Botanical Journal of the Linnean Society, 2021, **197**, 61–84. With 5 figures.

Skipping the Dry Diagonal: spatio-temporal evolution of *Croton* section *Cleodora* (Euphorbiaceae) in the Neotropics

IRENE MASA-IRANZO[1], ISABEL SANMARTÍN[1,†], MARIA BEATRIZ R. CARUZO[2] and RICARDA RIINA[1,*,†]

[1]*Real Jardín Botánico, CSIC, Plaza de Murillo 2, Madrid 28014, Spain*
[2]*Universidade Federal de São Paulo, Departamento de Ecologia e Biologia Evolutiva, Rua Prof. Artur Riedel, 275, 09972-270, Diadema, SP, Brazil*

Received 16 August 2020; revised 30 November 2020; accepted for publication 21 January 2021

Croton is one of the largest angiosperm genera, with > 1200 species in the tropics worldwide. The arborescent *Croton* section *Cleodora* stands out for its disjunct geographical distribution with two main centres of diversity in the Amazonian and Atlantic Forest regions, separated by the Caatinga, Cerrado and Chaco biomes (the 'Dry Diagonal'). This disjunction is found in other Neotropical lineages and attributed to Neogene geological and climatic events. We inferred a nearly complete phylogenetic reconstruction of *Croton* section *Cleodora* based on DNA sequences of nuclear ITS and five plastid regions (*rps16*, *trnH-psbA*, *trnL-F*, *trnT-L* and *ycf1*). We further estimated divergence times and reconstructed ancestral ranges using Bayesian methods. Our results show that *Croton* section *Cleodora* is monophyletic with two main clades; we also confirm the phylogenetic adscription of eight *Croton* spp. recently described or assigned to section *Cleodora* based on morphology. Divergence of section *Cleodora* from its sister clade occurred *c.* 25 Mya, and diversification within this group started 20 Mya. Biogeographic analyses suggest the section originated in the Amazonian region, from where it dispersed to other forested Neotropical regions, including the Atlantic Forest. Divergence between Amazonian and Atlantic Forest taxa appears to have been triggered by the formation of the South American Dry Diagonal.

ADDITIONAL KEYWORDS: Amazonia – Andes – Atlantic Forest – Bayesian biogeography – Caatinga – Cerrado – Chaco.

INTRODUCTION

Understanding the factors and processes underlying current patterns of biodiversity and their evolution in space and time is the main objective of biogeography (Lomolino *et al.*, 2016). One of the most intriguing patterns in biogeography is the huge concentration of species in tropical regions compared to temperate and subtropical regions. In particular, the Neotropics is considered the most diverse region of the planet (Gentry, 1982; Dirzo & Raven, 2003; Antonelli & Sanmartín, 2011; Antonelli *et al.*, 2018), and in the Neotropics the Amazon Basin, the Andes and the Atlantic Forest stand out as the main centres of diversity (Myers *et al.*, 2000; Morrison *et al.*, 2001; Barthlott *et al.*, 2005; da Silva, Rylands & da Fonseca, 2005). At the same time, these regions are some of the most threatened areas on Earth, with their conservation considered a high priority (Myers *et al.*, 2000; Ribeiro *et al.*, 2009).

Although an enormous progress has been made in the last decades to improve the knowledge on systematics and distribution patterns of Neotropical plant groups, specially stemming from the use of molecular data (Givnish *et al.*, 2011; Van Ee, Riina & Berry, 2011; Michelangeli *et al.*, 2013; Calió *et al.*, 2017; de Faria *et al.*, 2017; Guo *et al.*, 2017; Lucas *et al.*, 2018; Bogarín *et al.*, 2019), there are still numerous groups where a significant part of their taxonomy and phylogenetic knowledge remains incomplete (Cardoso *et al.*, 2017; Ulloa *et al.*, 2017). Biogeographic and diversification analyses of important Neotropical plant lineages have highlighted the need to integrate sound molecular phylogenetic hypotheses with a deeper understanding

*Corresponding author. E-mail: rriina@rjb.csic.es
†Co-senior authors.

of the geological and palaeoclimatic history (Table 1) of the region (Antonelli *et al.*, 2009; Hoorn *et al.*, 2010; Antonelli & Sanmartín, 2011; Arakaki *et al.*, 2011; Perret *et al.*, 2013; Givnish *et al.*, 2014; Hernández-Hernández *et al.*, 2014; Lagomarsino *et al.*, 2016; Canal *et al.*, 2019; Lavor *et al.*, 2019). The previous conception of a biota shaped by Pleistocene glacial events (Haffer, 1969) has been gradually replaced by a more complex picture, in which several sequential events from the Early Palaeocene to the Pliocene led to changes in migration, speciation and extinction rates, such as the uplift of the northern Andes, the formation of the Central American Landbridge, or the appearance of mega-wetlands in western Amazonia resulting from changes in the Amazon River system (Hoorn *et al.*, 2010; Bacon *et al.*, 2013, 2016; Antonelli *et al.*, 2018; Thode, Sanmartín & Lohmann, 2019). Some of these events have been under debate regarding their age. For example, a geological time frame for the formation of the Panama Isthmus and the closure of the Central American Seaway is still controversial (Montes *et al.*, 2015; O'Dea *et al.*, 2016; Jaramillo, 2018). Biological studies can help provide evidence to date the above-mentioned events, and therefore the construction of biomes (Baker *et al.*, 2014), as long as the geological events are not used in the phylogenetic dating (e.g. Sanmartín, Enghoff & Ronquist, 2001; Bacon *et al.*, 2015).

One of the richest genera in number of species in the tropics is *Croton* L. (Euphorbiaceae), with > 1200 species worldwide (Berry *et al.*, 2005) and *c.* 770 Neotropical species (Van Ee *et al.*, 2011; Ulloa *et al.*, 2017). Although Neotropical *Croton* lineages have been classified in sections using molecular phylogenetics as a basis (Van Ee *et al.*, 2011), little progress has been made in the understanding of phylogenetic relationships in its 31 sections or clades, especially those with the highest species diversity (Van Ee *et al.*, 2011; Sodré *et al.*, 2019).

Croton has always been regarded as a genus more common in dry-mesic open habitats (Van Ee *et al.*, 2011), however, some of its richest clades occupy predominantly wet forested habitats (Riina, Berry & Van Ee, 2009; Caruzo *et al.*, 2011). In this work, we revisit *Croton* section *Cleodora* (Klotzsch) Baill., one of the main arborescent lineages of the genus in the Neotropics (Caruzo *et al.*, 2011). Besides their predominantly arborescent growth form, the species in this section are characterized by the presence of latex, extrafloral nectaries, bisexual tirsoid-racemose inflorescences, staminate flowers with 15–25 stamens, pistillate flowers with imbricate or quincuncial aestivation (rarely valvate) and multifid styles fused proximally forming a basal column (Caruzo & Cordeiro, 2013) (Fig. 1). Among the species of section *Cleodora*, *Croton cajucara* Benth. ('sacaca'; Fig. 1G) stands out as an important medicinal plant, commonly used by the inhabitants of the Amazon Basin (Salatino, Salatino & Negri, 2007; Secco, 2008; Nascimento *et al.*, 2017).

In the most recent taxonomic revision of section *Cleodora*, Riina *et al.* (2018) described six new species and reassigned, using morphology alone, two species previously included by Van Ee *et al.* (2011) in other sections of *Croton*, thus increasing the number of species in section *Cleodora* from 18 to 27. Based on patterns of species diversity and distribution, Riina *et al.* (2018) also suggested that section *Cleodora* could have originated in the Amazon region. Species in this group are distributed from Mexico and Central America to south-eastern Brazil (Fig. 2, Appendix S1 in the Supporting Information), with the Amazon Basin and the Atlantic Forest as the main centres of diversity (Freitas dos Santos, Riina & Caruzo, 2017; Riina *et al.*, 2018). The section exhibits a disjunct north-south distribution with the Dry Diagonal as a geographical and climatic barrier. This wide corridor of seasonally dry vegetation encompasses three biomes that are different in species composition, ecology and history (de Queiroz *et al.*, 2017): the Argentinian-Paraguayan Chaco, the Cerrado and the Caatinga (Prado & Gibbs, 1993; Neves *et al.*, 2015; da Silva, Leal & Tabarelli, 2017), and it includes the largest area of seasonally dry tropical forest (SDTF) in the Americas (DRYFLOR, 2016; Fernandes, Cardoso & de Queiroz, 2020). Similar geographical disjunctions around the Dry Diagonal are found in other Neotropical plant and animal lineages and have been attributed to major geological and climatic events that took place during the Neogene, such as the Late Oligocene Warming Event (LOWE), the Mid-Miocene Climate Optimum (MMCO) or the Mid-Pliocene Warming Event (MPWE) (Table 1; Orme, 2007; Batalha-Filho *et al.*, 2013; Sobral-Souza *et al.*, 2015; Prates *et al.*, 2017; Thode *et al.*, 2019; Peres *et al.*, 2020). Formation of the Dry Diagonal probably began in the Late Eocene, coincident with a global cooling event that brought about colder and drier climates in central and southern South America. On the other hand, molecular phylogenetic studies and the fossil record support the existence of frequent re-connections between the Atlantic Forest and the Amazonian region via forest patches across the Dry Diagonal during the Miocene and Plio-Pleistocene (Prado & Gibbs, 1993; Prates *et al.*, 2017; Fine & Lohman, 2018; Peres *et al.*, 2020).

The only previous molecular phylogenetic analysis of section *Cleodora* (Caruzo *et al.*, 2011), including 15 of the 18 recognized species at the time, was based on three genetic regions (the nuclear ITS and the plastid interspacers *trnH-psbA* and *trnL-F*), and recovered two main clades. *Croton* subsection

Table 1. Summary of the main geological and climatic events and formation of main biomes in tropical America relevant in the evolutionary context of *Croton* section *Cleodora*. Events are ordered by their estimated age. Geological and biological studies supporting the timing of these events are provided

Event	Estimated age (Mya)	Geological period	References
Amazonian forest	< 50–60	Palaeogene	Burnham & Johnson, 2004; Hoorn *et al*., 2010
GAARlandia land bridge	35–33	Eocene–Oligocene boundary	Iturralde-Vinent, 2006; Iturralde-Vinent & MacPhee, 1999
Formation of Dry Diagonal	33–28	Late Eocene–Early Oligocene (full formation in the Pleistocene)	Prado & Gibbs, 1993; Prado, 2000; Batalha-Filho *et al*., 2013; Sobral-Souza, Lima-Ribeiro & Solferini, 2015
Late Oligocene Warming Event (LOWE)	25–23	Late Oligocene	Zachos *et al*., 2001
Pebas system	23–10	Early–Middle Miocene	Wesselingh & Salo, 2006; Wesselingh, 2006
Main uplift of Northern and Central Andes	20–10	Miocene	Gregory-Wodcicki, 2000; Hoorn *et al*., 2010; Mulch *et al*., 2010; Garzione *et al*., 2008, 2014; Leier *et al*., 2013; Saylor & Horton, 2014
Mid-Miocene Climate Optimum (MMCO)	17–15	Middle Miocene	Zachos *et al*., 2001
Central American Landbridge (CAM)	15–12/3.0–4.0	Middle Miocene/Pliocene	Controversial: Coates *et al*., 1992; Montes *et al*., 2015; Bacon *et al*., 2015, 2016; O'Dea *et al*., 2016; Jaramillo, 2018
Andean forest	10	Late Miocene	Hoorn *et al*., 2010
Cerrado vegetation	10–4	Late Miocene	Simon *et al*., 2009; Pennington & Hughes, 2014
Acre system	7–5	Late Miocene–Pliocene	Hoorn *et al*., 2010
Caatinga woodlands	10–4.5/2.6	Late Miocene–Pliocene/ Pleistocene	Controversial: de Queiroz, 2006; Pennington & Ratter 2006; Collevatti *et al*., 2013; de Queiroz *et al*., 2017
Final uplift of Panama Isthmus	3.5–2.8	Mid-Pliocene	O'Dea *et al*., 2016
Pleistocene Arch forest corridor (PAFC)	2.6–10 000	Pleistocene	Prado & Gibbs, 1993

Sphaerogyni Caruzo included most of the species found in the Atlantic Forest of south-eastern Brazil; and *Croton* subsection *Spruceani* Caruzo included species mainly distributed in central and north-western South America, Central America and Mexico.

In this study, we build a new phylogenetic hypothesis of *Croton* section *Cleodora* as a mean to reconstruct its spatio-temporal evolution. Taxonomic sampling was expanded relative to Caruzo *et al*. (2011) to include all species currently recognized in the section, excepting *C. fernandezii* Riina & Caruzo, and we also increased the number of plastid markers. Our main goals were: (1) to infer a robuts and well-sampled phylogenetic tree of section *Cleodora* (26 out of the 27 species, 96% total diversity) based on a dataset of DNA sequences from the nuclear ITS and five plastid markers (*rps16*, *trnH-psbA*, *trnL-F*, *trnT-L* and *ycf1*); (2) to use molecular data to confirm the adscription of eight species recently assigned to the section based on morphological characters; (3) to reconstruct the spatio-temporal evolution of the section by estimating lineage divergence times, ancestral ranges and migration, speciation and geographical extinction events using Bayesian statistical inference and (4) to examine the hypothesis that the Amazon region is the ancestral geographical range of section *Cleodora*, and whether its remarkable geographical disjunction is linked to the formation of the South American Dry Diagonal, as has been suggested in other plant and animal lineages with similar disjunctions.

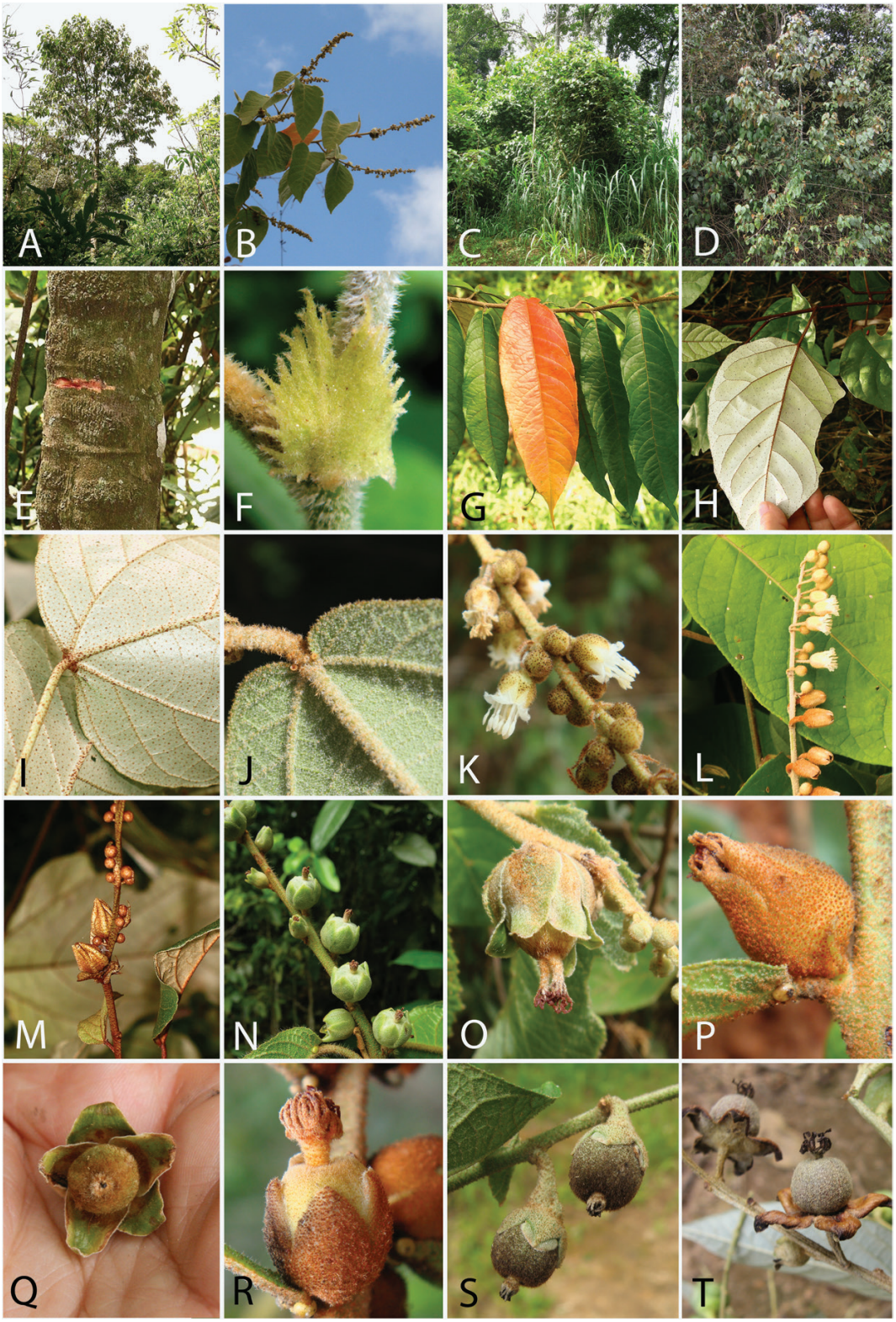


Figure 1. Members of *Croton* section *Cleodora* highlighting its main morphological characters. A, Arborescent habit of *Croton organensis*. B, Branches with inflorescences in *C. heterocalyx*. C, Habit and habitat of *C. rottlerifolius*. D, Arborescent habit of *C. nirguensis*. E, Wound on the trunk of *C. salutaris* showing its reddish latex. F, Foliose and lacerate stipule of

