basin with weak support. The remainder of this second main clade is a mix of species mostly from the Amazon and Cerrado regions.

The *Scheelea*-like clade, which has weak support (node 10; PP = 62), comprises species from Central America, the Chocó and the Amazon and includes the 'core *Scheelea*' clade (node 11; PP = 100), which has only one widespread species, *A. phalerata*, which also occurs in the Cerrado. Next we find a subclade including *A. guacuyule* (Liebm. ex Mart.) Zona, *A. cohune* Mart. and *A. cuatrecasana* (Dugand) A.J.Hend., Galeano & Bernal (node 12; PP = 100) and *A. nucifera* Karsten and *A. allenii* from Central America and the Chocó (node 13; PP = 90). Additionally, we have a strongly supported group (node 14; PP = 99) of *Orbignya*-like species, including *A. colenda* (Cook) Balslev & A.J.Hend. from the Chocó and *A. ferruginea* Burreit and *A. racemosa* Spruce from the western Amazon. The remainder of this clade forms a well-supported subclade including two samples of *A. butyracea*, paraphyletic to their conspecifics nested with *A. amygdalina*.

MOLECULAR DATING AND FOSSIL CALIBRATION

The majority of divergence events yielding current diversity of *Attalea* occurred between 15 and 5 Mya (chronogram, Fig. 2, see node numbering to follow description; Supporting Information, Fig. S7 for all values; Fig. S8 for collapsed nodes; Table S7 for dates for crown nodes of well-supported clades). The *Attalea*-like clade from the Atlantic forest is estimated to have diverged from the remainder of *Attalea s.l.* at 22.9 Mya (node 1). In the *Attalea-like* clade, most of the divergence is inferred to have taken place during the last 10 Myr. The *Scheelea*-like and *Orbignya*-like clades are estimated to have originated at 20 Mya (node 2). In the *Orbignya*-like clade, divergence of *A. crassispatha* from the ancestor of the remaining clade is estimated to have happened *c.* 17 Mya (node 3). The major divergence event giving rise to the *Maximiliana*-like clade is estimated to be *c.* 15 Mya, followed by the origin of *A. barreirensis* from the cerrado *c.* 13 Mya (node 4). Most species occurring in the Cerrado are sister to Amazonian clades and are recently diverged (< 10 Mya). In the *Scheelea*-like clade, the origin of the *A. racemosa* plus *A. colenda* clade (Chocó plus western Amazon) is estimated at 17 Mya (node 5) and the origin of the clade uniting *Attalea*-like and the *Orbignya*-like species from the Chocó is estimated to have happened *c.* 16 Mya (node 6).

DETECTION OF RECOMBINATION

Among the six markers used, only WRKY16 and WRKY21 had signals of recombination events. WRKY16 provided evidence for six events involving 12 samples representing ten species: *A. maripa*, *A. guacuyule*, *A. insignis*, *A. brejinhoensis* (Glassman) Zona, *A. barreirensis*, *A. speciosa*, *A. phalerata*, *A. nucifera*, *A. microcarpa* and *A. humilis* Mart. ex Spreng., of which only two events were accepted using our criteria (the first involving *A. brejinhoensis*_79, *A. insignis*_72 and *A. speciosa*_88 and the second involving *A. phalerata*_73, *A. insignis*_56 and *A. maripa*_63). WRKY21 had two recognized recombinant events involving six samples of five species: *A. barreirensis*, *A. phalerata*, *A. butyracea*, *A. oleifera* and *A. funifera*. We decided to exclude both recombination events involving WRKY21 because they do not hold under the criteria for event acceptance defined and described in the methods and explicitly marked in Supporting Information (Table S5). The identity of the putative parental species depends on the event and in most cases is uncertain, and in those cases the most probable close relative was chosen (i.e. *A. guacuyule*, *A. insinis*, *A. speciosa*, *A. barreirensis*, *A. maripa*, *A. microcarpa*, *A. nucifera*, *A. butyracea*, *A. phalerata* and *A. funifera*); see Supporting Information (Table S5) for details. We are aware that we were rather conservative in accepting recombination events; therefore, we have made the data available for further assessment.

RECONSTRUCTIONS OF ANCESTRAL AREAS

Ancestral area reconstruction is tentative; except for a few nodes, the average support in the chronogram

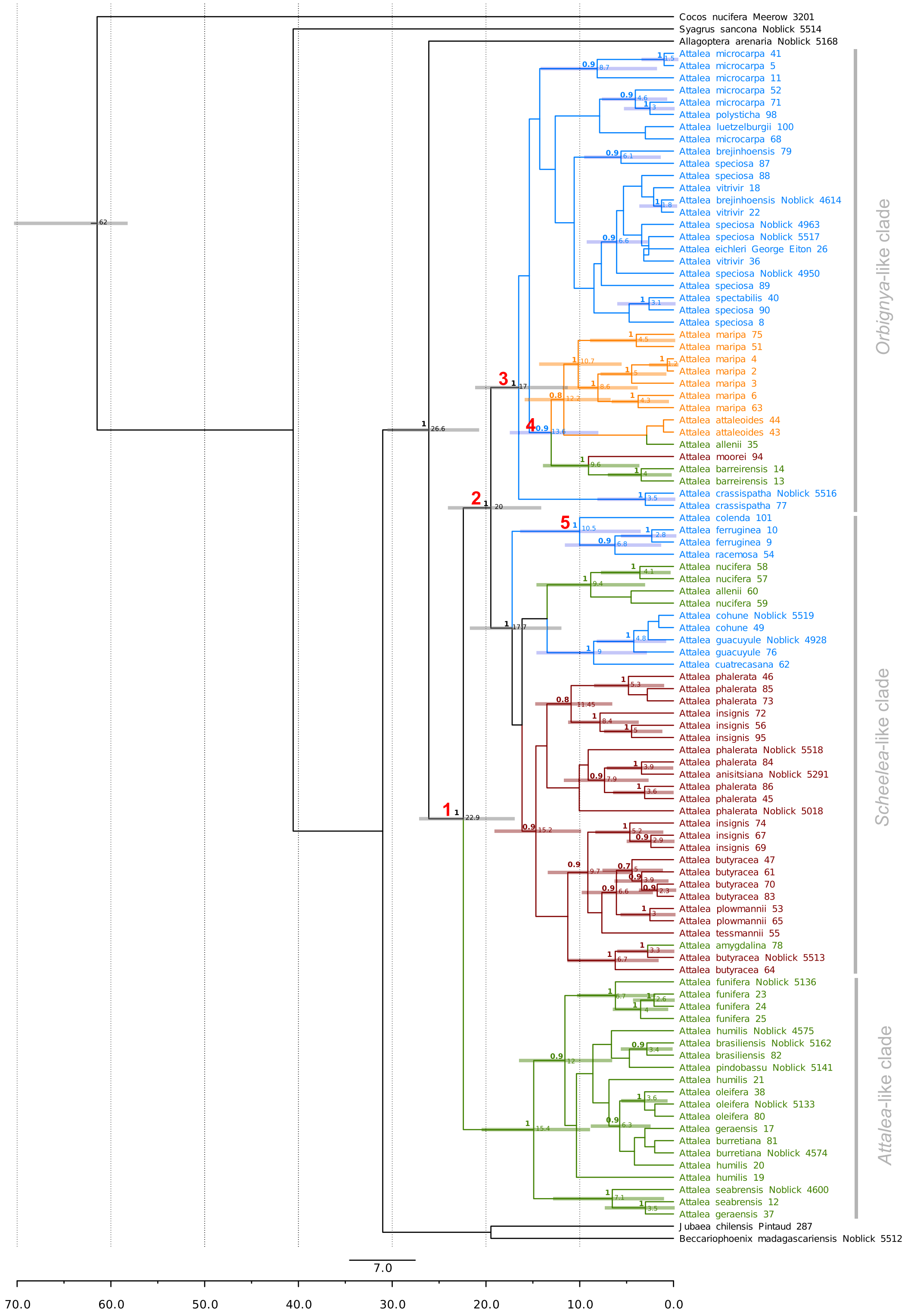

Figure 2. Maximum clade credibility tree showing the MCMC chronogram for *Attalea*. Values above the branches are posterior probabilities and values on the nodes are estimated ages. The 95% highest posterior densities are shown as bars in each node. Numbers in bold red represent clades mentioned in the main text to help readers to follow results. Colours represent the groups in *Attalea* previously recognized as independent genera: *Orbignya* (blue), *Maximiliana* (orange), *Scheelea* (red) and *Attalea s.s.* (green). The time scale is shown in millions of years.

was 50% or lower. Our analysis (Fig. 3) suggests that the common ancestor of *Attalea* was probably an Atlantic coastal forest lineage *c.* 30 Mya. We infer three major divergence events yielding lineages endemic to the cerrado, the first *c.* 15 Mya and the second when the ancestor of the *A. speciosa* complex diverged from an Amazonian lineage at *c.* 6 Mya. The third event involving the ancestor of *A. barreirensis* sister to *A. moorei* (synonym *A. cephalotus* Poepp. ex Mart.) occurred *c.* 10 Mya. The strongly supported exclusively Atlantic forest clade originated *c.* 21 Mya. A long-distance dispersal event to Hispaniola *c.* 18 Mya might explain the origin of *A. crassispatha*, endemic to the island. An exclusively Amazonian clade comprising mostly *Orbignya*-like species diverged *c.* 15 Mya, later yielding the Cerrado lineage mentioned above.

Colonization events from the Chocó region back to the Amazon might have led to a widespread clade comprising *A. maripa*, *A. allenii* from Panama and *A. attaleoides* from the central Amazon (Fig. 3). A mostly Chocó clade colonized the Brazilian Amazon again *c.* 15 Mya yielding the clade comprising *A. phalerata* and *A. insignis* there. Colonization events back and forth from the Chocó to the Amazon might have led to the widespread clade comprising *A. butyracea* and *A. insignis* from Colombia from 10 Mya onward. A mostly Chocó-region clade comprising *A. nucifera* and *A. allenii* from Colombia that is sister to *A. cohune* and *A. guacuyule* originated around 16 Mya. Another colonization from the Chocó region to the Amazon occurred around 10 Mya, leading to the clade comprising *A. ferruginea*, *A. racemosa* and *A. colenda*. Colonizations back and forth between Choco and Amazon are represented in Figure 3 with filled red circles and were considered every time there was a change from Chocó (Area B) to Amazon (Area D) or the reverse.

DISCUSSION

Attalea s.l. is well established as a monophyletic group (Meerow *et al.*, 2009, 2014), which our expanded data set corroborates. In our trees, *Allagoptera* is resolved as sister to *Attalea* in contrast to Meerow *et al.* (2014), who recovered a sister relationship with *Cocos*, although with weak support (PP = 0.7). Although the support for the relationship between *Attalea* and *Allagoptera* is strong, caution should be observed because the outgroup chosen here is a small sample of the subtribe. Also, *Parajubaea*, the strongly supported sister of *Allagoptera* in Meerow *et al.* (2009, 2014), was not included.

We can recognize four distinct groups in *Attalea* that roughly correspond to the four previously recognized genera, but they are not monophyletic. Thus there is not strong evidence supporting the separation of *Attalea* into the traditional segregate genera. These four groups have been distinguished by morphological characters related to the staminate flower (Glassman, 1999; Dransfield *et al.*, 2005), but our results suggest that these characters are homoplastic, as suggested by Dransfield *et al.* (2008). A new search for good morphological characters for lineages of *Attalea* in the light of our molecular phylogenetic analysis would be worthwhile. There is also evidence that environmental constraints at the local scale including dispersal limitation, topography and soil clay content have impacted this clade, as two *Attalea* spp. in the central Amazon are segregated along those environmental gradients (Freitas *et al.*, 2012).