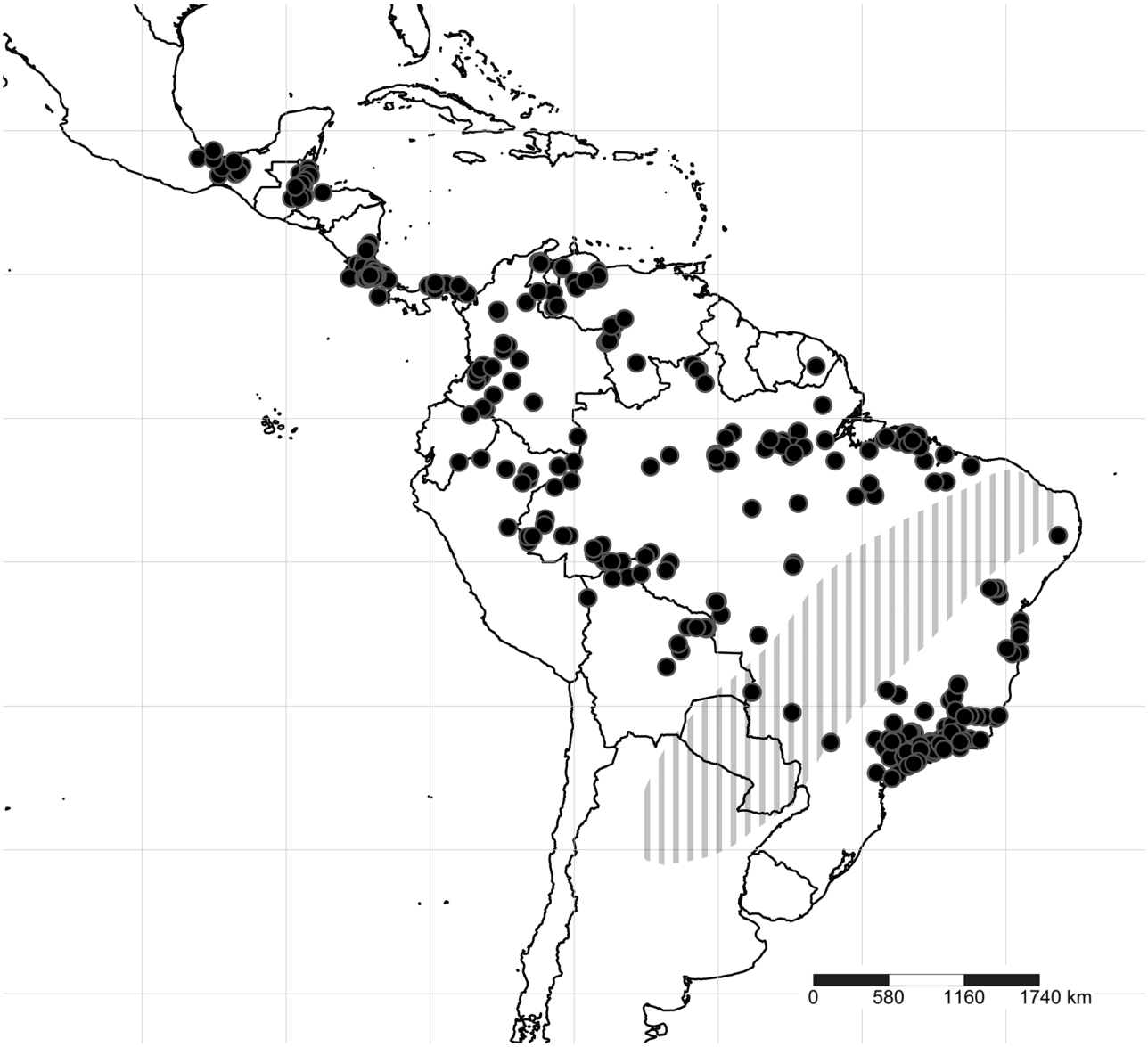


Figure 2. Map showing the known geographical distribution of *Croton* section *Cleodora* based on specimens examined in Caruzo & Cordeiro (2013) and Riina *et al*. (2018). Grey stripes represent the approximate area and location of the South American Dry Diagonal (see text).

C. fragrans. G, Leaves of *C. cajucara*, including a senescent leaf with the typical orange colour common in the genus. H, Silver leaves on the underside with a dense indument of lepidote trichomes in *C. campanulatus*. I, Petiolar-basilaminar glands with silver underside (lepidote trichomes), in *C. salutaris*. J, Petiolar glands and indument of stellate trichomes in *C. santaritensis*. K, Part of the inflorescence of *C. heterocalyx* showing the staminate flowers. L, Inflorescence of *C. spruceanus* showing pistillate flowers in the lower part and staminate flowers in the upper portion of the inflorescence. M, Lower portion of the inflorescence of *C. campanulatus* showing two pistillate flowers (larger size) and staminate flower buds (smaller and globose); indument is ferruginous-lepidote. N, Young fruits of *C. sphaerogynus*. O, Young fruit of *C. heterocalyx*. P, Pistillate flower of *C. stellatoferrugineus*. Q, Fruit of *C. heterocalyx*, without the styles, the quincuncial aestivation of the sepals is clearly visible. R, Fruit of *C. stellatoferrugineus* showing the typical stylar column characteristic of section *Cleodora*. S, Fruits of *C. rottlerifolius*, showing the persistent calyx. T, Fruits of *C. hemiargyreus* with the persistent calyx. Photographs: R. Riina and M.B.R. Caruzo; photograph D provided by W. Meier.

MATERIAL AND METHODS

Taxon sampling

Our sampling included 42 species (60 specimens): 26 species (44 specimens) from section *Cleodora sensu* Riina *et al.* (2018); 14 species (14 specimens) from other *Croton* clades and two species (two specimens) of *Brasiliocroton* P.E.Berry & Cordeiro, as the outgroup (Appendix 1). For the representation of other *Croton* clades, we used species of sections *Adenophylli* (Griseb.) Riina, B.W.van Ee & P.E.Berry, *Corylocroton* G.L.Webster, *Cuneati* (G.L.Webster) Riina & P.E. Berry, *Cupreati* Riina, *Moacroton* (Croizat) B.W.Van Ee & P.E.Berry, *Pachypodi* B.W.Van Ee, *Quadrilobi* Müll. Arg. and *Sampatik* (G.L.Webster) Riina, following the classification of Van Ee *et al.* (2011), and three species representing the Old World clade. In section *Cleodora*, we tried to sequence more than one specimen per species when possible to cover most of the species genetic variability and morphological plasticity and to assess the monophyly of species. Voucher information and GenBank accession numbers for the sequences used are provided in Appendix 1.

DNA extraction, amplification, sequencing and alignment

DNA extraction was done using the CTAB protocol (Doyle, 1991), using silica-gel dried leaf tissue or from herbarium samples. We amplified the multi-copy nuclear marker ITS and five regions of the plastidial genome: the non-coding spacer *trnH-psbA*, the intergenic spacers *trnL-F* and *trnT-L* and the coding regions *rps16* and *ycf1*. Primers and protocols used in the PCR amplification of each of the molecular markers are given in Table 2 and Appendix S2 of the Supporting Information, respectively. PCR products were purified using ExoSAP-IT PCR Product Cleanup Reagent (ThermoFisher) and sent for sequencing at MACROGEN (Macrogen Inc., Seoul, South Korea), using the same amplification primers (Table 2). Nucleotide sequences for each genetic region were assembled and edited in Geneious Prime v.2019.0.3 (Biomatters Ltd, Auckland, New Zealand) and automatically aligned with MAFFT online version (Katoh, Rozewicki & Yamada, 2017), using the default parameters. Manual adjustments were made to each alignment matrix in Geneious Prime v.2019.0.3, applying a similarity criterion (Simmons, 2004). Summary statistics for each alignment matrix (number of constant, variable and potentially phylogenetically informative characters according to the maximum parsimony criterion) were estimated in PAUP v.4.0b10 (Swofford, 2002) and included in Table 3.

Phylogenetic inference

Phylogenetic relationships were reconstructed using Bayesian inference implemented in MrBayes v.3.2.6 (Ronquist *et al.*, 2012) hosted on the CIPRES Science Gateway (Miller, Pfeiffer & Schwartz, 2010). Substitution models for analysis of each individual marker were selected using the Akaike information criterion implemented in jModelTest v.2.1.6

Table 2. Primers used in this study for the PCR amplification and sequencing of the five genetic regions selected for inferring phylogenetic relationships in *Croton* section *Cleodora* (ITS, *rps16*, *trnH-psbA*, *trnLF*, *trnT-L* and *ycf1*)

Primer name	Primer sequence (5′ to 3′)	Reference
ITS-I	GTCACTGAACCTTATCATTTAG	Urbatsch, Baldwin & Donoghue (2000)
ITS2	GCCRAGATATCCGTTGCCGAG	Cheng *et al.* (2016)
ITS3	YGACTCTCGGCAACGGATA	Cheng *et al.* (2016)
ITS4	CCGCTTAKTGATATGCTTAAA	Cheng *et al.* (2016)
trnH(GUH)	ACTGCCTTGATCCACTTGGC	Hamilton, 1999
psbAF	GTTATGCATGAACGTAATGCTC	Sang, Crawford & Stuessy (1997)
trnLF-c	CGAAATCGGTAGACGCTACG	Taberlet *et al.* (1991)
trnLF-d	GGGGATAGAGGGACTTGAAC	Taberlet *et al.* (1991)
trnLF-e	GGTTCAAGTCCCTCTATCCC	Taberlet *et al.* (1991)
trnLF-f	ATTTGAACTGGTGACACGAG	Taberlet *et al.* (1991)
rps16F	AAACGATGTGGTARAAAGCAAC	Shaw *et al.* (2007)
rps16R	AACATCWATTGCAASGATTCGATA	Shaw *et al.* (2007)
trnTLA2	CAAATGCGATGCTCTAACCT	Shaw *et al.* (2007)
trnTLB	TCTACCGATTTCGCCATATC	Shaw *et al.* (2007)
ycf1F	ATACATGTCGAAGTGATGGAAAA	Dong *et al.* (2015)
ycf1R	TCTCGACGAAAGTCCGATTGTTGTGAAT	Dong *et al.* (2015)

Table 3. Summary statistics for the aligned molecular data matrices used in our analyses, including sequences from GenBank and newly generated ones. **trnH-psbA* was only sequenced for *Cleodora* and *Cuneati* sections.

Dataset	NA	AL	VC	PIS (%)	MNS
ITS	60	672	270	194 (29)	m1
trnL-F	82	956	307	165 (17)	m2
trnTL	77	474	311	107 (23)	m2
rps16	49	872	280	131 (15)	m2
ycf1	44	1101	455	222 (20)	m2
*trnH-psbA**	42	401	68	50 (11)	m2
Plastid	54	3804	1010	313 (8)	m2
Plastid + ITS	60	4476	1280	507 (11)	m1 & m2

AL, aligned length; MNS, model of nucleotide substitution; m1, GTR + I + G; m2, GTR + G; NA, number of accessions; PIS, number of potentially parsimony informative sites; VC, number of variable characters.

(Guindon & Gascuel, 2003; Darriba *et al.*, 2012), also hosted on CIPRES; these were SYM+I+G for ITS, GTR+G for *rps16*, TPM1uf+G for *trnH-psbA*, TVM+G for *trnL-F*, TVM+G for *trnT-L* and GTR+G for *ycf1*. Because MrBayes does not implement the complexity of models evaluated in jModelTest, we used the nearest, slightly more complex model, which was GTR+I+G for ITS and GTR+G for each plastid marker. Bayesian inference is also known to be more robust to over-parametrization and more sensitive to infra-parametrization than the maximum likelihood optimization used in jModelTest (Ronquist & Deans, 2010). Two analyses of four chains each were run for 10×10^7 generations, sampling every 1000th generation. Convergence was assessed with the potential scale reduction factor and by monitoring the standard deviation of split frequencies (< 0.01). Tracer v.1.7.1 was used to monitor the mixing of Markov chains (effective sample size; Rambaut *et al.*, 2014). A 50% majority-rule consensus tree was constructed after discarding the first 25% of samples as burn-in.

To assess topological incongruence among markers, we ran Bayesian analyses on each individual matrix and in the combined plastid matrix. Topological incongruence was defined as the presence of well-supported clades, with posterior probability (PP) value ≥ 95% (Alfaro, Zoller & Lutzoni, 2003) that show conflicting relationships between trees. Since no cases of strong topological incongruence were detected among any of the individual plastid or combined plastid topologies versus the ITS topology (individual trees not shown), we built a concatenated matrix with six partitions (ITS, *rps16*, *trnH-psbA*, *trnL-F*, *trnT-L* and *ycf1*); this matrix was analysed allowing each marker its own substitution model and with the overall mutation rate allowed to differ among partitions (command *ratepr* = *variable* in MrBayes). Dating and biogeographic analyses were performed on the tree topology generated by the analysis of the concatenated nuclear-plastidial matrix combining all six markers, but including a single representative accession per species to avoid estimating artefactual speciation events, i.e. within-species divergence events (populations/individuals) modelled as speciation events in the analyses.

Divergence time estimation

Lineage divergence times were estimated using Bayesian relaxed clock models implemented in BEAST v.1.8.4 (Drummond *et al.*, 2012) hosted on CIPRES (Miller *et al.*, 2010). The matrix used was the concatenated matrix (six regions), allowing different molecular evolutionary models for each molecular marker. As the tree growth prior, we used the birth-death model with incomplete sampling (Stadler, 2009) because we have a nearly complete sampling of the species of section *Cleodora*. As the molecular clock model, we used the uncorrelated lognormal relaxed molecular clock (UCLN), which allows variations in the rate of molecular evolution between the branches of the tree according to a lognormal probabilistic distribution (Drummond *et al.*, 2012). We ran two MCMC chains for 10×10^7 generations, sampling every 1000th generation, and used TRACER v.1.7.1 to monitor convergence and adequate mixing (ESS > 200; Rambaut *et al.*, 2014). A maximum clade credibility (MCC) tree (burn-in 25%) was constructed in TreeAnnotator v.1.8.4 (Drummond & Rambaut, 2016).