The lack of resolution and weak support in parts of the tree and the evidence of recombination probably represent the effects of rapid divergence and hybridization. All would increase the phylogenetic similarity between putative species decreasing the tree resolution and possibly lead to over-description of species based on solely morphological characters. Some well-known cases of hybridization, such as *A. speciosa* × *A. vitrivir* (Henderson, 1995; Glassman, 1999) and *A. oleifera* × *A. funifera* (= *A.* × *pindobassu* Bondar, in the Atlantic forest), were not confirmed by our data. The same might be expected for the event involving *A. oleifera* and *A. pindobassu* [referred to as *A. burretiana* Bondar × *A. acaulis* Burret in Bondar (1942) and *A. burretiana* × *A. funifera* in Glassman (1999)]. Overall, we were conservative in accepting recombination events and we are aware we have sampled a small portion of the genome of these species and cannot rule out recombination as a major contributor to the lack of resolution. A more inclusive study, focused at the population level, would be a sensible next step to discover within-group relationships.

There is a clear split between the monophyletic Atlantic forest clade and the species occurring in the remaining areas (node 2; Fig. 1). Based on the chronogram and in the AAR, our results corroborate the evidence that the most recent ancestor of *Attalea* is an Atlantic forest clade (node 1 in Fig. 3) (i.e. Meerow *et al.*, 2014) and is in line with the hypothesis that its centre of origin was in the Atlantic forest. However, this remains a hypothesis, as the likelihood score (0.43) in the AAR is quite low.

Most of the divergence in *Attalea*, including the origin of the three main clades, happened in the last 20 Myr. This recent divergence coincides with major shifts in South American climatic and geographical configuration, such as the development of the Pebas Lake system in the western Amazon (Hoorn *et al.*,

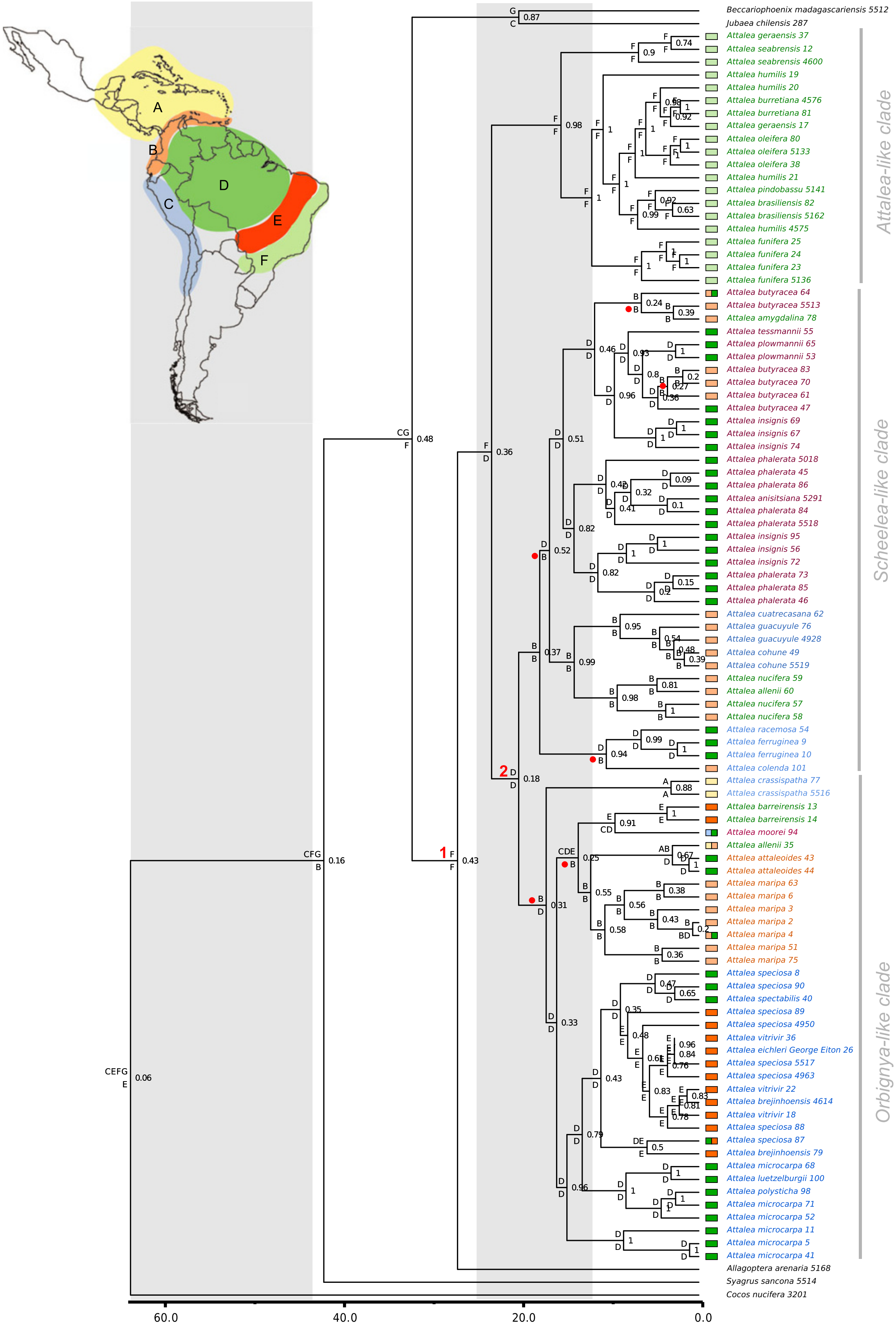
A
B
C
D
E
F
Beccariophoenix madagascariensis 5512
Jubaea chilensis 287
Attalea geraensis 37
Attalea seabrensis 12
Attalea seabrensis 4600
Attalea humilis 19
Attalea humilis 20
Attalea burretiana 4576
Attalea burretiana 81
Attalea geraensis 17
Attalea oleifera 80
Attalea oleifera 5133
Attalea oleifera 38
Attalea humilis 21
Attalea pindobassu 5141
Attalea brasiliensis 82
Attalea brasiliensis 5162
Attalea humilis 4575
Attalea funifera 25
Attalea funifera 24
Attalea funifera 23
Attalea funifera 5136
Attalea butyracea 64
Attalea butyracea 5513
Attalea amygdalina 78
Attalea tessmannii 55
Attalea plowmannii 65
Attalea plowmannii 53
Attalea butyracea 83
Attalea butyracea 70
Attalea butyracea 61
Attalea butyracea 47
Attalea insignis 69
Attalea insignis 67
Attalea insignis 74
Attalea phalerata 5018
Attalea phalerata 45
Attalea phalerata 86
Attalea anisitsiana 5291
Attalea phalerata 84
Attalea phalerata 5518
Attalea insignis 95
Attalea insignis 56
Attalea insignis 72
Attalea phalerata 73
Attalea phalerata 85
Attalea phalerata 46
Attalea cuatrecasana 62
Attalea guacuyule 76
Attalea guacuyule 4928
Attalea cohune 49
Attalea cohune 5519
Attalea nucifera 59
Attalea allenii 60
Attalea nucifera 57
Attalea nucifera 58
Attalea racemosa 54
Attalea ferruginea 9
Attalea ferruginea 10
Attalea colenda 101
Attalea crassispatha 77
Attalea crassispatha 5516
Attalea barreirensis 13
Attalea barreirensis 14
Attalea moorei 94
Attalea allenii 35
Attalea attaleoides 43
Attalea attaleoides 44
Attalea maripa 63
Attalea maripa 6
Attalea maripa 3
Attalea maripa 2
Attalea maripa 4
Attalea maripa 51
Attalea maripa 75
Attalea speciosa 8
Attalea speciosa 90
Attalea spectabilis 40
Attalea speciosa 89
Attalea speciosa 4950
Attalea vitrivir 36
Attalea eichleri George Eiton 26
Attalea speciosa 5517
Attalea speciosa 4963
Attalea vitrivir 22
Attalea brejinhoensis 4614
Attalea vitrivir 18
Attalea speciosa 88
Attalea speciosa 87
Attalea brejinhoensis 79
Attalea microcarpa 68
Attalea luetzelburgii 100
Attalea polysticha 98
Attalea microcarpa 71
Attalea microcarpa 52
Attalea microcarpa 11
Attalea microcarpa 5
Attalea microcarpa 41
Allagoptera arenaria 5168
Syagrus sancona 5514
Cocos nucifera 3201
Attalea-like clade
Scheelea-like clade
Orbignya-like clade
G C 0.87
CG F 0.48
CFG B 0.16
CEFG E 0.06
1 F F 0.43
F D 0.36
2 D D 0.18
F F 0.98
60.0
40.0
20.0
0.0