To calibrate relative lineage divergence times, we used three secondary calibration points based on the most recent dated phylogenetic tree for *Croton* (Arévalo *et al.*, 2017). As calibration prior distribution, we used a normal or Gaussian distribution, with mean and standard deviation covering the mean and the confidence interval, the 95% highest posterior density credibility interval (HPD), of the posterior estimates by Arévalo *et al.* (2017). The root node, the divergence between *Brasiliocroton* and *Croton* was assigned a normal distribution with mean (M) = 57.2 Mya and

standard deviation (SD) = 1.1, 95% HPD = 55.39–59.01 Mya. The second calibration point, the divergence of the clade comprising *Croton* sections *Pachypodi*, *Sampatik*, *Quadrilobi*, *Corylocroton* and *Moacroton*, was assigned a normal distribution with M = 42 Myr, SD = 4, 95% HPD = 35.42–48.58 Myr. Finally, the third calibration point, the divergence of section *Adenophylli* from the rest of *Croton*, was assigned a prior distribution with M = 38 Mya, SD = 4, 95% HPD = 31.42–44.58 Mya.

Biogeographical analyses

We inferred ancestral geographical ranges and the history of migration, speciation and local extinction events using dispersal extinction cladogenesis (DEC, Ree & Smith, 2008). This parametric method allows the inference of anagenetic events, i.e rates of range expansion (dispersal) and range contraction (local extinction), along time-calibrated branches, and the likelihood range inheritance scenarios at speciation nodes describing the split of the ancestral range between the two descendants. The possibility to include cladogenetic events in the inference of both widespread and narrow-distributed ancestors makes DEC particularly appropriate when dealing with continental biogeographic scenarios, where areas are adjacent to one another and lineages are likely to expand their ranges and later speciate by range division (Sanmartín, 2020). The original DEC implementation (Ree & Smith, 2008) used a maximum likelihood framework for biogeographic inference. We used here instead the Bayesian implementation of DEC in the open software RevBayes (Höhna *et al.*, 2016) by Landis, Freyman & Baldwin (2018). A Bayesian approach to DEC allows us to account for uncertainty in the inferred model parameters, the rates of dispersal and local extinction and the marginal posterior probabilities of ancestral ranges. Moreover, the Bayesian framework permits full statistical comparison of the marginal odds for two competing hypothesis using Bayesian factors.

As the phylogenetic tree for the DEC analysis, we used the MCC tree from BEAST, which has branch lengths in units of time since divergence (Myr) and also root-to-tip path lengths equal, unlike the MrBayes phylogram. The function *drop.tip* from the package *ape* (Paradis *et al.*, 2019) implemented in R (R Development Core Team, 2017) was used to prune this tree to leave only one specimen per species, except in the case of *C. sphaerogynus* Baill., which was recovered as non-monophyletic (see Results). We also reduced the taxon sampling outside section *Cleodora* to its closest clades, sections *Cuneati* and *Cupreati*, since the aim of the analysis was the spatio-temporal evolution of section *Cleodora*, not *Croton*. Moreover, most *Croton* sections are poorly represented in terms of taxon sampling in our analysis, and typically occur in several Neotropical areas. Sections *Cuneati* and *Cupreati*, however, are narrowly distributed: monotypic *Cupreati* is Andean (montane forest), and *Cuneati* (11 species) is mostly Amazonian, except for two species (Riina *et al.*, 2010). Sister to these groups is the large Old World clade (*c.* 450 species) distributed in tropical Africa, Asia and Australia (Berry *et al.*, 2005; Van Ee *et al.*, 2011).

Area definition was based on current distribution patterns of species in section *Cleodora*. We used the available database of georeferenced locations for all species in this section from our previous taxonomic studies (Caruzo & Cordeiro, 2013; Riina *et al.*, 2018). A map representing all individual occurrences was constructed using the online tool SimpleMappr (Shorthouse, 2010) (Fig. 2). Operational areas for the biogeographic analysis were also defined considering criteria such as integrating the geological history of South America (Hoorn *et al.*, 2010; Antonelli & Sanmartín, 2011), and maximizing congruence with previous plant biogeographic analysis in this region (Perret *et al.*, 2013; Chazot *et al.*, 2019; Lavor *et al.*, 2019; Thode *et al.*, 2019). The four delimited areas were: (A) Atlantic Forest; (B) Amazonia; (C) Andes and (D) Central America and Mexico.

The RevBayes DEC analysis used the default settings in the tutorials (Landis *et al.*, 2018), except that we did not scale the rate of dispersal by geographical distance as in Lavor *et al.* (2019); we were not interested in the influence of geographical distance on dispersal rates and considered that long- and medium-/short-distance dispersal should be initially equally probable. We also modified the priors of the baseline migration and extinction rates, parameters 'rate-bg' and 'extinction_rate', to adopt more informative distributions based on the posterior estimates from a preliminary run (see next). To reduce the complexity in our analysis, the size of widespread ancestral ranges was limited to a maximum of three areas using the function 'max_areas', given also that none of the current species in section *Cleodora* is distributed in more than three areas. Cladogenetic range evolution was modelled as resulting from two different types of range inheritance scenarios, sympatry for single areas ('s') and allopatry ('a'), the latter includes in RevBayes DEC vicariance and peripheral isolate speciation. Founder speciation events, represented by the parameter 'j', were no included because of recent criticisms to the DEC+J model (Matzke, 2016), as ignoring the contribution of branch length information; this often results in negligible extinction rates (close to zero) and lower dispersal rates relative to DEC. At the extreme, DEC+J inferences can be explained (almost) entirely by the cladogenetic events (Ree & Sanmartín, 2018).

Analyses were run for 10 000 generations (in RevBayes, movement proposals affect all parameter

simultaneously unlike in MrBayes, so one generation in RB is equivalent to *c*. 1000 in MrBayes). We ran two different analyses. One analysis (M0) was run with the parameter for the baseline dispersal rate ('rate_bg') modelled as constant over time, and cladogenetic events modelled as a simplex, with equal probability for 's' and 'a' range inheritance scenarios. A more complex analysis (M1) was run, in which cladogenetic events were modelled through a Dirichlet distribution, with probability of allopatry equal to 1 minus the probability of simpatry. We also increased the complexity of the M1 model relative to M0 by using a stratified dispersal model. We divided the phylogenetic tree into time slices and scaled the baseline dispersal rate by some discrete values according to a hypothesis of geological connectivity among areas through time. This stratified model allowed us to examine a similar hypothesis to Thode *et al*. (2019). These authors used two different dispersal-scaled rate matrices in which the late Mid-Miocene (13 Mya) acted as the bound/temporal limit between the two intervals. Although major geological and global climatic changes occurred during the entire Cenozoic (Table 1), the Mid–Late Miocene is often considered a turning point for the Neotropics. For example, the formation of the Central American Landbridge, dated *c*. 15–12 Mya by some authors (Table 1), is suggested as an important biotic migration route between North America and South America for many animals and plants (Bacon *et al*., 2015). Although full completion of the Panama Isthmus allowing land dispersal probably occurred later (O'Dea *et al*., 2016; Table 1), for seed plant lineages this was less an impediment (Bacon *et al*., 2013). Similarly, the Late Miocene marks the closing of the Pebas System (10 Mya) and its successor, the Acre system (7–5 Mya; Table 1), which allowed biotic exchange between the Andes and the eastern Amazonian region (Antonelli *et al*., 2009; Hoorn *et al*., 2010). As mentioned above, the Dry Diagonal began to form during the Eocene/Oligocene transition, but increasingly drier and cooler climates after the MMCO (17–15 Mya, Table 1) led to the replacement of the extensive woodlands that characterized the Paleogene by more xeric vegetation. The new more xeric biomes of the Dry Diagonal probably acted as a dispersal barrier between the Amazonia and Atlantic Forest regions for many tropical plant lineages (Prado & Gibbs, 1993; Peres *et al*., 2020).

Rather than using a fixed age for the boundary between the two time slices (Thode *et al*., 2019), we followed the approach of Landis *et al*. (2018). We defined a uniform prior probability distribution bounded by 18 and 7 Mya, which spans the dates of the aforementioned events during the Mid–Late Miocene. During the analysis, the MCMC sampled the breakpoint value, when the model switches between the two dispersal-scaled rate matrices, within this time interval or uniform probability distribution. This allows us to reflect the uncertainty in the biogeographic connectivity model, i.e. we do not know which of these event/s were the most relevant for the evolution of section *Cleodora*, and the geological ages for the events themselves are also uncertain.

The two models M0 and M1 and their fit to the data were evaluated using Bayes Factor Comparison, using path sampling and stepping-stone power posteriors to estimate the marginal likelihood of each of the models (Blanquart & Lartillot, 2006; Baele *et al*., 2012). In a preliminary run, the power posterior analysis failed to complete due to conflict with sampled prior values too distant from the empirical data (likelihood of roughly 0). To solve this, we used more informative, biologically realistic priors for some parameters, using the posterior estimates from the analysis of the empirical data. Thus, in the final analyses, rate_bg was set to a loguniform prior bounded by mininum = 0.0001 and maximum = 0.1 events/Myr, compared to 100 in the RevBayes tutorial; the 'extirpation_rate' was set to a lognormal prior with logmean = 0.0 and $\log_{SD}$ = 0.5, giving a 95% HPD range = 0.29–3.43 events/Myr, compared to default 0.38–2.01. The final M0 and M1 analyses comprised 100 steps, with 1000 generations/each and a burn-in of 10 000 generations in the first step (to obtain the empirical posterior). The script to carry out this analysis is provided in the Supporting Information, Appendix S3.

RESULTS

DNA SEQUENCING AND ALIGNMENT

Summary statistics for each individual marker and for the combined plastid and concatenate nuclear-matrix datasets are presented in Table 3. In total, our data matrix comprised 238 sequences: 198 were newly generated for this study, and 40 were downloaded from GenBank (Appendix 1); 122 PCR amplification attempts were unsuccessful (missing data). The number of accessions (and the percentage of amplification success) varied among markers: *trnL-F* was the most successful and *trnH-psbA*, the least. Data alignment was relatively straightforward, with only a few manual adjustments after automatic alignment. The marker with the largest number of variable sites was *ycf1*, and the marker containing the greatest number of potentially parsimony informative sites was ITS (29%), followed by *trnT-L* (23%) and *ycf1* (20%) (Table 3). The *trnL-F* region has been the plastid region most often used in systematic studies of *Croton*, but here it shows a lower percentage of informative sites (17%)

compared to plastid regions *trnT-L* and *ycf1* (never used previously for *Croton*). Of the three matrices analysed (ITS, plastid combined and plastid + ITS), the plastid combined matrix provided the lowest degree of phylogenetic information (8%) (Table 3).

Phylogenetic analyses

Concerning relationships among subsections in *Croton,* the ITS phylogenetic analysis (Supporting Information, Appendix S4) shows low to moderate resolution for several of the basal nodes. In contrast, the monophyly of section *Cleodora* (PP = 1) and its division into two subclades, corresponding to subsections *Spaherogyni* (PP = 1) and *Spruceani* (PP = 0.91), were strongly supported. Relationships in subsection *Sphaerogyni* show better resolution and clade support than those in subsection *Spruceani*; in the latter, most species appear in a large polytomy, with *C. amazonicus* Müll. Arg. as the sister group (PP = 0.91).

The phylogenetic tree generated by the combined plastid matrix (Supporting Information, Appendix S5) supports the monophyly of section *Cleodora* (PP = 1) and, unlike the ITS tree, many of the basal nodes receive strong support (PP = 1). In *Cleodora*, two well-supported subclades/subsections are recovered (PP = 1) (Supporting Information, Appendix S5), but in contrast to ITS, subsection *Sphaerogyni* is not well resolved. In subsection *Sphaerogyni,* the three accessions of *C. sphaerogynus* did not form a clade; similarly, *C. hoffmannii* is recovered as paraphyletic, although support is low (Supporting Information, Appendix S5). In subsection *Spruceani*, *C. fragrans* Kunth appears as sister to the rest of species in the subsection, and its position receives stronger support (PP = 1, Appendix S4) than *C. amazonicus* in the ITS phylogenetic analysis (PP = 0.91, Appendix S5).

Phylogenetic analysis of the concatenate nuclear-plastid matrix generated the best-resolved tree, with strong clade support for all nodes (PP = 1), including section *Cleodora* and its main subclades (Fig. 3). Therefore, this tree was used for the subsequent phylogenetic dating and biogeographical analyses, and it forms the basis for the discussion of phylogenetic relationships that follow.

Among the other clades of *Croton*, section *Cuneati* was recovered as the sister clade of section *Cleodora* (PP = 1). In section *Cleodora*, subsection *Spruceani* is strongly supported (PP = 1); *C. fragrans* is recovered as the sister group of the other species with strong support (PP = 1), but in general species relationships are poorly resolved [with the exception of *C. orinocensis* Müll.Arg.-*C. perstipulatus* G.L.Webster ex Caruzo & Secco (PP = 1; Fig. 3)]. Subsection *Sphaerogyni* (PP = 1) shows better resolution, with two main subclades: one formed by *C. hemiargyreus* Müll.Arg., *C. salutaris* Casar. and *C. campanulatus* Caruzo & Cordeiro (PP = 1) and one comprising the remaining species. In the latter subclade, *C. gigantifolius* P.E.Berry & Secco forms the sister group (PP = 1) of a large subclade, with generally poor resolution and clade support for species relationships (Fig. 3). As in the ITS and combined plastid tree, most conspecific accessions are grouped into clades, supporting the monophyly of species (including *C. hoffmannii* Müll. Arg.). However, a few species were recovered as unresolved (*C. spruceanus* Benth.) or paraphyletic; the accesions of *C. sphaerogynus* are grouped within a clade with those of *C. stellatoferrugineus* Caruzo & Cordeiro (Fig. 3), with strong support (PP = 0.97). This relationship agrees with the pattern shown by the ITS tree (Supporting Information, Appendix S4).

The seven species sampled from the eight new additions to the section by Riina *et al*. (2018), which are sequenced here for the first time, were recovered in our phylogenetic analyses as part of section *Cleodora* with all markers (indicated with asterisk in Fig. 3, Supporting Information, Appendices S4, S5). Six of the species are included in subsection *Spruceani* (*C. amazonicus*, *C. nirguensis* Riina & Meier, *C. santaritensis* Huft, *C. javarisensis* Secco, *C. perstipulatus*, *C. loretensis* Riina & Caruzo), with only *C. viroleoides* P.E.Berry & Secco and *C. gigantifolius* recovered as part of subsection *Sphaerogyni* (Fig. 3, Supporting Information, Appendices S4, S5). Only *C. fernandezii*, a poorly known Amazonian species with a few herbarium records, remains unsampled for section *Cleodora*.

Divergence time estimates

Species relationships in the MCC chronogram (Fig. 4) were similar to the MrBayes non-clock trees (Fig. 3, Supporting Information, Appendices S4, S5) and exhibited similarly strong clade support values (PP > 0.95) for the basal nodes. The divergence of section *Cleodora* from sister-section *Cuneati* is dated in the Late Oligocene (24.52 Mya, 95% HPD = 18.80–30.60 Mya), with strong support (PP = 0.99); the split between these two clades and section *Cupreati* is dated in the Middle Oligocene (27.13 Mya, 95% HPD = 21.05–33.31 Mya; PP = 0.99). As in the non-clock trees, the Old World clade of *Croton* was recovered as sister of these three sections, with divergence estimated around the Early Oligocene (31.4 Mya, 95% HPD = 25.33–37.61 Mya, PP = 0.99; Fig. 4). The most recent common ancestor (MRCA) of the extant species of section *Cleodora*, the divergence between subsections *Sphaerogyni* and *Spruceani*, is dated in the Early Miocene (19.7 Mya, 95% HPD 14.54–25.40 Mya; PP = 0.99; Fig. 4) and the age of crown node of these two clades (MRCAs of *Sphaerogyni* and

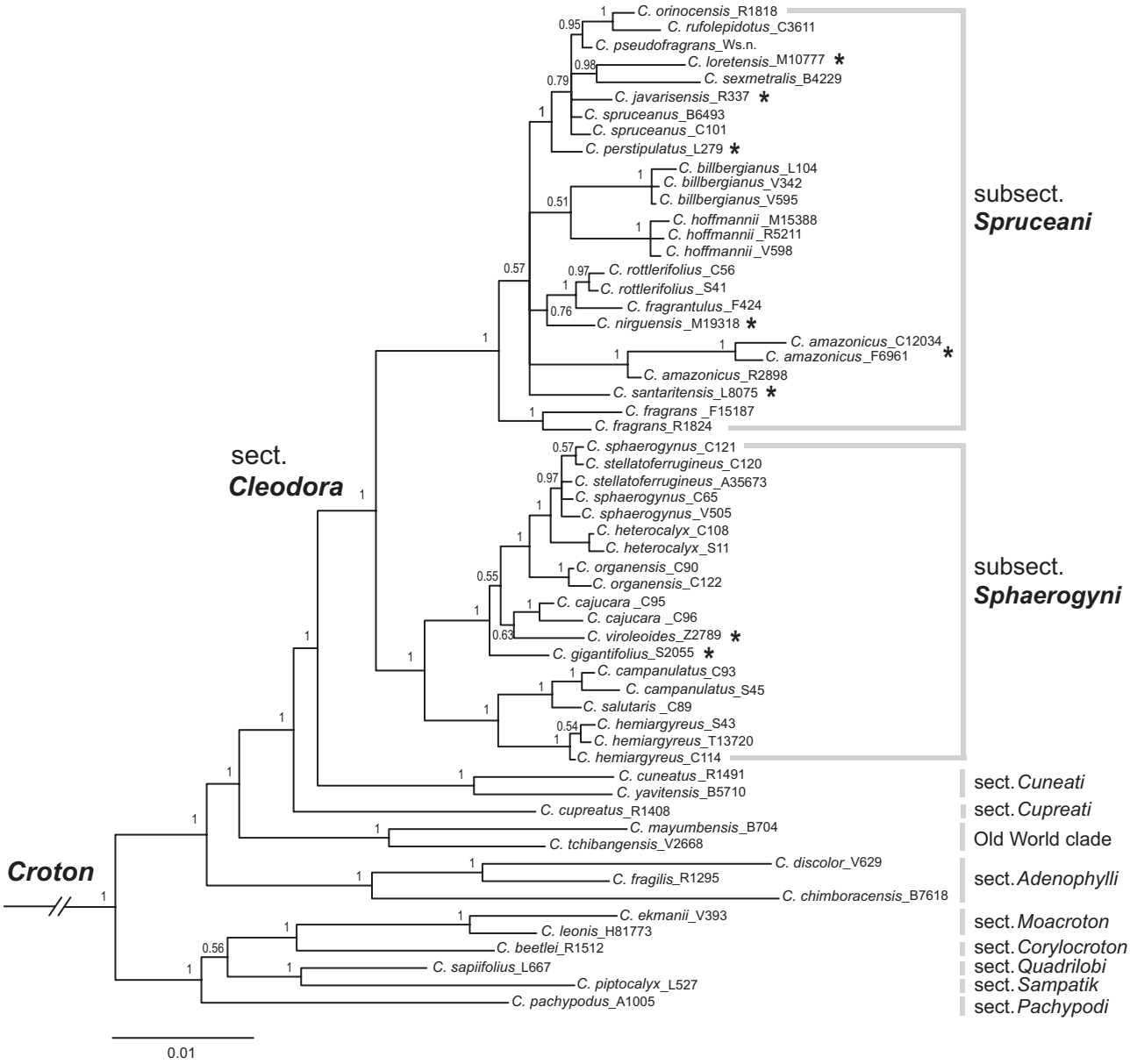


Figure 3. Majority-rule consensus tree of the MrBayes analysis of combined ITS, *rps16*, *trnH-psbA*, *trnL-F*, *trnT-L* and *ycf1* data of *Croton* section *Cleodora* and related taxa. The numbers represent support values with Bayesian posterior probability (PP). Species sampled for the first time in a phylogenetic analysis, including recently described species, are indicated with an asterisk. The collector (initial of last name) and the collection number are included in the name of the taxa. The outgroup (*Brasiliocroton*) was removed from the figure.

Spruceani) dates to the Middle Miocene, 15 Mya (95% HPD 9.40–19.61 Mya; PP = 0.99) and 12 Mya (95% HPD 8.40–16.31 Mya; PP = 1), respectively (Fig. 4).

Biogeographic reconstruction

Comparison of marginal likelihood values of the two biogeographic models with Bayes Factors gave ss and ps values of –47.179 and 47.178, and –51.942 and 51.942, for M1 and M0, respectively; BF = 2 (M1–M0) = 9.52, indicating strong evidence (> 6; Kass & Raftery, 1995) for the stratified M1 model. Marginal PP values for ancestral areas in this reconstruction (Fig. 5) was moderate for the most basal nodes and higher for the most recent speciation events.

According to this model, the split of sections *Cleodora–Cuneati* from section *Cupreati* (node 1, Fig. 5) is inferred as a vicariance event between the Amazon (area B) and

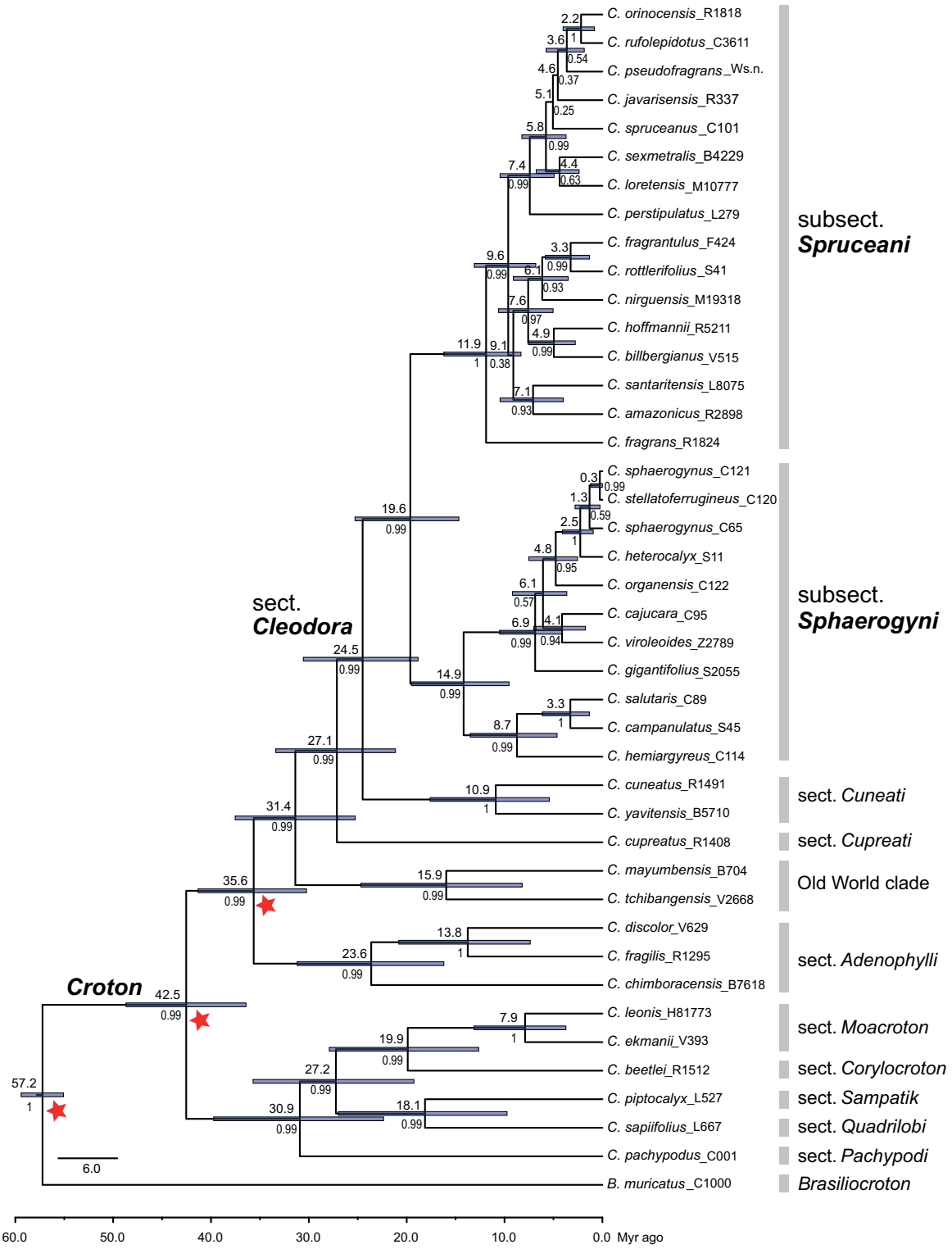


Figure 4. Maximum clade credibility (MCC) tree for *Croton* section *Cleodora* and related taxa, showing phylogenetic relationships and lineage divergence times, inferred by BEAST v.1.8.4. The red stars indicate the position of three secondary calibrations points. Numbers above nodes indicate mean age estimates (Myr), those below nodes show posterior probability values (PP). Blue bars at nodes indicate the associated 95% HPD credibility intervals for those nodes with PP > 0.50.

the Andes region (area C) within a widespread ancestor (marginal PP = 0.62). The ancestor of *Cleodora–Cuneati* (node 2, PP = 0.85) was reconstructed as restricted to the Amazon area; similarly, the MRCAs of sections *Cuneati* and *Cleodora* were inferred as Amazonian (nodes 3 and 4, 0.93 and 0.68, respectively, Fig. 5). The ancestor of subsection *Spruceani* is also reconstructed as distributed in the Amazonian region, but with much stronger support (node 5, 0.91, Fig. 5). The ancestor of subsection *Sphaerogyni* was inferred as widespread over the Amazonian and Atlantic Forest regions (node 6, Fig. 5), implying a dispersal event from the Amazon Basin to the Atlantic Forest during the 5 Myr subtending branch (20–15 Mya), although again support is not strong (0.65). The first cladogenetic event in *Sphaerogyni* implied a vicariance event between the Amazonia (B) and the Atlantic Forest (A), giving rise to two subclades, one of which diversified in the Atlantic Forest (node 7), whereas the second subclade extended its range from Amazonia into the Atlantic Forest (node 8, 0.86), giving rise to two endemic clades, *C. organensis* Baill.–*C. stellatoferrugineus* and *C. viroleoides–C. cajucara* through a second vicariance event (node 9, Fig. 5). In subsection *Spruceani* (node 5), most speciation events seem to have involved area B, the Amazonia, but at least in one of the subclades, several migration (range expansion) events led to subsequent diversification. Dispersal events from the Amazonia (B) to the Central American and Mexican region (D, node 10) and from this region to the Andes (C; node 11) are inferred, with some uncertainty in the sequence of events (Fig. 5). A dispersal event from the Andes (C) to the Atlantic Forest (A) is reconstructed in the clade *C. rottlerifolius* Baill.–*C. fragrantulus* Croizat (node 12). Migration from the Amazonia to the Andes explains the distribution of *C. sexmetralis* in a mostly Amazonian subclade (node 13; Fig. 5).

The DEC reconstruction with the non-scaled model M0 (Supporting Information, Appendix S6) was similar to that obtained with M1 (Fig. 5), except for slightly lower values of marginal posterior probabilities for the ancestral areas at the basalmost nodes, e.g. nodes 2, 4, 5 and 6. The largest difference is in the MRCA of section *Spruceani* (node 5), which is reconstructed as Amazonian, but with much weaker clade support (0.62, Supporting Information, Appendix S6) compared to M1 (0.91, Fig. 5).

DISCUSSION

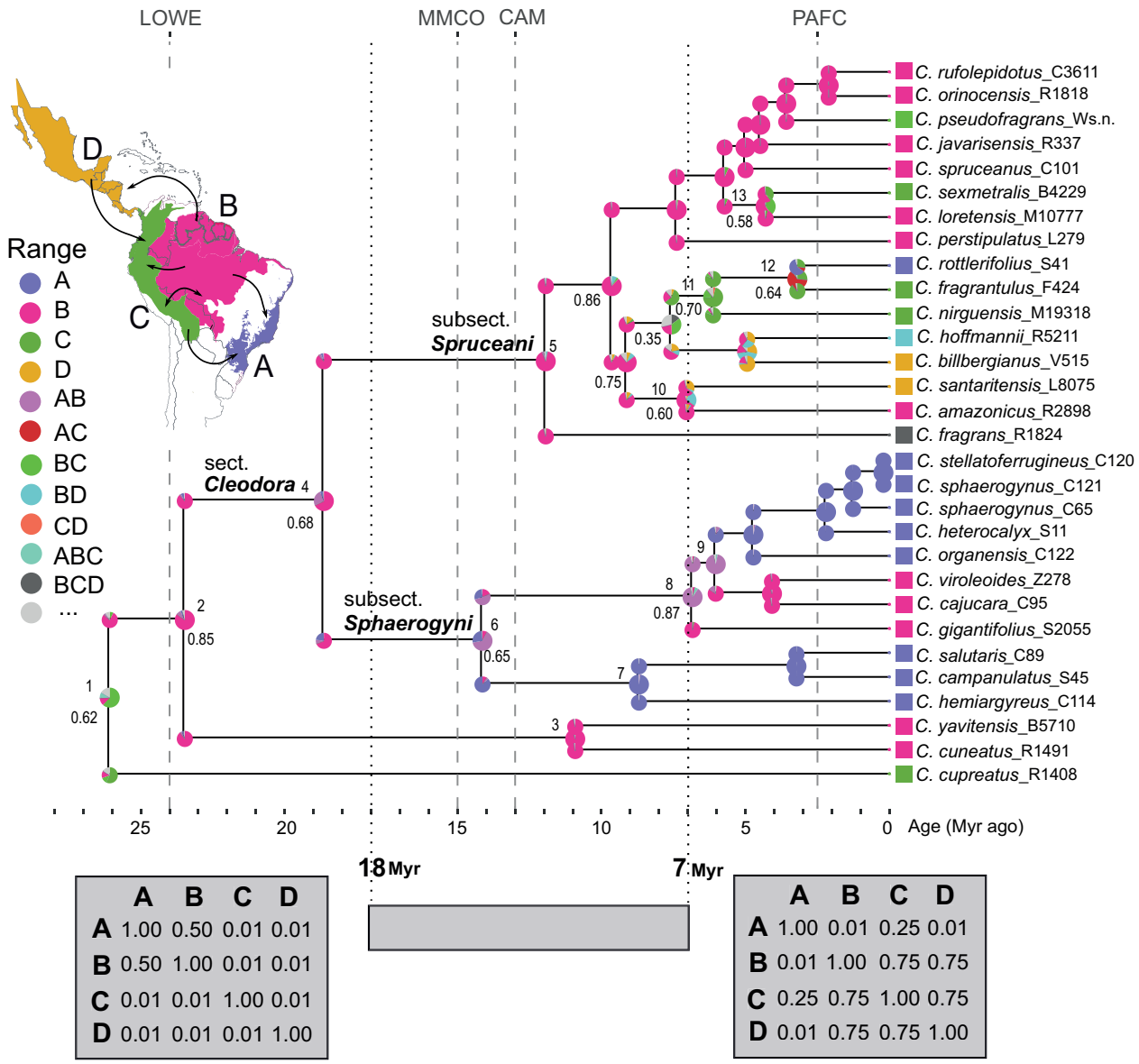
A NEW MOLECULAR PHYLOGENETIC TREE FOR *CROTON* SECTION *CLEODORA*

The plastid markers *rps16*, *trnT-L* and *ycf1*, used here for the first time in *Croton*, proved phylogenetically informative, especially if used in combination with ITS and the other two plastid regions (*trnH-psbA*, *trnL-F*) previously used for the genus (Table 2; Fig. 3, Supporting Information, Appendices S4, S5). However, none of the six markers, alone or in combination, allowed us to completely solve the most distal nodes in the phylogenetic tree, especially those in subsection *Spruceani*. A multigene approach, e.g. using genomic target sequencing (Villaverde *et al*., 2018), might be needed to solve or better understand the most recent speciation events in section *Cleodora*.