Figure 3. Maximum clade credibility tree showing the MCMC chronogram for *Attalea* performed in BEAST with the ancestral area reconstruction. Values on the nodes are likelihood support for the node and letters correspond to the areas assigned for each species in Lagrange. Central America including Greater Antilles (A), Chocó including Colombian and Venezuelan valleys (B), Andes (C), Amazon (D), Cerrado (E) and Atlantic Forest (F). Madagascar (G) is omitted from the map. Grey blocks mark the time frames set for the analysis: 62–42, 42–25, 25–11 and 11 Mya to the present. Colours represent the groups in *Attalea* previously recognized as independent genera: *Orbignya* (blue), *Maximiliana* (orange), *Scheelea* (red) and *Attalea s.s.* (green). Numbers and circles in bold red represent clades mentioned in the main text to help readers to follow results. The time scale is shown in millions of years.

2010a), sequences of uplift in the Andes (Hartley, 2003; Roig-Juñent *et al.*, 2006) and the increase in aridity across the continent (Le Roux, 2012). The increase in aridity was a result of climatic changes driven by geological events such as the successive uplift of the Andes from south to north and the opening of the Drake Passage from shallow water (43 Mya) through to full flux (30 Mya), leading to climatic changes inland (Barker & Burrell, 1977; Scher & Martin, 2006) from a warm and humid to a drier and cooler climate. This change drove the replacement of the tropical rain forest by a much drier forest and opened the possibility of colonization from the wetter Amazon to the drier Cerrado (Pennington *et al.*, 2004).

The Cerrado flora is probably recent and derived from adjacent biomes, including rain forest, as a study in Fabaceae suggests (Schrire, Lavin & Lewis, 2005). Evidence for this is also found for *Viguiera* (Schilling *et al.*, 2000), *Manihot* (Chacón *et al.*, 2008) and *Astrocaryum* (Roncal *et al.*, 2013). Indeed, the Cerrado and the adjacent rain forest have high floristic affinities: *c.* 80% of Cerrado woody species belong to rain-forest genera (Pennington, Richardson & Lavin, 2006), which could be explained by a transition between these two biomes, also supported by a dated phylogenetic analysis of *Andira* (Simon *et al.*, 2009). There is indirect evidence of colonization of the newly formed dry formations across South America followed by *in situ* diversification. This was reported for some genera typically found in the current seasonally dry tropical forests including *Coursetia* (20 Mya), *Poissonia* (18 Mya) and *Ruprechtia* (8–12 Mya) (Pennington *et al.*, 2004).

The split between two Amazonian clades (central and western Chocó) *c.* 20 Mya (node 2, Fig. 3) coincides with the development of the Pebas aquatic system, which is hypothesized to have covered most of the western Amazon basin (Hoorn *et al.*, 2010a). The *Orbignya*-like clade comprises species from the central Amazon, the Cerrado and *A. crassispatha*, endemic to Hispaniola. The exclusively Amazonian *Orbignya*-like clade originated *c.* 15 Mya, with colonization events from the Chocó back to the Amazon (filled red circles; Fig. 3). This back-and-forth colonization led to a widespread clade including species once classified as *Maximiliana* (e.g. *A. maripa*, *A. allenii* from Panama and *A. attaleoides* from the central Amazon basin). Colonization events between Chocó and the Amazon and a rapid increase in cladogenesis *c.* 10 Mya were also found in *Astrocaryum* (Roncal *et al.*, 2013). Around the same period, an increase in diversification rates was found in *Geonoma* (Arecaceae; Roncal *et al.*, 2011) and *Guatteria* (Annonaceae; Erkens *et al.*, 2007).

The *Scheelea*-like clade comprises species from the western Amazon and northern Andean valleys and two widespread species (*A. butyracea* and *A. phalerata*, which are paraphyletic). The species shared between Chocó and the Amazon may reflect the absence of a major barrier for dispersal until *c.* 10–5 Mya when the northern Andes rose to their full height (Hartley, 2003). After the full uplift and subsequent formation of the drier inter-Andean valleys this connection would have been broken.