Our nearly complete phylogenetic reconstruction of section *Cleodora* agrees well with previous studies (Caruzo *et al*., 2011; Van Ee *et al*., 2011), supporting the monophyly of the section (Fig. 3). The topology of the plastid phylogenetic analysis is similar to those obtained by Riina *et al*. (2010) and Caruzo *et al*. (2011), showing section *Cuneati* as sister group of section *Cleodora*. These two sections share the arborescent habit, the presence of stellate to lepidote trichomes and petiolar nectary glands, and they occupy similar forest habitats in the Neotropics (Riina *et al*., 2010; Caruzo *et al*., 2011). The ITS phylogenetic tree, however, differs from that of Caruzo *et al*. (2011), in which section *Cleodora* was placed in a polytomy in *Croton*. This is probably because, unlike Caruzo *et al*. (2011), our sampling of *Croton* did not include several sections of possible hybrid origin, e.g. section *Cyclostigma* Griseb. (see Van Ee *et al*., 2008; Riina *et al*., 2009).

Both plastid and ITS topologies are congruent with previous studies (Caruzo *et al*., 2011; Van Ee *et al*., 2011) supporting the division of section *Cleodora* into subsections *Sphaerogyni* and *Spruceani*, but with stronger clade support and a nearly complete taxon sampling; the only omission being *C. fernandezii* (Supporting Information, Appendices S4, S5). The main morphological characters used to separate the subsections are: calyx aestivation of the pistillate flower, which is quincuncial (Fig. 1M–Q) in subsection *Sphaerogyni* versus imbricate (Fig. 1L, S) in subsection *Spruceani*; and fruit shape, which is globose or ellipsoid in subsection *Spruceani* (Fig. 1S) and subglobose or often trigonal in subsection *Sphaerogyni* (Fig. 1T) (Caruzo *et al*., 2011; Riina *et al*., 2018). *Croton fernandezii* was classified in subsection *Sphaerogyni* by Riina *et al*. (2018) because of its quincuncial calyx aestivation, but this remains to be confirmed by molecular data.

The sectional assignment of new species to section *Cleodora* hypothesized by Riina *et al*. (2018) based solely on morphology was supported by our results. Regarding the subsectional placement, only two species were not correctly classified by Riina *et al*. (2018). They assigned *C. amazonicus* and *C. perstipulatus* to subsection *Sphaerogyni*, whereas our phylogenetic analysis recovers these species in



	A	B	C	D
A	1.00	0.50	0.01	0.01
B	0.50	1.00	0.01	0.01
C	0.01	0.01	1.00	0.01
D	0.01	0.01	0.01	1.00

	A	B	C	D
A	1.00	0.01	0.25	0.01
B	0.01	1.00	0.75	0.75
C	0.25	0.75	1.00	0.75
D	0.01	0.75	0.75	1.00

Figure 5. Model M1 reconstruction of the spatio-temporal evolution of *Croton* section *Cleodora* and closest allies, using Bayesian dispersal extinction cladogenesis (DEC). The phylogenetic tree is the MCC tree from BEAST shown in Figure 4. A pie chart on the node indicates alternative ancestral geographical ranges, with pie slices indicating their marginal posterior probabilities (MPP). Values below pie charts indicate the MPP for the ancestral range receiving the highest probability; only values < 0.90 are shown. Numbers above pie charts indicate the nodes mentioned in the text. The smaller pie charts on the branches indicate the geographical range immediately after the speciation event; a pie chart on the node indicates how this range has changed as a result of range expansion and contraction events along the branch. Colours represent the geographical range state according to the legend. Letters on the map indicate the operational areas according to the legend. The two time slices with dispersal-scaler values are included in the two tables below the tree. The two dotted vertical lines and the grey rectangle indicate the time bounds of the interval within which the switch between the two time slices is modelled to occur, spanning the Mid–Late Miocene (i.e. 18–7 Mya). The dashed vertical lines (with labels at the top) represent geological or global climatic events that may have played a role in the spatio-temporal evolution of section *Cleodora*. Arrows on the map indicate the main events of dispersal and vicariance between areas according to the ancestral reconstructions with higher marginal probability. LOWE: Late Oligocene Warming Event, MMCO: Mid-Miocene climatic optimum, CAM: formation of the Central American Landbridge, PAFC: Pleistocene Arch forest corridor.

subsection *Spruceani*. A key character to separate the two subsections is floral aestivation, but this could not be properly assessed for *C. amazonicus* and *C. perstipulatus* due to poor or incomplete herbarium specimens (Riina *et al.*, 2018).

Our phylogenetic reconstruction of section *Cleodora* represents the most complete and resolved of any major Neotropical clades of *Croton*, for which only a partial taxon sampling have been included in previous phylogenetic studies. The three additional plastid markers used here (*rps16*, *trnT-L* and *ycf1*) are likely to be useful in future phylogenetic analysis of other diverse and taxonomically challenging *Croton* clades (e.g. *Adenophylli*, *Geiseleria*, *Barhamia*), in need of a phylogenetic framework for evolutionary research and species delimitation.

The grouping of the accessions of *Croton stellatoferrugineus* and *C. sphaerogynus* into a single clade in the combined analysis (Fig. 3) and the ITS tree (Supporting Information, Appendix S4) agrees with their close morphological affinity (Caruzo *et al.*, 2010) and their recent age of divergence (*c.* 1 Mya, Fig. 4). However, despite their similarities, the two species are easily told apart, using morphology and the striking differences in habitat and distribution (Supporting Information, Appendix S1; Caruzo *et al.*, 2010). *Croton stellatoferrugineus* occupies seasonally dry forests in the interior of Minas Gerais, whereas *C. sphaerogynus* occurs in moist forests in seashore plains ('restinga' forests) in south-eastern Brazil. On the other hand, some accessions of *C. sphaerogynus* appear grouped with *C. organensis* and *C. heterocalyx* Baill. in the plastid tree with strong clade support (0.95, Supporting Information, Appendix S5). Further population-level sampling of *C. sphaerogynus* and related species and additional genetic data may help solve their taxonomic delimitation.

Spatio-temporal evolution of *Croton* section *Cleodora* in the Neotropics

The Atlantic Forest and the Amazonia are tropical forest biomes that are home to an enormous biodiversity and are currently threatened by human action and climate change (Ribeiro *et al.*, 2009; Cardoso *et al.*, 2017; Antonelli *et al.*, 2018; Peres *et al.*, 2020). The biogeographic disjunction between these two forest biomes, attributed to the formation of the Dry Diagonal as an ecological barrier, have been identified in different organisms including frogs (Fouquet *et al.*, 2012; Castroviejo-Fisher *et al.*, 2014; Padial *et al.*, 2014; Pirani *et al.*, 2020), lizards (Rodrigues *et al.*, 2014; Prates *et al.*, 2017) and plants (Freitas *et al.*, 2016; Thode *et al.*, 2019).

The Dry Diagonal started to form during the Palaeogene, driven by global climate cooling at the Late Eocene–Early Oligocene transition (33–28 Mya; LEEOC; Zachos *et al.*, 2001). Later warming events, such as the LOWE and the MMCO (Table 1), contributed to further divide an initially continuous forest mass into eastern and western sections, isolated by dry-adapted, open land vegetation (Bigarella, Andrade-Lima & Riehs, 1975; Orme, 2007; Fouquet *et al.*, 2012; Batalha-Filho *et al.*, 2013; Sobral-Souza *et al.*, 2015; Peres *et al.*, 2020). Although the Dry Diagonal probably acted as a climatic/geographical barrier for regional biotic interchanges between Amazonia and Atlantic biomes since its formation, fossil record and past climatic reconstructions suggest that several forest corridors allowed migration events at different times during the Neogene (i.e. Miocene, Pliocene and Pleistocene), and involving different parts of the Dry Diagonal (i.e. north, central and south; Fine & Lohmann, 2018; Peres *et al.*, 2020). According to Fine & Lohmann (2018), more stable and longer (*c.* 50 Mya) forest corridors would have connected western Amazonia and the south-western Atlantic Forest in the south, whereas less stable and shorter ones (5–10 Mya) connected these biomes in the north.

The Dry Diagonal reached its climax with the formation of the Caatinga and the Cerrado biomes in the Late Miocene–Pliocene (Table 1). The Cerrado was mainly assembled during the Late Miocene (*c.* 10 Mya, Simon *et al.*, 2009), following the expansion of colder and drier climates worldwide around this period (Meseguer *et al.*, 2015), although some lineage divergences are dated in the Pliocene (Pennington & Hughes, 2014). The age of the Caatinga has long been debated. Some authors consider its origins to be Quaternary (e.g. de Queiroz, 2006), whereas others suggest a much older beginning for this region (e.g. Pennington, Prado & Pendry, 2000; Pennington *et al.*, 2004; Pennington & Ratter, 2006; Collevatti *et al.*, 2013). A recent review of molecular phylogenetic evidence (de Queiroz *et al.*, 2017) supports a heterochronous origin, with the SDTF woody lineages of the 'crystalline Caatinga' dating from the Late Miocene to the Pliocene, whereas endemic species from the sandy and karstic areas in the 'sedimentary Caatinga' are mostly Pleistocene in age.

The biogeographic history of *Croton* section *Cleodora* reconstructed here (Fig. 5, Supporting Information, Appendix S6) supports the role of the Dry Diagonal as a major biogeographic barrier for moist forest-adapted plants, with species diversity in subsections *Sphaerogyni* and *Spruceani* centred on the Atlantic Forest and the Amazonia, respectively. However, unlike *Amphilophium* Kunth (Bignoniaceae) (Thode *et al.*, 2019), the divergence between these two subsections (*c.* 20 Mya) postdates the Palaeogene formation of the Dry Diagonal (33–28 Mya, Table 1). Similar scenarios

(post-Dry Diagonal divergence) have been found in other groups of animals and plants such as lizards in the genus *Anolis* (Prates *et al.*, 2017), the plant genera *Stigmaphyllon* A.Juss. (Malpighiaceae) and *Attalea* Kunth (Arecaceae) (Freitas *et al.*, 2016; de Almeida & van der Berg, 2020), and the New World suboscine birds (Batalha-Filho *et al.*, 2013). This last study, in particular, identified multiple splits of Atlantic and Amazonian forest lineages with different ages, all post-dating the Dry Diagonal formation.

Our inferred biogeographic scenario supports an Amazonian origin for *Croton* section *Cleodora* at 19.7 Mya, in the Early Miocene (node 4, Fig. 5). The MRCA of *Croton* subsection *Spruceani* is also reconstructed as Amazonian (node 6) with strong support. Migration events from the Amazonian region to the Atlantic Forest explain the distribution of the MRCA of subsection *Sphaerogyni* in the Atlantic Forest, at 14.9 Mya (Fig. 5). Prado & Gibbs (1993) suggested that connectivity between the Amazonia and the Atlantic Forest across the Dry Diagonal was re-established along patches of river gallery forests during the warming events of the Early–Mid-Miocene (18–12 Mya), and similar connections during the MMCO (Table 1) have been found in other plant lineages (Peres *et al.*, 2020). The second most probable ancestral range in M1 for the MRCA of section *Cleodora* is a widespread ancestor occurring in both the Amazonia and the Atlantic Forest (Fig. 5). This would imply an earlier dispersal event between these two regions along the stem-branch of *Cleodora*, starting at the Palaeogene-Neogene boundary (24.5 Mya); however, this alternative reconstruction receives considerably weaker support (0.26) compared to 0.68 for an Amazonian crown-*Cleodora* (Fig. 5).

In any case, our reconstruction suggests that the split between the Amazonian *Spruceani* and Atlantic Forest *Sphaerogyni* pre-dated the formation of the two main biomes that now form the Dry Diagonal, the Caatinga and the Cerrado. The Caatinga, Cerrado and the Chaco biomes are characterized by sclerophyllous open woodlands and grasslands with a network of forest patches and corridors (Costa, 2003). Plant lineages in these biomes exhibit adaptations to drought, e.g. the Caatinga is rich in groups such as Cactaceae and Euphorbiaceae (Fernandes *et al.*, 2020), whereas the Cerrado is rich in fire-adapted plants (Simon *et al.*, 2009), characterized, e.g. by an underground xylopodium. Thus, we hypothesize that the ancestor of section *Cleodora* was probably moist-adapted, and that the vicariance event that initiated diversification in this section was linked to the intensification of drier conditions in the area of the current Dry Diagonal from the Mid-Cenozoic onwards, until the final establishment of the Caatinga and Cerrado vegetation in the Late Miocene–Pliocene.

The establishment of drier climates in the Neotropics could also explain the first vicariance event in our phylogenetic reconstruction. The split of section *Cupreati* in the Andes and sections *Cleodora* and *Cuneati* in the Amazonian region (Fig. 5, Supporting Information, Appendix S6) is dated as 27 Mya; however, its confidence interval (95% HPD: 30–25 Mya, Fig. 4) spans the LEEOC cooling event. This event initiated an aridification trend that led to the contraction of the early Palaeogene pan-Amazonian forest, which, together with initial uplift in the northern-central Andes (Hoorn *et al.*, 2010), might have contributed to the isolation of section *Cupreati* and the MRCA of *Cleodora–Cuneati*.

Unlike in *Amphilophium* (Thode *et al.*, 2019), in which most migration events occurred in the Atlantic Forest clade, dispersal to other geographical regions in *Croton* section *Cleodora* was mainly inferred in the Amazonian clade, subsection *Spruceani*. The timing of these events, however, is congruent with that of Thode *et al.* (2019). A dispersal event to Central America–Mexico was inferred along the branch 12.5–9.0 Mya (Fig. 5), in agreement with recent studies, pointing to a late-Middle Miocene closure of the Central American Seaway (15–12 Mya; Bacon *et al.*, 2015; Montes *et al.*, 2015; but see O'Dea *et al.*, 2016). The dispersal event from Central America–Mexico to the Andes, dated at 7.5–5.0 Mya (Fig. 5), also preceded the final closure of the Isthmus of Panama (*c.* 3.5 Mya; Coates *et al.*, 1992; O'Dea *et al.*, 2016) and could be associated with the uplift of the northern Andes in the Late Miocene–Pliocene (10–5 Mya; Table 1). The migration event between the Amazonian region and the Andes, dated 6–4 Mya (Fig. 5), could also be connected with this uplift, and the subsequent closure of the Acre system, at 7–5 Mya (Hoorn *et al.*, 2010). The western Amazonian mega-wetlands of the Acre and Pebas systems seem to have acted as a geographic barrier between the Amazonian forests and the eastern slopes of the Andes until the end of the Miocene, when they were gradually replaced by terra firme rainforest, following the uplift of the eastern part of the northern Andes (Antonelli *et al.*, 2009; Hoorn *et al.*, 2010).

The migration event from the Andes to the Atlantic Forest that preceded the split of *C. fragrantulus* and *C. rottlerifolius*, two of the few species in the section occurring in SDTFs, is dated at 3.25 Mya (Fig. 5, node 12) when the open Caatinga and Cerrado biomes were already fully formed. One possibility is that the ancestor of *C. fragrantulus* and *C. rottlerifolius* used the so-called 'Pleistocene Arch' (Prado & Gibbs, 1993; Mogni, Oakley & Prado, 2015); a climatic corridor for the dispersal of SDTF lineages between south-eastern Brazil and the Andes across the Dry Diagonal during the Late Pleistocene (33–25 Kya) and Last Glacial Maximum (18–12 Kya) (Prado & Gibbs, 1993; Mogni

et al., 2015). There is, however, one problem with this hypothesis. The estimated age of the MRCA of *C. fragrantulus* and *C. rottlerifolius* is 3.25 Myr, and therefore the dispersal event from the Andes to the Atlantic Forest took place before the Pleistocene (2.6–10.0 Mya) in our reconstruction (Fig. 5); the confidence interval, however, is large (5.83–1.32 Mya, Fig. 4) and spans the Early-Mid-Pleistocene period.