Our analysis agrees with the hypothesis that *A. crassispatha* colonized Hispaniola *c.* 18 Mya, much later than when the Great Antilles Avian Ridge (GAAR) is thought to have existed, and much earlier than the timing traditionally accepted for Panama channel closure. In fact, comparing divergence dates between phylogenetic analyses, Cody *et al.* (2010) found that divergence dates between Central American and South American clades were older for plants than for animals, indicating that plants were less dispersal-limited and colonized South America much earlier than the supposed Panama Chanel closure timing. On the other hand, the GAAR, which is thought to have connected the Greater Antilles and South America from 35–32 Mya, probably allowed biotic interchange between those two landmasses (Iturralde-Vinent & MacPhee, 1999). After this event, the Caribbean islands are thought to have become isolated from South America until *c.* 5–2 Mya when the Panama Channel finally closed (Bacon *et al.*, 2013), indirectly linking these two regions again. There is evidence of an earlier tectonic collision between South America and Panama *c.* 23–25 Mya (Farris *et al.*, 2011), which is in line with evidence of a more complex history between North, Central and South America, with biotic interchange occurring in pulses at *c.* 23–20 and 8–6 Mya (Bacon *et al.*, 2015). Since *A. crassispatha* diverged from the Amazonian *Orbignya*-like clade *c.* 18 Mya when no connection is thought to have existed between the Greater Antilles and South America, a long-distance dispersal event is likely, as postulated by Cody *et al.* (2010). Post-GAAR colonization by rafting (Fabre *et al.*, 2014) is also possible.

Attalea is an extremely important genus not only economically, but also as a cultural expression of the people of the American tropics, and humans are likely to have partially shaped the distribution of *Attalea*. Despite the absence of literature regarding the influence of pre-Colombian and modern people on the distribution of the genus, Levis *et al.* (2012) found that the density and richness of useful trees, including palms, were artificially increased near main rivers in the Amazon. This may have significantly impacted the distribution of *Attalea* and promoted hybridization events resulting from secondary contact between formerly vicariant species. Although we cannot rule out the influence of humans on the distribution of *Attalea*, in this study we conclude that most of the divergence within the three main clades happened in the last 10 Myr (Fig. 1) roughly the same time as the arrival of humans (11 Mya) and is consistent with evidence of a recent burst in diversification rates for Meliaceae (Koenen *et al.*, 2015). This recent increase in divergence is also consistent with the increase of aridity and the consolidation of the dry diagonal including the Chaco, Cerrado and Caatinga (Pennington *et al.*, 2004). Nevertheless, deciduous-forest lineages of Meliaceae did the opposite of what we are suggesting for *Attalea* (i.e. they migrated from drier ecosystems to tropical rain forests). In *Attalea*, many species occurring in the cerrado such as *A. vitrivir*, *A. brejinhoensis* and *A. eichleri* (Drude) A.J.Hend. are younger than and sister to Amazonian clades, indicating colonization from Amazonia to the new niche formed by the dry diagonal of South America.

Finally, some *Attalea* spp. appear to be polyphyletic, the most striking cases being *A. allenii*, *A. insignis*, *A. phalerata* and *A. butyracea*. *Attalea allenii* from Colombia is sister to *A. nucifera* from the inter-Andean Valleys, whereas *A. allenii* from Panama is sister to *A. attaleoides* from the central Amazon. The only support (PP = 93) for monophyly between the two samples of *A. allenii* is in the WRKY16 alignment (considering that we were not able to amplify those two samples for WRKY21). The distance between these two samples is probably real, as both clades are resolved with high support in the MCC chronogram. Perhaps this pattern reflects the Andean uplift segregating the older inter-Andean–Colombian clade (9.4 Mya) from the remaining Panama–central Amazon clade (3.4 Mya). Another possibility is a hybridization event between *A. allenii* and other species that we were not able to detect.

The polyphyletic species *A. phalerata* and *A. insignis* resolve in the *Scheelea*-like clade, which split into two subclades *c.* 16 Mya (Fig. 2). *Attalea insignis* segregates between geological formations (samples 67, 69 and 74 on an older formation and samples 56, 72 and 95 on a younger formation). The species is restricted to the alluvial terraces in the western Amazon; its isolation there is consistent with the presence of a geological feature known as the Purus arch, which is hypothesized to have divided the Amazon and Solimões drainage basins (Wanderley-Filho *et al.*, 2010). This split is also consistent with a geological formation reported by Higgins *et al.* (2012) that is supposed to have separated the Pebas and the Iça formation in the western Amazon and is evidenced by significant floristic differences between those two areas (Higgins *et al.*, 2011).

The substantial problems with species delimitation that we have revealed in this study add another layer of complexity to tracing the history of *Attalea*. In *A. butyracea* and *A. phalerata*, two widespread and polyphyletic species, taxonomic investigation is needed, although morphology is consistent throughout samples covering their distribution arc. Additionally, in the Kew Checklist the two species are divided into 13 and six different taxa, respectively (see Supporting Information, Table S8 for correspondence). Cryptic biological species may well be at the root of the problem; further investigation should focus on the population level coupled with other sources of data to help reveal the species delimitations and the processes that yielded this complexity.

CONCLUSIONS

Using an inclusive sample of the genus, we reject the monophyly of the previous recognized segregate genera within *Attalea.* Three main clades can be recognized; they roughly correspond to previously recognized but artificial genera. These clades, however, largely lie in ecogeographical biomes, such as the Atlantic forest, Amazon, Cerrado and northern Mesoamerican regions. The Atlantic forest clade, sister to the remaining two mostly Amazonian clades, is consistent with the previously proposed Atlantic forest origin for *Attalea*. An alternative scenario of a widespread clade that went through geologically induced vicariance is possible. The geologically recent development of dry forest biomes also played a significant role in the speciation of *Attalea*. The polyphyly of some species (e.g. *A. allenii*, *A. insignis*, *A. butyracea* and *A. phalerata*), the evidence of recombination events and the overall low resolution at recent nodes are impediments to finalizing a stable taxonomy for the genus. Although it is clear that more effort is needed to resolve interspecific relationships in *Attalea*, this study offers a significant starting point for future progress in the systematics and biogeography of this important group.

ACKNOWLEDGEMENTS

C.G.F. wishes to thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for financial support (481297/2011-1 granted to F.R.C.C.), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior/Programa de Pós Doutorado no Exterior (Processo no. 3138-13-5) for a Postdoc fellowship granted to C.G.F., and Flávia Santana, Thomas Couvreur, Glória Galeano, Rodrigo Bernal, Finn Borchsenius, Thaise Emilio, Henrik Balslev, Sarah Reis and Camila Duarte for sample donations.

REFERENCES

Anderson AB, May PH, Balick MJ. 1991. *The subsidy from nature: palm forests, peasantry, and development on an Amazon frontier*. New York: Columbia University Press.

Asmussen C, Dransfield J, Deickmann V, Barfod AS, Pintaud J, Baker WJ. 2006. A new subfamily classification of the palm family (Arecaceae): evidence from plastid DNA phylogeny. *Botanical Journal of the Linnean Society* **151:** 15–38.

Bacon CD, Mora A, Wagner WL, Jaramillo CA. 2013. Testing geological models of evolution of the Isthmus of Panama in a phylogenetic framework. *Botanical Journal of the Linnean Society* **171:** 287–300.

Bacon CD, Silvestro D, Jaramillo C, Smith BT, Chakrabarty P, Antonelli A. 2015. Biological evidence supports an early and complex emergence of the Isthmus of Panama. *Proceedings of the National Academy of Sciences of the United States of America* **112:** 1–6.

Baker W, Savolainen V, Asmussen C, Chase MW, Dransfield J, Forest F, Harley M, Uhl N, Wilkinson M. 2009. Complete generic-level phylogenetic analyses of palms (Arecaceae) with comparisons of supertree and supermatrix approaches. *Systematic Biology* **58:** 240–256.

Baker WJ, Couvreur TLP. 2013a. Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. Historical biogeography. *Journal of Biogeography* **20:** 1–12.

Baker WJ, Couvreur TLP. 2013b. Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. II. Diversification history and origin of regional assemblages. *Journal of Biogeography* **40:** 286–298.

Balslev H, Grandez C, Zambrana P, Møller AL, Hansen SL. 2008. Palmas (Arecaceae) útiles en los alrededores de Iquitos, Amazonía Peruana. *Revista Peruana de Biología* **15:** 121–132.

Barker PF, Burrell J. 1977. The opening of Drake Passage. *Marine Geology* **25:** 15–34.

Barrett CF, Baker WJ, Comer JR, Conran JG, Lahmeyer SC, Leebens-Mack JH, Li J, Lim GS, Mayfield-Jones DR, Perez L, Medina J, Pires JC, Santos C, Stevenson D, Zomlefer WB, Davis JI. 2015. Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. *New Phytologist* **209:** 855–870.

Bernal R, Galeano G. 2010. Uses and commercial prospects for the wine palm, *Attalea butyracea*, in Colombia. *Ethnobotany* **8:** 255–268.

Bjorholm S, Svenning J, Baker W. 2006. Historical legacies in the geographical diversity patterns of New World palm (Arecaceae) subfamilies. *Botanical Journal of the Linnean Society* **151:** 113–125.

Bondar G. 1942. New palms of Bahia. *Publications of the Field Museum of Natural History. Botanical Series* **22:** 457–463.

Cannon SB, Mitra A, Baumgarten A, Young ND, May G. 2004. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biology* **4:** 1–21.

Chacón J, Madriñán S, Debouck D, Rodriguez F, Tohme J. 2008. Phylogenetic patterns in the genus *Manihot* (Euphorbiaceae) inferred from analyses of nuclear and chloroplast DNA regions. *Molecular Phylogenetics and Evolution* **49:** 260–267.

Cody S, Richardson JE, Rull V, Ellis C, Pennington RT. 2010. The great American biotic interchange revisited. *Ecography* **33:** 326–332.

Couvreur TLP, Forest F, Baker W. 2011. Origin and global diversification patterns of tropical rain forests: inferences from a complete genus-level phylogeny of palms. *BMC Biology* **9:** 44.

Crisp MD, Trewick S, Cook LG. 2011. Hypothesis testing in biogeography. *Trends in Ecology & Evolution* **26:** 66–72.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9:** 772.

Doyle J, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin, Botanical Society of America* **19:** 11–15.

Dransfield J, Uhl N, Asmussen C, Baker W. 2005. A new phylogenetic classification of the palm family, Arecaceae. *Kew Bulletin* **60:** 559–569.

Dransfield J, Uhl N, Asmussen C, Baker W, Harley M, Lewis CE. 2008. *Genera palmarum*. Kew: Royal Botanic Gardens.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29:** 1969–1973.

Eiserhardt WL, Pintaud JC, Asmussen-Lange C, Hahn WJ, Bernal R, Balslev H, Borchsenius F. 2011a. Phylogeny and divergence times of Bactridinae (Arecaceae, Palmae) based on plastid and nuclear DNA sequences. *Taxon* **23:** 1–14.

Eiserhardt WL, Svenning JC, Kissling WD, Balslev H. 2011b. Geographical ecology of the palms (Arecaceae): determinants of diversity and distributions across spatial scales. *Annals of Botany* **108:** 1391–1416.

Erkens RHJ, Chatrou LW, Maas JW, van der Niet T, Savolainen V. 2007. A rapid diversification of rainforest trees (*Guatteria*; Annonaceae) following dispersal from Central into South America. *Molecular Phylogenetics and Evolution* **44:** 399–411.

Fabre PH, Vilstrup JT, Raghavan M, Sarkissian CD, Willerslev E, Douzery EJP, Orlando L. 2014. Rodents of the Caribbean: origin and diversification of hutias unraveled by next-generation museomics. *Biology Letters* **10:** 20140266.

Farris DW, Jaramillo C, Bayona G, Restrepo-Moreno SA, Montes C, Cardona A, Mora A, Speakman RJ, Glascock MD, Valencia V. 2011. Fracturing of the Panamanian Isthmus during initial collision with South America. *Geology* **39:** 1007–1010.

Freitas C, Costa FRC, Svenning JC, Balslev H. 2012. Topographic separation of two sympatric palms in the central Amazon – does dispersal play a role? *Acta Oecologica* **39:** 128–135.