Alternatively, a forest corridor of SDTF vegetation might have existed before the Pleistocene. The Pliocene (5.33–2.60 Mya) was an epoch of rapid climate change. The Early Pliocene (5.33–3.60 Mya) was characterized by warm and humid conditions, which peaked in the Pliocene climatic optimum (4.35 Mya), with a short period of drier and colder climates 4.1–4.0 Mya. A global aridity trend climaxed in the MPWE, 3.2–2.8 Mya), when temperatures rose 2–3 °C worldwide, after which a decline in temperatures in the Late Pliocene culminated with the first glaciation at 2.6 Mya. The fossil record shows that these climatic changes were accompanied by major changes in floras and faunas (Willis & McDonald, 2011; Fedorov *et al.*, 2013; Jiménez-Moreno *et al.*, 2019). The expansion and later contraction of a SDTF corridor during the climatic oscillations of the Pliocene might have allowed the migration of dry-adapted arborescent plant lineages, such as the ancestor of *C. fragrantulus* and *C. rottlerifolius* across the Dry Diagonal.

Transitions from evergreen rain forest to seasonally dry biomes have been found in other Neotropical plant lineages (Bacon *et al.*, 2017; Lohmann *et al.*, 2013). A recent study (Zizka *et al.*, 2020) tested the effect of biome transitions among evergreen rain forest and SDTF on the diversification of Malvaceae subfamily Bombacoideae and showed that transition among these biomes are common and that the permeability was directional, with transitions from evergreen rain forest to seasonally dry biomes, being more common than the reverse. Although we have not analysed biome transitions in this study, our analyses suggest that transitions from wet to dry forests have probably been more common in the evolution of *Croton* section *Cleodora*.

CONCLUSIONS

This study provides the first nearly complete (> 96%) phylogenetic analysis of *Croton* section *Cleodora* based on six plastid and nuclear markers. Our spatio-temporal reconstruction supports a major role for geological, climatic and biome formation events during the Neogene in the formation of current distribution patterns in this section. The geographical disjunction between Amazonian and Atlantic Forest centres of diversity in *Cleodora* was probably driven by the appearance of the dryland biomes forming the South American Dry Diagonal, which agrees with patterns observed in other Neotropical plant and animal lineages. Section *Cleodora* appears to have originated in the Amazon region in the Early Miocene, from where it dispersed to other forested areas, including SDTFs, in the Neotropics during the Late Miocene and Pliocene.

ACKNOWLEDGEMENTS

This study stems from a project carried out by Irene Masa-Iranzo to fulfil the requirements of the Masters Degree in 'Biodiversity in Tropical Areas and its Conservation' (CSIC-UIMP), jointly supervised by RR and IS. We thank the staff technicians at RJB for help during laboratory work, with special mention to Itziar Cantero Pacheco, for guiding and helping in all laboratory stages. IS and RR were supported by the Spanish Ministry of Economy and Competitiveness (MINECO), the European Regional Development Fund (FEDER), and the State Research Agency (AEI) through grants CGL2015-67849 (MINECO/FEDER), CGL2015-73621-JIN (AEI/FEDER), and PID2019-108109GB-I00 (AEI/FEDER). Publication fee granted by the CSIC Open Access Publication Support Initiative through its Unit of Information Resources for Research (URICI).

REFERENCES

Alfaro ME, **Zoller S**, **Lutzoni F. 2003.** Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Molecular Biology and Evolution* **20:** 255–266.

de Almeida RF, **van den Berg C. 2020.** Biogeography of *Stigmaphyllon* (Malpighiaceae) and a meta-analysis of vascular plant lineages diversified in the Brazilian Atlantic Rainforests point to the Late Eocene origins of this megadiverse biome. *Plants* **9:** 1569.

Antonelli A, **Nylander JA**, **Persson C**, **Sanmartín I. 2009.** Tracing the impact of the Andean uplift on Neotropical plant evolution. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 9749–9754.

Antonelli A, **Sanmartín I. 2011.** Why are there so many plant species in the Neotropics? *Taxon* **60:** 403–414.

Antonelli A, **Zizka A**, **Carvalho FA**, **Scharn R**, **Bacon CD**, **Silvestro D**, **Condamine FL. 2018.** Amazonia is the primary source of Neotropical biodiversity. *Proceedings of the National Academy of Sciences of the United States of America* **115:** 6034–6039.

Arakaki M, **Christin PA**, **Nyffeler R**, **Lendel A**, **Eggli U**, **Ogburn RM**, **Spriggs E**, **Moore MJ**, **Edwards EJ. 2011.**

Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proceedings of the National Academy of Sciences of the United States of America* **108:** 8379–8384.

Arévalo R, **Van Ee BW**, **Riina R**, **Berry PE**, **Wiedenhoeft AC. 2017.** Force of habit: shrubs, trees and contingent evolution of wood anatomical diversity using *Croton* (Euphorbiaceae) as a model system. *Annals of Botany* **119:** 563–579.

Bacon CD, **Molnar P**, **Antonelli A**, **Crawford AJ**, **Montes C**, **Vallejo-Pareja MC. 2016.** Quaternary glaciation and the Great American Biotic Interchange. *Geology* **44:** 375–378.

Bacon CD, **Mora AS**, **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 CA**, **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:** 6110–6115.

Baele G, **Lemey P**, **Bedford T**, **Rambaut A**, **Suchard MA**, **Alekseyenko AV. 2012.** Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution* **29:** 2157–2167.

Baker PA, **Fritz SC**, **Dick CW**, **Eckert AJ**, **Horton BK**, **Manzoni S**, **Ribas CC**, **Garzione CN**, **Battisti DS. 2014.** The emerging field of geogenomics: constraining geological problems with genetic data. *Earth Science Review* **135:** 38–47.

Barthlott W, **Mutke J**, **Rafiqpoor D**, **Kier G**, **Kreft H. 2005.** Global centers of vascular plant diversity. *Nova Acta Leopoldina* **92:** 61–83.

Batalha-Filho H, **Fjeldså J**, **Fabre PH**, **Miyaki CY. 2013.** Connections between the Atlantic and the Amazonian forest avifaunas represent distinct historical events. *Journal of Ornithology* **154:** 41–50.

Berry PE, **Hipp AL**, **Wurdack KJ**, **Van Ee B**, **Riina R. 2005.** Molecular phylogenetics of the giant genus *Croton* and tribe Crotoneae (Euphorbiaceae *sensu stricto*) using ITS and *trnL-trnF* DNA sequence data. *American Journal of Botany* **92:** 1520–1534.

Bigarella JJ, **Andrade-Lima D**, **Riehs PJ. 1975.** Considerações a respeito das mudanças paleoambientais na distribuição de algumas espécies vegetais e animais no Brasil. *Anais da Academia Brasileira de Ciências* **47:** 411–464.

Blanquart S, **Lartillot N. 2006.** A Bayesian compound stochastic process for modeling nonstationary and nonhomogeneous sequence evolution. *Molecular Biology and Evolution* **23:** 2058–2071.

Bogarín D, **Pérez-Escobar OA**, **Karremans AP**, **Fernández M**, **Kruizinga J**, **Pupulin F**, **Smets E**, **Gravendeel B. 2019.** Phylogenetic comparative methods improve the selection of characters for generic delimitations in a hyperdiverse Neotropical orchid clade. *Scientific Reports* **9:** 1–17.

Burnham RJ, **Johnson KR. 2004.** South American palaeobotany and the origins of Neotropical rainforests. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **359:** 1595–1610.

Calió MF, **Lepis KB**, **Pirani JR**, **Struwe L. 2017.** Phylogeny of Helieae (Gentianaceae): resolving taxonomic chaos in a Neotropical clade. *Molecular Phylogenetics and Evolution* **106:** 192–208.

Canal D, **Köster N**, **Celis M**, **Croat TB**, **Borsch T**, **Jones KE. 2019.** Out of Amazonia and back again: historical biogeography of the species-rich Neotropical genus *Philodendron* (Araceae) 1. *Annals of the Missouri Botanical Garden* **104:** 49–68.

Cardoso D, **Särkinen T**, **Alexander S**, **Amorim AM**, **Bittrich V**, **Celis M**, **Daly DC**, **Fiaschi P**, **Funk VA**, **Giacomin LL**, **Goldenberg R**, **Heiden G**, **Iganci J**, **Kelloff CL**, **Knapp S**, **Cavalcante H**, **de Lima A**, **Machado FP**, **dos Santos RM**, **Mello-Silva R**, **Michelangeli FA**, **Mitchell J**, **Moonlight P**, **Rodrigues de Moraes PL**, **Mori SA**, **Nunes TS**, **Pennington TD**, **Pirani JR**, **Prance GT**, **de Queiroz LP**, **Rapini A**, **Riina R**, **Rincón CAV**, **Roque N**, **Shimizu G**, **Sobral M**, **Stehmann JR**, **Stevens WD**, **Taylor CM**, **Trovóa M**, **van den Berg C**, **van der Werff H**, **Viana PL**, **Zartman CE**, **Forzza RC. 2017.** Amazon plant diversity revealed by a taxonomically verified species list. *Proceedings of the National Academy of Sciences of the United States of America* **114:** 10695–10700.

Caruzo MBR, **Cordeiro I. 2013.** Taxonomic revision of *Croton* section *Cleodora* (Euphorbiaceae). *Phytotaxa* **121:** 1–41.

Caruzo MBR, **Cordeiro I**, **Berry PE**, **Riina R. 2010.** A new species of *Croton* section *Cleodora* (Euphorbiaceae *s.s.*) from Minas Gerais, Brazil. *Phytotaxa* **3:** 27–33.

Caruzo MBR, **Van Ee BW**, **Cordeiro I**, **Berry PE**, **Riina R. 2011.** Molecular phylogenetics and character evolution of the 'sacaca' clade: novel relationships of *Croton* section *Cleodora* (Euphorbiaceae). *Molecular Phylogenetics and Evolution* **60:** 193–206.

Castroviejo-Fisher S, **Guayasamin JM**, **Gonzalez-Voyer A**, **Vilà C. 2014.** Neotropical diversification seen through glassfrogs. *Journal of Biogeography* **41:** 66–80.

Chazot N, **Willmott KR**, **Lamas G**, **Freitas AV**, **Piron-Prunier F**, **Arias CF**, **Mallet J**, **de-Silva DL**, **Elias M. 2019.** Renewed diversification following Miocene landscape turnover in a Neotropical butterfly radiation. *Global Ecology and Biogeography* **28:** 1118–1132.

Cheng T, **Xu C**, **Lei L**, **Li C**, **Zhang Y**, **Zhou S. 2016.** Barcoding the kingdom Plantae: new PCR primers for ITS regions of plants with improved universality and specificity. *Molecular Ecology Resources* **16:** 138–149.

Coates AG, **Jackson JB**, **Collins LS**, **Cronin TM**, **Dowsett HJ**, **Bybell LM**, **Jung P**, **Obando JA. 1992.** Closure of the Isthmus of Panama: the near-shore marine record of Costa Rica and western Panama. *Geological Society of America Bulletin* **104:** 814–828.

Collevatti RG, **Ribeiro MDSL**, **Diniz Filho JAF**, **Oliveira GD**, **Dobrovolski R**, **Terribile LC. 2013.** Stability of Brazilian seasonally dry forests under climate change: inferences for long-term conservation. *American Journal of Plant Sciences* **4:** 792–805.

Costa LP. 2003. The historical bridge between the Amazon and the Atlantic Forest of Brazil: a study of molecular phylogeography with small mammals. *Journal of Biogeography* **30:** 71–86.

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

Dirzo R, **Raven PH. 2003.** Global state of biodiversity and loss. *Annual Review of Environment and Resources* **28:** 137–167.

Doyle J. 1991. DNA protocols for plants. In: Hewitt GM, Johnston AW, Young JPW, eds. *Molecular techniques in taxonomy*. Heidelberg: Springer-Verlag, 283–293.

Dong W, **Xu C**, **Li C**, **Sun J**, **Zuo Y**, **Shi S**, **Cheng T**, **Guo J**, **Zhou S. 2015.** *ycf1*, the most promising plastid DNA barcode of land plants. *Scientific Reports* **5:** 8348.

Drummond AJ, **Rambaut A. 2016.** TreeAnnotator v.1.8.4 [computer program]. Available at: http://beast.bio.ed.ac.uk/. Accessed 4 May 2019.

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

DRYFLOR. **2016.** Plant diversity patterns in Neotropical dry forests and their conservation implications. *Science* **353:** 1383–1387.

de Faria AD, **Pirani JR**, **Ribeiro JELDS**, **Nylinder S**, **Terra-Araujo MH**, **Vieira PP**, **Swenson U. 2017.** Towards a natural classification of Sapotaceae subfamily Chrysophylloideae in the Neotropics. *Botanical Journal of the Linnean Society* **185:** 27–55.

Fedorov AV, **Brierley CM**, **Lawrence KT**, **Liu Z**, **Dekens PS**, **Ravelo AC. 2013.** Patterns and mechanisms of early Pliocene warmth. *Nature* **496:** 43–49.

Fernandes MF, **Cardoso D**, **de Queiroz LP. 2020.** An updated plant checklist of the Brazilian Caatinga seasonally dry forests and woodlands reveals high species richness and endemism. *Journal of Arid Environments* **174:** 104079.

Fine PV, **Lohmann LG. 2018.** Importance of dispersal in the assembly of the Neotropical biota. *Proceedings of the National Academy of Sciences of the United States of America* **115:** 5829–5831.

Fouquet A, **Recoder R**, **Teixeira M Jr**, **Cassimiro J**, **Amaro RC**, **Camacho A**, **Damasceno R**, **Carnaval AC**, **Moritz C**, **Rodrigues MT. 2012.** Molecular phylogeny and morphometric analyses reveal deep divergence between Amazonia and Atlantic Forest species of *Dendrophryniscus*. *Molecular Phylogenetics and Evolution* **62:** 826–838.

Freitas C, **Meerow AW**, **Pintaud JC**, **Henderson A**, **Noblick L**, **Costa FR**, **Barbosa CE**, **Barrington D. 2016.** Phylogenetic analysis of *Attalea* (Arecaceae): insights into the historical biogeography of a recently diversified Neotropical plant group. *Botanical Journal of the Linnean Society* **182:** 287–302.

Freitas dos Santos R, **Riina R**, **Caruzo MBR. 2017.** Diversity of arborescent lineages of Crotoneae (Euphorbiaceae) in the Brazilian Atlantic Rain Forest. *Plant Systematics and Evolution* **303:** 1467–1497.

Garzione CN, **Auerbach DJ**, **Smith JJS**, **Rosario JJ**, **Passey BH**, **Jordan TE**, **Eiler JM. 2014.** Clumped isotope evidence for diachronous surface cooling of the Altiplano and pulsed surface uplift of the central Andes. *Earth and Planetary Science Letters* **393:** 173–181.

Garzione CN, **Hoke GD**, **Libarkin JC**, **Withers S**, **MacFadden B**, **Eiler J**, **Mulch A. 2008.** Rise of the Andes. *Science* **320:** 1304–1307.

Gentry AH. 1982. Neotropical floristic diversity: phytogeographical connections between Central and South America, Pleistocene climatic fluctuations, or an accident of the Andean orogeny? *Annals of the Missouri Botanical Garden* **69:** 557–593.

Givnish TJ, **Barfuss MH**, **Van Ee B**, **Riina R**, **Schulte K**, **Horres R**, **Gonsiska PA**, **Jabaily RS**, **Crayn DM**, **Smith JAC**, **Winter K. 2011.** Phylogeny, adaptive radiation, and historical biogeography in Bromeliaceae: insights from an eight-locus plastid phylogeny. *American Journal of Botany* **98:** 872–895.

Givnish TJ, **Barfuss MH**, **Van Ee B**, **Riina R**, **Schulte K**, **Horres R**, **Gonsiska PA**, **Jabaily RS**, **Crayn DM**, **Smith JAC**, **Winter K. 2014.** Adaptive radiation, correlated and contingent evolution, and net species diversification in Bromeliaceae. *Molecular Phylogenetics and Evolution* **71:** 55–78.

Gregory-Wodzicki KM. 2000. Uplift history of the central and northern Andes, a review. *Geological Society of America Bulletin* **112:** 1091–1105.