Futey MK, Gandolfo MA, Zamaloa MC, Cúneo R, Cladera G. 2012. Arecaceae fossil fruits from the Paleocene of Patagonia, Argentina. *The Botanical Review* **78:** 205–234.

Glassman SF. 1999. *A taxonomic treatment of the palm subtribe Attaleinae (tribe Cocoeae) 59*. Urbana, IL: University of Illinois Press.

Hartley A. 2003. Andean uplift and climate change. *Journal of the Geological Society* **160:** 7–10.

Henderson A. 1995. *The palms of the Amazon*. Oxford: Oxford University Press.

Higgins MA, Asner GP, Perez E, Elespuru N, Tuomisto H, Ruokolainen K, Alonso A. 2012. Use of Landsat and SRTM data to detect broad-scale biodiversity patterns in northwestern Amazonia. *Remote Sensing* **4:** 2401–2418.

Higgins MA, Ruokolainen K, Tuomisto H, Llerena N, Cardenas G, Phillips OL, Vásquez R, Räsänen M. 2011. Geological control of floristic composition in Amazonian forests. *Journal of Biogeography* **38:** 2136–2149.

Hoorn C, Roddaz M, Dino R, Soares E, Uba C, Ochoa-Lozano D, Mapes R. 2010a. The Amazonian Craton and its influence on past fluvial systems (Mesozoic–Cenozoic, Amazonia). In: Hoorn C, Wesselingh F, eds. *Amazonia: landscape and species evolution*. Oxford: Blackwell Publishing, 103–122.

Hoorn C, Wesselingh FP, ter Steege H, Bermudez MA, Mora A, Sevink J, Sanmartín I, Sanchez-Meseguer A, Anderson CL, Figueiredo JP, Jaramillo C, Riff D, Negri FR, Hooghiemstra H, Lundberg J, Stadler T, Särkinen T, Antonelli A. 2010b. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science* **330:** 927–931.

Hopkins MJG. 2007. Modeling the known and unknown plant biodiversity of the Amazon Basin. *Journal of Biogeography* **34:** 1400–1411.

Iturralde-Vinent M, MacPhee RDE. 1999. Paleogeography of the Caribbean region: implications for Cenozoic biogeography. *Bulletin of the American Museum of Natural History* **238:** 1–95.

Jordan TE, Farjat AD, Echavarrı L, Idleman BD, Reynolds JH. 2005. Age, distribution, tectonics, and eustatic controls of the Paranense and Caribbean marine transgressions in southern Bolivia and Argentina. *Journal of South American Earth Sciences* **19:** 495–512.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30:** 3059–3066.

Koenen EJM, Clarkson JJ, Pennington TD, Chatrou LW. 2015. Recently evolved diversity and convergent radiations of rainforest mahoganies (Meliaceae) shed new light on the origins of rainforest hyperdiversity. *New Phytologist* **207:** 327–339.

Latrubesse EM, Cozzuol M, da Silva-Caminha SAF, Rigsby CA, Absy ML, Jaramillo C. 2010. The Late Miocene paleogeography of the Amazon Basin and the evolution of the Amazon River system. *Earth-Science Reviews* **99:** 99–124.

Le Roux JP. 2012. A review of Tertiary climate changes in southern South America and the Antarctic Peninsula. Part 1: oceanic conditions. *Sedimentary Geology* **247:** 1–20.

Levis C, Souza P, Schietti J, Emilio T, Pinto J, Clement CR, Costa FRC. 2012. Historical human footprint on modern tree species composition in the Purus–Madeira interfluve, Central Amazonia. *PLoS ONE* **7:** e48559.

Martin D, Rybicki E. 2000. RDP: detection of recombination amongst aligned sequences. *Bioinformatics* **16:** 562–563.

Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P. 2010. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26:** 2462–2463.

Maynard Smith J. 1992. Analyzing the mosaic structure of genes. *Journal of Molecular Evolution* **34:** 126–129.

Meerow AW, Noblick L, Borrone JW, Couvreur TLP, Mauro-Herrera M, Hahn WJ, Kuhn DN, Nakamura K, Oleas NH, Schnell RJ. 2009. Phylogenetic analysis of seven WRKY genes across the palm subtribe Attaleinae (Arecaceae) identifies *Syagrus* as sister group of the coconut. *PLoS ONE* **4:** e7353.

Meerow AW, Noblick L, Salas-Leiva DE, Sanchez V, Francisco-Ortega J, Jestrow B, Nakamura K. 2014. Phylogeny and historical biogeography of the cocosoid palms (Arecaceae, Arecoideae, Cocoseae) inferred from sequences of six WRKY gene family loci. *Cladistics* **31:** 1–26.

Miller MA, Pfeiffer W, Schwartz T 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA, 1–8.

Müller DK. 2005. SeqState. *Applied Bioinformatics* **4:** 65–69.

Myers N, Mittermeier RA, Mittermeier CG, Fonseca GAB, Kent J. 2000. Biodiversity hotspots for conservation priorities. *Nature* **403:** 853–858.

Noblick L. 2013. Leaflet anatomy verifies relationships within *Syagrus* (Arecaceae) and aids in identification. *PhytoKeys* **26:** 75–99.

Noblick L, Meerow A. 2015. The transfer of the genus *Lytocaryum* to *Syagrus*. *Palms* **59:** 57–62.

Oliveira-Filho AT, Ratter JA. 2002. Vegetation physiognomies and woody flora of the cerrado biome. In: Oliveira PS, Marquis RJ, eds. *The cerrados of Brazil: ecology and natural history of a Neotropical savanna*. New York: Columbia University Press, 91–120.

Padidam M, Sawyer S, Fauquet CM. 1999. Possible emergence of new geminiviruses by frequent recombination. *Virology* **265:** 218–225.

Pennington RT, Lavin M, Prado DE, Pendry CA, Pell SK, Butterworth CA. 2004. Historical climate change and speciation: neotropical seasonally dry forest plants show patterns of both Tertiary and Quaternary diversification. *Philosophical Transactions of the Royal Society B: Biological Sciences* **359:** 515–538.

Pennington RT, Richardson JE, Lavin M. 2006. Insights into the historical construction of species-rich biomes from dated plant phylogenies, neutral ecological theory and phylogenetic community structure. *New Phytologist* **172:** 605–616.

Pinedo AS, Martins RC, de Oliveira RC, Gomes SM. 2016. Leaf anatomy in Allagoptera (Arecaceae). *Botanical Journal of the Linnean Society* doi:10.1111/boj.12439.

Pintaud JC. 2008. An overview of the taxonomy of *Attalea* (Arecaceae). *Revista Peruana de Biología* **15:** 53–62.

Ree RH, Smith SA. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* **57:** 4–14.

Ribeiro JF, Walter BMT. 1998. Fitofisionomias do bioma cerrado. In: Sano SM, Almeida SP, eds. *Cerrado: ambiente e flora*. Brasília: EMBRAPA, 91–166.

Roig-Juñent S, Domínguez MC, Flores GE, Mattoni C. 2006. Biogeographic history of South American arid lands: a view from its arthropods using TASS analysis. *Journal of Arid Environments* **66:** 404–420.

Roncal J, Blach-Overgaard A, Borchsenius F, Balslev H, Svenning JC. 2011. A dated phylogeny complements macroecological analysis to explain the diversity patterns in *Geonoma* (Arecaceae). *Biotropica* **43:** 324–334.

Roncal J, Borchsenius F, Asmussen C, Balslev H. 2010. Divergence times in the tribe Geonomateae (Arecaceae) coincide with Tertiary geological events. In: Seberg O, Peterson AT, Barfod A, Davis AM, eds. *Diversity, phylogeny and evolution in the monocotyledons*. Aarhus: Århus University Press, 245–265.

Roncal J, Francisco-Ortega J, Asmussen CB, Lewis CE. 2005. Molecular phylogenetics of tribe Geonomeae (Arecaceae) using nuclear DNA sequences of phosphoribulokinase and RNA polymerase II. *Systematic Botany* **30:** 275–283.

Roncal J, Kahn F, Millan B, Couvreur TL, Pintaud JC. 2013. Cenozoic colonization and diversification patterns of tropical American palms: evidence from *Astrocaryum* (Arecaceae). *Botanical Journal of the Linnean Society* **171:** 120–139.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61:** 539–542.

Ross CA, Liu Y, Shen QJ. 2007. The *WRKY* gene family in rice *Oryza sativa*. *Journal of Integrative Plant Biology* **49:** 827–842.

Scher HD, Martin EE. 2006. Timing and climatic consequences of the opening of Drake Passage. *Science* **312:** 428–430.

Schilling EE, Da Costa FB, Lopes NP, Heise PJ. 2000. Brazilian species of *Vigueira* (Asteraceae) exhibit low levels of ITS sequence variation. *Edinburgh Journal of Botany* **57:** 323–332.

Schrire BD, Lavin M, Lewis GP. 2005. Global distribution patterns of the Leguminosae: insights from recent phylogenies. In: Friis I, Balslev H, eds. *Plant diversity and complexity patterns: local, regional, and global dimensions*. Proceedings of an International Symposium. Copenhagen: Royal Danish Academy of Sciences and Letters, 367–422.

Simmons MP, Ochoterena H. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* **49:** 369–381.

Simon MF, Grether R, de Queiroz LP, Skema C, Pennington RT, Hughes CE. 2009. Recent assembly of the cerrado, a Neotropical plant diversity hotspot, by *in situ* evolution of adaptations to fire. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 20359–20364.

Terborgh J. 1992. Maintenance of diversity in tropical forests. *Biotropica* **24:** 283–292.

Tomlinson P. 2006. The uniqueness of palms. *Botanical Journal of the Linnean Society* **151:** 5–14.

Wanderley-Filho J, Eiras J, Cunha P, van der Ven P. 2010. The Paleozoic Solimões and Amazonas basins and the Acre foreland basin of Brazil. In: Hoorn C, Wesselingh FP, eds. *Amazonia: landscape and species evolution*. Oxford: Blackwell Publishing, 29–37.

WCSP. 2015. *World checklist of selected plant families*. Kew: Royal Botanic Gardens.

Wu KL, Guo ZJ, Wang HH, Li J. 2005. The WRKY family of transcription factors in rice and arabidopsis and their origins. *DNA Research* **12:** 9–26.

Zak MR, Cabido M. 2002. Spatial patterns of the Chaco vegetation of central Argentina: integration of remote sensing and phytosociology. *Applied Vegetation Science* **5:** 213–226.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Figure S1. Bayesian 50% majority rule consensus tree of *Attalea* spp. and outgroup for the WRKY6 nuclear marker.

Figure S2. Bayesian 50% majority rule consensus tree of *Attalea* spp. and outgroup for the WRKY7 nuclear marker.

Figure S3. Bayesian 50% majority rule consensus tree of *Attalea* spp. and outgroup for the WRKY12 nuclear marker.

Figure S4. Bayesian 50% majority rule consensus tree of *Attalea* spp. and outgroup for the WRKY16 nuclear marker.

Figure S5. Bayesian 50% majority rule consensus tree of *Attalea* spp. and outgroup for the WRKY19 nuclear marker.

Figure S6. Bayesian 50% majority rule consensus tree of *Attalea* spp. and outgroup for the WRKY21 nuclear marker.

Figure S7. Maximum clade credibility tree showing the MCMC chronogram for *Attalea*.

Figure S8. Maximum clade credibility tree showing the MCMC chronogram for *Attalea*.

Table S1. Silica-dried samples from fresh collected and donated samples of the genus *Attalea* used to extract DNA.

Table S2. Number of *Attalea* samples included for each marker and final supermatrix with the size of the fragment amplified.

Table S3. Voucher information and GenBank accession numbers for *Attalea* samples included in the study.

Table S4. WRKY primer sequences (5′→3′).

Table S5. Recombination events showing the number of methods supporting and the decision making towards the acceptance of each event for each marker with recombination signs.

Table S6. Dispersal cost matrices used as an input to Lagrange for ancestral area reconstruction showing each time frame at the headings of each table section.

Table S7. Dates for crown nodes of well-supported clades in the MCC tree.

Table S8. The 125 records retrieved from the Kew Checklist showing the nomenclature correspondence between this study and the Kew Checklist for *Attalea*.