Guindon S, **Gascuel O. 2003.** A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Systematic Biology* **52:** 696–704.

Guo X, **Tang CC**, **Thomas DC**, **Couvreur TL**, **Saunders RM. 2017.** A mega-phylogeny of the Annonaceae: taxonomic placement of five enigmatic genera and support for a new tribe, Phoenicantheae. *Scientific Reports* **7:** 1–11.

Haffer J. 1969. Speciation in Amazonian forest birds. *Science* **165:** 131–137.

Hamilton MB. 1999. Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molecular Ecology* **8:** 521–523.

Hernández-Hernández T, **Brown JW**, **Schlumpberger BO**, **Eguiarte LE**, **Magallón S. 2014.** Beyond aridification: multiple explanations for the elevated diversification of cacti in the New World Succulent Biome. *New Phytologist* **202:** 1382–1397.

Höhna S, **Landis MJ**, **Heath TA**, **Boussau B**, **Lartillot N**, **Moore BR**, **Huelsenbeck JP**, **Ronquist F. 2016.** RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology* **65:** 726–736.

Hoorn C, **Wesselingh FP**, **Ter Steege H**, **Bermudez MA**, **Mora A**, **Sevink J**, **Sanmartín I**, **Meseguer AS**, **Anderson CL**, **Figueiredo JP**, **Jaramillo C. 2010.** Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science* **330:** 927–931.

Iturralde-Vinent MA. 2006. Meso–Cenozoic Caribbean paleogeography: implications for the historical biogeography of the region. *International Geology Review* **48:** 791–827.

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

Jaramillo C. 2018. Evolution of the Isthmus of Panama: biological, paleoceanographic and paleoclimatological implications. In: Hoorn C, Perrigo A, Antonelli A, eds. *Mountains, climate and biodiversity*. Oxford: Wiley Blackwell, 323–338.

Jiménez-Moreno G, **Pérez-Asensio JN**, **Larrasoaña JC**, **Sierro FJ**, **Garcia-Castellanos D**, **Salazar Á**, **Salvany JM**, **Ledesma S**, **Mata MP**, **Mediavilla C. 2019.** Early Pliocene climatic optimum, cooling and early glaciation deduced by terrestrial and marine environmental changes in SW Spain. *Global and Planetary Change* **180:** 89–99.

Kass RE, **Raftery AE. 1995.** Bayes factors. *Journal of the American Statistical Association* **90:** 773–795.

Katoh K, **Rozewicki J**, **Yamada KD. 2017.** MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics* **20**: 1160–1166.

Lagomarsino LP, **Condamine FL**, **Antonelli A**, **Mulch A**, **Davis CC. 2016.** The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). *New Phytologist* **210:** 1430–1442.

Landis MJ, **Freyman WA**, **Baldwin BG. 2018.** Retracing the Hawaiian silversword radiation despite phylogenetic, biogeographic, and paleogeographic uncertainty. *Evolution* **72:** 2343–2359.

Lavor P, **Calvente A**, **Versieux LM**, **Sanmartín I. 2019.** Bayesian spatio-temporal reconstruction reveals rapid diversification and Pleistocene range expansion in the widespread columnar cactus *Pilosocereus*. *Journal of Biogeography* **46:** 238–250.

Leier A, **McQuarrie N**, **Garzione C**, **Eiler J. 2013.** Stable isotope evidence for multiple pulses of rapid surface uplift in the central Andes, Bolivia. *Earth and Planetary Science Letters* **371:** 49–58.

Lohmann LG, **Bell CD**, **Calió MF**, **Winkworth RC. 2013.** Pattern and timing of biogeographical history in the Neotropical tribe Bignonieae (Bignoniaceae). *Botanical Journal of the Linnean Society* **171:** 154–170.

Lomolino MW, **Riddle BR**, **Whittaker RJ**, **Brown JH. 2016.** *Biogeography*, *5th edn*. Sunderland: Sinauer Associates; Oxford: Oxford University Press.

Lucas EJ, **Amorim BS**, **Lima DF**, **Lima-Lourenço AR**, **Lughadha EN**, **Proença CEB**, **Rosa PO**, **Rosário AS**, **Santos LL**, **Santos MF**, **Souza MC. 2018.** A new infra-generic classification of the species-rich Neotropical genus *Myrcia s.l*. *Kew Bulletin* **73:** 9.

Matzke NJ. 2016. Trait-dependent dispersal models for phylogenetic biogeography, in the R package BioGeoBEARS. *Integrative and Comparative Biology* **56:** E330.

Meseguer AS, **Lobo JM**, **Ree RH**, **Beerling DJ**, **Sanmartín I. 2015.** Integrating fossils, phylogenies, and niche models into biogeography to reveal ancient evolutionary history: the case of *Hypericum* (Hypericaceae). *Systematic Biology* **64:** 215–232.

Michelangeli FA, **Guimaraes PJ**, **Penneys DS**, **Almeda F**, **Kriebel R. 2013.** Phylogenetic relationships and distribution of new world Melastomeae (Melastomataceae). *Botanical Journal of the Linnean Society* **171:** 38–60.

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

Mogni VY, **Oakley LJ**, **Prado DE. 2015.** The distribution of woody legumes in Neotropical dry forests: the Pleistocene Arc Theory 20 years on. *Edinburgh Journal of Botany* **72:** 35–60.

Montes C, **Cardona A**, **Jaramillo C**, **Pardo A**, **Silva JC**, **Valencia V**, **Niño H. 2015.** Middle Miocene closure of the Central American seaway. *Science* **348:** 226–229.

Morrison JC, **Olson DM**, **Loucks CJ**, **Dinerstein E**, **Allnutt TF**, **Wikramanayake ED**, **Ricketts TH**, **Burgess ND**, **Kura Y**, **Powell GVN**, **Lamoreux JF**, **Underwood EC**, **Wettengel WW**, **D'Amico JA**, **Hedao P**, **Itoua I**, **Kassem KR**, **Strand HE. 2001.** Terrestrial ecoregions of the world: a new map of life on Earth. *BioScience* **51:** 933–938.

Mulch A, **Uba CE**, **Strecker MR**, **Schoenberg R**, **Chamberlain CP. 2010.** Late Miocene climate variability and surface elevation in the central Andes. *Earth and Planetary Science Letters* **290:** 173–182.

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

Nascimento AM, **Maria-Ferreira D**, **Dal Lin FT**, **Kimura A**, **de Santana-Filho AP**, **Werner MFDP**, **Iacomini M**, **Sassaki GL**, **Cipriani TR**, **De Souza LM. 2017.** Phytochemical analysis and anti-inflammatory evaluation of compounds from an aqueous extract of *Croton cajucara* Benth. *Journal of Pharmaceutical and Biomedical Analysis* **145:** 821–830.

Neves DM, **Dexter KG**, **Pennington RT**, **Bueno ML**, **Oliveira Filho AT. 2015.** Environmental and historical controls of floristic composition across the South American Dry Diagonal. *Journal of Biogeography* **42:** 1566–1576.

O'Dea A, **Lessios HA**, **Coates AG**, **Eytan RI**, **Restrepo-Moreno SA**, **Cione AL**, **Collins LS**, **de Queiroz A**, **Farris DW**, **Norris RD**, **Stallard RF. 2016.** Formation of the Isthmus of Panama. *Science Advances* **2:** 1600883.

Orme AR. 2007. Tectonism, climate, and landscape change. *The Physical Geography of South America* **1:** 23–44.

Padial JM, **Grant T**, **Frost DR. 2014.** Molecular systematics of terraranas (Anura: Brachycephaloidea) with an assessment of the effects of alignment and optimality criteria. *Zootaxa* **3825:** 1–132.

Paradis E, **Blomberg S**, **Bolker B**, **Brown J**, **Claude J**, **Cuong HS**, **Desper R. 2019.** Package 'ape'. *Analyses of Phylogenetics and Evolution* 2–4. Available at: https://cran.stat.unipd.it/web/packages/ape/ape.pdf

Pennington RT, **Dick CW. 2004.** The role of immigrants in the assembly of the South American rainforest tree flora. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **359:** 1611–1622.

Pennington RT, **Hughes CE. 2014.** The remarkable congruence of New and Old World savannah origins. *New Phytologist* **204:** 4–6.

Pennington RT, **Prado DE**, **Pendry CA. 2000.** Neotropical seasonally dry forests and Quaternary vegetation changes. *Journal of Biogeography* **27:** 261–273.

Pennington RT, **Ratter JA. 2006.** *Neotropical savannas and seasonally dry forests: plant diversity, biogeography, and conservation*. Boca Raton: CRC Press.

Peres EA, **Pinto-da-Rocha R**, **Lohmann LG**, **Michelangeli FA**, **Miyaki CY**, **Carnaval AC. 2020.** Patterns of species and lineage diversity in the Atlantic Rainforest of Brazil. In: Rull V, Carnaval AC, eds. *Neotropical diversification: patterns and processes*. Cham: Springer, 415–520.

Perret M, **Chautems A**, **Araujo AO**, **Salamin N. 2013.** Temporal and spatial origin of Gesneriaceae in the New World inferred from plastid DNA sequences. *Botanical Journal of the Linnean Society* **171:** 61–79.

Pirani RM, **Peloso PL**, **Prado JR**, **Polo ÉM**, **Knowles LL**, **Ron SR**, **Rodrigues MT**, **Sturaro MJ**, **Werneck FP. 2020.** Diversification history of clown tree frogs in Neotropical rainforests (Anura, Hylidae, *Dendropsophus leucophyllatus* group). *Molecular Phylogenetics and Evolution* **150:** 106877.

Prado DE. 2000. Seasonally dry forests of tropical South America: from forgotten ecosystem to a new phytogeographic unit. *Edinburgh Journal of Botany* **57:** 437–461.

Prado DE, **Gibbs PE. 1993.** Patterns of species distribution in the dry seasonal forests of South America. *Annals of the Missouri Botanical Garden* **80:** 902–927.

Prates I, **Melo-Sampaio PR**, **de Oliveira Drummond L**, **Teixeira M Jr**, **Rodrigues MT**, **Carnaval AC. 2017.** Biogeographic links between southern Atlantic Forest and western South America: rediscovery, re-description, and phylogenetic relationships of two rare montane anole lizards from Brazil. *Molecular Phylogenetics and Evolution* **113:** 49–58.

de Queiroz LP. 2006. The Brazilian Caatinga: phytogeographical patterns inferred from distribution data of the Leguminosae. In: Pennington RT, Ratter JA, eds. *Neotropical savannas and seasonally dry forests: plant diversity, biogeography, and conservation*. Boca Raton: CRC Press, 121–157.

de Queiroz LP, **Cardoso D**, **Ferreira Fernandes M**, **Freire Moro M. 2017.** Diversity and evolution of flowering plants of the Caatinga Domain. In: da Silva JMC, Leal IR, Tabarelli M, eds. *Caatinga: the largest tropical dry forest region in South America*. Cham: Springer, 23–65.

R Development Core Team. **2017.** *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rambaut A, **Suchard MA**, **Xie D**, **Drummond AJ. 2014.** *Tracer v.1.7.1*. Available at: http: //beast.bio.ed.ac.uk/Tracer. Accessed 4 May 2019.

Ree RH, **Sanmartín I. 2018.** Conceptual and statistical problems with the DEC+ J model of founder-event speciation and its comparison with DEC via model selection. *Journal of Biogeography* **45:** 741–749.

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

Ribeiro MC, **Metzger JP**, **Martensen AC**, **Ponzoni FJ**, **Hirota MM. 2009.** The Brazilian Atlantic Forest: how much is left, and how is the remaining forest distributed? Implications for conservation. *Biological Conservation* **142:** 1141–1153.

Riina R, **Berry PE**, **Secco RS**, **Meier W**, **Caruzo MBR. 2018.** Reassessment of *Croton* sect. *Cleodora* (Euphorbiaceae) points to the Amazon Basin as its main center of diversity. *Annals of the Missouri Botanical Garden* **103:** 330–350.

Riina R, **Berry PE**, **Van Ee BW. 2009.** Molecular phylogenetics of the dragon's blood *Croton* section *Cyclostigma* (Euphorbiaceae): a polyphyletic assemblage unraveled. *Systematic Botany* **34:** 360–374.

Riina R, **Van Ee BW**, **Wiedenhoeft AC**, **Cardozo A**, **Berry PE. 2010.** Sectional rearrangement of arborescent clades of *Croton* (Euphorbiaceae) in South America: evolution of arillate seeds and a new species, *Croton domatifer*. *Taxon* **59:** 1147–1160.

Ronquist F, **Deans AR. 2010.** Bayesian phylogenetics and its influence on insect systematics. *Annual Review of Entomology* **55:** 189–206.

Ronquist F, **Teslenko M**, **Van der Mark P**, **Ayres D**, **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.

Rodrigues MT, **Bertolotto CEV**, **Amaro RC**, **Yonenaga-Yassuda Y**, **Freire EMX**, **Pellegrino KCM. 2014.** Molecular phylogeny, species limits, and biogeography of the Brazilian endemic lizard genus *Enyalius* (Squamata: Leiosauridae): an example of the historical relationship between Atlantic Forests and Amazonia. *Molecular Phylogenetics and Evolution* **81:** 137–146.

Salatino A, **Salatino MLF**, **Negri G. 2007.** Traditional uses, chemistry and pharmacology of *Croton* species (Euphorbiaceae). *Journal of the Brazilian Chemical Society* **18:** 11–33.

Sang T, **Crawford DJ**, **Stuessy TF. 1997.** Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany* **84:** 1120–1136.

Sanmartín I. 2020. Breaking the chains of parsimony: the development of parametric approaches in historical biogeography. In: Cox CB, Ladle RJ, Moore PD, eds. *Biogeography: an ecological and evolutionary approach*, *10th edn*. Oxford: Wiley Blackwell, 283–287.

Sanmartín I, **Enghoff H**, **Ronquist F. 2001.** Patterns of animal dispersal, vicariance and diversification in the Holarctic. *Biological Journal of the Linnean Society* **73:** 345–390.

Saylor JE, **Horton BK. 2014.** Nonuniform surface uplift of the Andean plateau revealed by deuterium isotopes in Miocene volcanic glass from southern Peru. *Earth and Planetary Science Letters* **387:** 120–131.

Shaw J, **Lickey EB**, **Schilling EE**, **Small RL. 2007.** Comparison of whole chloroplast genome sequences to choose noncoding

regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany* **94:** 275–288.

Shorthouse DP. 2010. SimpleMappr, an online tool to produce publication-quality point maps. Available at: https://www.simplemappr.net. Accessed 9 November 2020.

Secco R de S. 2008. *Sinopse das espécies de* Croton *L. (Euphorbiaceae) na Amazônia Brazileira: um ensaio taxonômico*. Belém: Museu Paraense Emilio Goeldi (Coleção Adolfo Ducke).

da Silva JMC, **Leal IR**, **Tabarelli M. 2017.** *Caatinga: the largest tropical dry forest region in South America*. Cham: Springer.

da Silva JMC, **Rylands AB**, **da Fonseca GA. 2005.** The fate of the Amazonian areas of endemism. *Conservation Biology* **19:** 689–694.

Simmons MP. 2004. Independence of alignment and tree search. *Molecular Phylogenetics and Evolution* **31:** 874–879.

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.

Sobral-Souza T, **Lima-Ribeiro MS**, **Solferini VN. 2015.** Biogeography of Neotropical Rainforests: past connections between Amazon and Atlantic Forest detected by ecological niche modeling. *Evolutionary Ecology* **29:** 643–655.

Sodré RC, **de Sales MF**, **Berry PE**, **da Silva MJ. 2019.** Taxonomic synopsis of *Croton* section *Geiseleria* (Euphorbiaceae) in Brazil, including description of a new species. *Phytotaxa* **417:** 1–105.

Stadler T. 2009. On incomplete sampling under birth–death models and connections to the sampling-based coalescent. *Journal of Theoretical Biology* **261:** 58–66.

Swofford DL. 2002. *PAUP 4.0 b10: phylogenetic analysis using parsimony*. Sunderland: Sinauer Associates.

Taberlet P, **Gielly L**, **Pautou G**, **Bouvet J. 1991.** Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* **17:** 1105–1109.

Thode VA, **Sanmartín I**, **Lohmann LG. 2019.** Contrasting patterns of diversification between Amazonian and Atlantic forest clades of Neotropical lianas (*Amphilophium*, Bignonieae) inferred from plastid genomic data. *Molecular Phylogenetics and Evolution* **133:** 92–106.

Ulloa CU, **Acevedo-Rodríguez P**, **Beck S**, **Belgrano MJ**, **Bernal R**, **Berry PE**, **Brako L**, **Celis M**, **Davidse G**, **Forzza RC**, **Gradstein SR. 2017.** An integrated assessment of the vascular plant species of the Americas. *Science* **358:** 1614–1617.

Urbatsch LE, **Baldwin BG**, **Donoghue MJ. 2000.** Phylogeny of the coneflowers and relatives (Heliantheae: Asteraceae) based on nuclear rDNA internal transcribed spacer (ITS) sequences and chloroplast DNA restriction site data. *Systematic Botany* **25:** 539–565.

Van Ee BW, **Berry PE**, **Riina R**, **Amaro JEG. 2008.** Molecular phylogenetics and biogeography of the Caribbean-centered *Croton* subgenus *Moacroton* (Euphorbiaceae *ss*). *The Botanical Review* **74:** 132–165.

Van Ee BW, **Riina R**, **Berry PE. 2011.** A revised infrageneric classification and molecular phylogeny of New World *Croton* (Euphorbiaceae). *Taxon* **60:** 791–823.

Villaverde T, **Pokorny L**, **Olsson S**, **Rincón-Barrado M**, **Johnson MG**, **Gardner EM**, **Wickett NJ**, **Molero J**, **Riina R**, **Sanmartín I. 2018.** Bridging the micro-and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. *New Phytologist* **220:** 636–650.

Wesselingh FP. 2006. Evolutionary ecology of the Pachydontinae (Bivalvia, Corbulidae) in the Pebas lake/wetland system (Miocene, western Amazonia). *Scripta Geologica* **133:** 395–417.

Wesselingh FP, **Salo JA. 2006.** A Miocene perspective on the evolution of the Amazonian biota. *Scripta Geologica* **133:** 439–458.

Willis KJ, **MacDonald GM. 2011.** Long-term ecological records and their relevance to climate change predictions for a warmer world. *Annual Review of Ecology, Evolution and Systematics* **42:** 267–287.

Zachos J, **Pagani M**, **Sloan L**, **Thomas E**, **Billups K. 2001.** Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292:** 686–693.

Zizka A, **Carvalho-Sobrinho JG**, **Pennington RT**, **Queiroz LP**, **Alcantara S**, **Baum DA**, **Bacon CD**, **Antonelli A. 2020.** Transitions between biomes are common and directional in Bombacoideae (Malvaceae). *Journal of Biogeography* **47:** 1310–1321.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix S1. Habitat, distribution and elevation data of the 27 species of *Croton* section *Cleodora* (based on Riina *et al*., 2018).

Appendix S2. Summary of the PCR programs used for each molecular marker.

Appendix S3. Scripts used in biogeographical analysis in RevBayes.

Appendix S4. Majority-rule consensus tree of the MrBayes analysis of ITS data of *Croton* section *Cleodora* and related taxa. The numbers represent support values with Bayesian posterior probability (PP). Species sampled for the first time in a phylogenetic analysis, including recently described species, are indicated with an asterisk. The collector (initial of last name) and the collection number are included in the name of the taxa. The outgroup (*Brasiliocroton*) was removed from the figure.

Appendix S5. Majority-rule consensus tree of the MrBayes analysis of combined plastid *rps16*, *trnH-psbA*, *trnL-F*, *trnT-L* and *ycf1* data for *Croton* section *Cleodora* and related taxa. The numbers represent support values with Bayesian posterior probability (PP). Species sampled for the first time in a phylogenetic analysis, including recently described species, are indicated with an asterisk. The collector (initial of last name) and the collection number are included in the name of the taxa. The outgroup (*Brasiliocroton*) was removed from the figure.
Appendix S6. Model M0 reconstruction of the spatio-temporal evolution of *Croton* section *Cleodora* and closest allies, using Bayesian dispersal extinction cladogenesis (DEC). The phylogenetic tree is the MCC tree from BEAST shown in Figure 4. A pie chart on the node indicates alternative ancestral geographical ranges, with pie slices indicating their marginal posterior probabilities (MPP). Values below nodal pie charts indicate the MPP for the ancestral range receiving the highest probability; only values < 0.90 are shown. Numbers above nodal pie charts indicate the nodes mentioned in the text. The smaller pie charts on the branches indicate the geographical range immediately after the speciation event; a pie on the node indicates how this range has changed as a result of range expansion and contraction events along the branch. Colours represent the geographical range state according to the legend. Letters on the map indicate the operational areas according to the legend. The dashed vertical lines indicate geological or global climatic events which may have played a role in the spatio-temporal evolution of section *Cleodora*. LOWE: Late Oligocene Warming Event, MMCO: Mid-Miocene climatic optimum, CAM: formation of the Central American Landbridge, PAFC: Pleistocene Arch forest corridor.

APPENDIX 1. TAXA, LOCALITIES, VOUCHERS AND GENBANK ACCESSION NUMBERS FOR ALL SEQUENCES ANALYSED. GENBANK ACCESSION NUMBERS FOR NEWLY GENERATED SEQUENCES ARE PRECEDED BY AN ASTERISK (*). TAXON; LOCALITY; VOUCHER; ITS; *rps16*; *trnH-psbA*; *trnL-F*; *trnT-L*; *ycf1*. MISSING DATA: –.

Brasiliocroton mamoninha P.E.Berry & Cordeiro; Brazil, Bahia; Lombardi 7148 (HRCB); *MW263126; –; –; *MW266672; –; –. *B. muricatus* Riina & Cordeiro; Brazil, Bahia; Carneiro-Torres 1000 (HUESF); KF208629; *MW266609; –; KF208632; *MW266706; *MW266590. *Croton amazonicus* Müll.Arg.; Colombia, Caquetá; Castroviejo 12034 (COL); *MW263129; –; *MW266644; *MW266679; –; –. *C. amazonicus* Müll.Arg.; Brazil, Amazonas, Presidente Figueiredo; Ferreira 6961 (INPA); *MW263130; –; *MW266645; *MW266680; –; –. *C. amazonicus* Müll.Arg.; Brazil, Maranhão; Rosa 2898 (DAV); *MW263131; –; *MW266646; *MW266681; *MW266712; –. *C. beetlei* Croizat; Bolivia, Santa Cruz; Riina 1512 (WIS); *MW263161; *MW266632; –; *MW266700; *MW266741; *MW266601. *C. billbergianus* Müll.Arg.; Costa Rica, Alajuela; Van Ee 342 (WIS); EU477998; –; *MW266647; –; –; –. *C. billbergianus* Müll.Arg.; Mexico, Veracruz; León 104 (MICH); *MW263132; *MW266615; –; *MW266682; *MW266713; –. *C. billbergianus* Müll.Arg.; Costa Rica, Limón; Van Ee 595 (WIS); *MW263133; *MW266614; *MW266648; *MW266683; *MW266714; *MW266594. *C. cajucara* Benth.; Brazil; Caruzo 96 (SP); HM044789; –; HM044814; HM044770; –; –. *C. cajucara* Benth.; Brazil; Caruzo 95 (SP); *MW263134; *MW266616; *MW266651; *MW266684; *MW266710; *MW266595. *C. campanulatus* Caruzo & Cordeiro; Brazil, Rio de Janeiro; Caruzo 93 (SP); *MW263138; –; *MW266652; *MW266678; *MW266715; –. *C. campanulatus* Caruzo & Cordeiro; Brazil, Rio de Janeiro; Santos 45 (RB); *MW263139; *MW266617; *MW266653; *MW266677; *MW266716; –. *C. chimboracensis* P.E.Berry & Riina; Ecuador, Chimborazo; Berry 7618 (WIS); AY971204; *MW266610; –; AY971293; *MW266708; *MW266591. *C. coryi* Croizat; U.S.A., Texas; Van Ee 520 (WIS); –; *MW266613; –; –; –; –. *C. cuneatus* Klotzsch; Peru; Riina 1491 (MICH); EU478005; –; *MW266640; –; *MW266742; –. *C. cupreatus* Croizat; Ecuador, Pichincha; Riina 1408 (WIS); EU586919; *MW266634; –; EU586974; *MW266744; *MW266603. *C. discolor* Willd.; Dominican Republic, Barahona; Van Ee 629 (WIS); *MW263127; *MW266611; –; *MW266674; *MW266709; *MW266592. *C. ekmanii* Urb.; Cuba, Holguín; Van Ee 393 (WIS); *MW263162; *MW266635; –; –; *MW266745; *MW266604. *C. fragilis* Kunth; Venezuela, Sucre, La Llanada Vieja; Riina 1295 (VEN); *MW263128; *MW266612; –; *MW266673; *MW266707; *MW266593. *C. fragrans* Kunth; Venezuela, Bolívar; Fernández 15187 (IVIC); *MW263135; –; –; –; –; –. *C. fragrans* Kunth; Venezuela, Cojedes; Riina 1824 (MICH); *MW263136; *MW266618; *MW266654; –; *MW266723; *MW266596. *C. fragrantulus* Croizat; Bolivia, Santa Cruz; Foster 424 (MO); HM044791; *MW266619; HM044820; HM044772; *MW266724; *MW266597. *C. gigantifolius* P.E.Berry & Secco; Brazil, Rondônia; Silva 2055 (CEN); *MW263137; –; *MW266649; –; *MW266711; –. *C. hemiargyreus* Müll.Arg.; Brazil, Minas Gerais; Caruzo 114 (SP); HM044793; –; HM044824; HM044774; –; –. *C. hemiargyreus* Müll. Arg.; Brazil, Pernambuco, Petrolina; Santos 43 (HTSA); *MW263140; *MW266620; *MW266658;

*MW266676; *MW266718; –. *C. hemiargyreus* Müll.Arg.; Brazil, Minas Gerais, Salto da Divisa; Thomas 13720 (MO); *MW263141; –; *MW266659; –; *MW266719; –. *C. heterocalyx* Baill.; Brazil, Bahia; Caruzo 108 (SP); HM044794; –; HM044825; HM044775; *MW266721; –. *C. heterocalyx* Baill.; Brazil, Espírito Santo, Guarapari; Santos 11 (RB); *MW263142; *MW266621; *MW266660; *MW266691; *MW266722; –. *C. hoffmannii* Müll.Arg.; Costa Rica, Cartago, Paraíso; Van Ee 598 (WIS); EF421773; *MW266623; –; –; *MW266729; –. *C. hoffmannii* Müll.Arg.; French Guiana, Saül; Mori 15388 (MO); *MW263143; –; *MW266655; *MW266685; *MW266727; –. *C. hoffmannii* Müll.Arg.; Costa Rica, Limón, Pococí; Rodríguez 5211 (MO); *MW263144; *MW266622; *MW266661; *MW266686; *MW266728; *MW266598. *C. javarisensis* Secco; Peru, Loreto; Rimachi 337 (DAV); *MW263145; *MW266624; *MW266662; *MW266688; *MW266730; –. *C. leonis* (Croizat) Van Ee & P.E. Berry; Cuba, Holguín; Hajb 81773 (MICH); EF421758; *MW266636; –; EF408140; *MW266746; *MW266605. *C. loretensis* Riina & Caruzo; Peru, Loreto; McDaniel 10777 (MO); *MW263146; –; *MW266663; *MW266689; *MW266731; –. *C. mayumbensis* J.Léonard; Gabon; Bissiengou 704 (WAG); *MW263163; –; –; *MW266702; *MW266747; *MW266606. *C. nirguensis* Riina & Meier; Venezuela, Carabobo; Meier 19318 (VEN); *MW263147; *MW266625; *MW266657; *MW266690; *MW266732; –. *C. organensis* Baill.; Brazil, Rio de Janeiro; Caruzo 90 (WIS); EU586914; *MW266626; HM044832; EU586969; *MW266725; –. *C. organensis* Baill.; Brazil, Rio de Janeiro, Itatiaia; Caruzo 122 (SP); *MW263148; –; *MW266664; *MW266692; *MW266726; –. *C. orinocensis* Müll.Arg.; Venezuela, Amazonas; Riina 1818 (MICH); *MW263149; *MW266627; *MW266665; *MW266693; *MW266733; *MW266599. *C. pachypodus* G.L. Webster; Colombia, Nariño; Ávila 1005 (COL); *MW263165; *MW266637; –; *MW266703; *MW266749; *MW266607. *C. perstipulatus* G.L.Webster ex Caruzo & Secco; Brazil, Acre; Lowrie 279 (INPA); *MW263150; –; *MW266666; *MW266694; *MW266734; –. *C. piptocalyx* Müll.Arg.; Brazil, São Paulo; Lima 527 (SP); *MW263167; *MW266639; –; *MW266705; *MW266751; *MW266608. *C. pseudofragrans* Croizat; Peru, Loreto; Woodward s.n. (MICH); HM044800; –; HM044837; HM044780; *MW266735; –. *C. rottlerifolius* Baill.; Brazil, São Paulo; Caruzo 56 (SP); EU586915; –; HM044838; EU586970; *MW266736; –. *C. rottlerifolius* Baill.; Brazil, São Paulo; Santos 41 (SP); *MW263151; *MW266628; *MW266656; *MW266687; *MW266737; –. *C. rufolepidotus* Caruzo & Riina; Colombia, Antioquía; Callejas 3611 (MO); *MW263152; –; *MW266667; *MW266695; –; –. *C. salutaris* Casar.; Brazil, Rio de Janeiro; Caruzo 89 (SP); HM044804; *MW266629; HM044840; HM044783; *MW266717; –. *C. santaritensis* Huft; Panama; Lee 8075 (MICH); *MW263153; –; *MW266668; *MW266696; *MW266738; –. *C. sapiifolius* Müll. Arg.; Brazil, Bahia; Lima 667 (CEPEC); *MW263166; *MW266638; –; *MW266704; *MW266750; –. *C. sexmetralis* Croizat; Venezuela, Táchira; Bono 4229 (MO); *MW263154; *MW266630; *MW266669; *MW266697; *MW266739; –. *C. sphaerogynus* Baill.; Brazil, Rio de Janeiro; Van Ee 505 (WIS); HM044805; –; –; HM044784; –; –. *C. sphaerogynus* Baill.; Brazil, Rio de Janeiro; Caruzo 65 (SP); *MW263156; –; *MW266642; –; *MW266720; –. *C. sphaerogynus* Baill.; Brazil, Rio de Janeiro; Caruzo 121 (SP); *MW263157; –; *MW266643; –; –; –. *C. spruceanus* Benth.; Colombia, Valle del Cauca; Baker 6493 (MO); HM044806; –; –; –; –; –. *C. spruceanus* Benth.; Brazil, Pará; Caruzo 101 (SP); *MW263158; *MW266631; *MW266670; *MW266698; *MW266740; *MW266600. *C. stellatoferrugineus* Caruzo & Cordeiro; Brazil, Minas Gerais; Anderson 35673 (NY); *MW263155; –; –; –; –; –. *C. stellatoferrugineus* Caruzo & Cordeiro; Brazil, Minas Gerais; Caruzo 120 (SP); *MW263159; –; *MW266671; –; –; –. *C. tchibangensis* Pellegr.; Gabon, Nyanga; Valkenburg 2668 (WAG); *MW263164; –; –; *MW266701; *MW266748; –. *C. viroleoides* P.E.Berry & Secco; Brazil, Rondônia; Zarucchi 2789 (INPA); *MW263160; –; *MW266650; *MW266699; –; –. *C. yavitensis* Croizat; Bolivia, Beni; Beck 5710 (LPB); EU586918; *MW266633; *MW266641; *MW266675; *MW266743; *MW266602.