Phylogenetic relationships of a major diversification of *Croton* (Euphorbiaceae) in the western Indian Ocean region

ELIZABETH A. HABER[1†], KENT KAINULAINEN[1], BENJAMIN W. VAN EE[2], BEN O. OYSERMAN[1‡] and PAUL E. BERRY[1*]

[1]*University of Michigan Herbarium and Department of Ecology and Evolutionary Biology, 3600 Varsity Drive, Ann Arbor, MI 48108, USA*
[2]*Department of Biology, Universidad de Puerto Rico, Recinto Universitario de Mayagüez, Mayagüez, PR 00680, Puerto Rico*

Croton is one of the most diverse genera of flowering plants in Madagascar, with >150 endemic species encountered in nearly all vegetation types. This study focuses on the evolutionary history of *Croton* in the western Indian Ocean region (WIOR), which includes Madagascar, the Comoro Islands and the Mascarene Islands, and aims to resolve species relationships and the biogeographic history of the genus there. We obtained nuclear ribosomal internal transcribed spacer and plastid (*matK*, *trnH-psbA* and *trnL-F*) sequence data for 71, or close to half, of the *Croton* spp. in the WIOR and 15 other Old World species and seven New World species. We analysed them phylogenetically using Bayesian and maximum likelihood criteria and performed an ancestral state reconstruction of their geographical distributions. Our results show general support for a single origin of *Croton* on Madagascar from Africa. This was followed by an extensive radiation into nearly all available habitats in Madagascar and subsequent dispersal events out of Madagascar, including two back colonizations to Africa, one or more dispersals to the Comoro Islands and a single dispersal to the Mascarene Islands.

ADDITIONAL KEYWORDS: Comoro Islands – Madagascar – Mascarene Islands – molecular phylogenetics.

INTRODUCTION

Croton L. (Euphorbiaceae) is one of the ten largest genera of flowering plants, comprising >1200 species (Govaerts, Frodin & Radcliffe-Smith, 2000; Frodin, 2004). As a result of a broad molecular survey by Van Ee, Riina & Berry (2011), *Croton* was separated into four subgenera, three of which are restricted to the New World, whereas the fourth, *C.* subgenus *Croton*, appears to be an entirely Old World clade. The possibility of a monophyletic Old World clade suggests a single origin for Old World *Croton* from the New World (Berry *et al.*, 2005).

In the Old World, Madagascar hosts one of the most diverse and threatened floras in the world (Ganzhorn *et al.*, 2001). As of early 2012, the estimated number of plant species on the island was 11 350 with 10 839 of these considered to be native (Madagascar Catalogue, 2013). With *c*. 150 species on Madagascar (all endemic there), *Croton* accounts for nearly 1.5% of all plant species on the island (Schatz, 2001). It is among the three most species-rich genera in Madagascar, with *Dombeya* Cav. (Malvaceae; *c*. 180 species) and *Diospyros* L. (Ebenaceae; *c*. 100 described and many undescribed species; Schatz, 2001). *Croton* can be found in essentially all vegetation types in Madagascar, ranging from the arid south-west to the moist north-east (Grubb, 2003) and from sea level to the highest peaks of the island. It is a major component of the dry forests in southern Madagascar (Sussman

*Corresponding author. E-mail: peberry@umich.edu
[†]Current address: Copernicus Institute of Sustainable Development, University of Utrecht, Heidelberglaan 2, 3584 CS Utrecht, The Netherlands
[‡]Current address: Department of Civil and Environmental Engineering, University of Wisconsin-Madison, 3207D Engineering Hall, 1415 Engineering Dr, Madison, WI 53706, USA

& Rakotozafy, 1994; Grubb, 2003) and in the eastern lowland rainforests (Dumetz, 1999; Klanderud *et al.*, 2010).

Croton is generally characterized by a dense covering of stellate or lepidote trichomes, conspicuous petiolar and/or laminar glands, clear to coloured sap, unisexual flowers and orange senescent leaves. *Croton* spp. in Madagascar are all woody trees or shrubs and many have opposite or pseudo-verticillate leaves, which is unusual outside the region. Several species also differ from most New World species in having well-developed petals in the pistillate flowers. Species delimitation in Malagasy *Croton* is in many cases unclear and notoriously difficult; this is due to incomplete material for assessing the range of characters and to individuals exhibiting intermediate morphological characters (Leandri, 1939).

Jacques Leandri was the foremost botanist to work on western Indian Ocean region (WIOR) *Croton* spp. He worked at the Paris herbarium, which has the most complete collection of botanical specimens from the region, and he also made several field excursions to Madagascar to observe the genus in situ. He produced a comprehensive treatment of WIOR *Croton* (Leandri, 1939) in which he recognized 94 species in 18 informal taxonomic groups, rather than trying to place them into the formal sections that were recognized by Henri Baillon (1858) and Müller Argovensis (1866). His groups were defined by characters such as the number of stamens in staminate flowers, the type of indumentum of leaf undersides, the presence or absence of petals in the pistillate flowers, and the size and shape of the leaves. He was clearly aware that more studies, both within and outside of the region, would be required to determine how accurately these groups might reflect true affinities among the species. By the time Leandri published his last papers on Malagasy *Croton* in 1973, he had increased the total of accepted *Croton* spp. from Madagascar, the Comoro Islands and the Mascarenes to 137 (Leandri, 1948, 1969, 1970a, b, 1972a, b, 1973a, b, c).

Grady Webster (1993) provided a revised worldwide synopsis of *Croton*, which relied heavily on the classification of Baillon (1858). Webster's synopsis divided the genus into 40 sections and a few subsections. Among these, nine sections were represented in the WIOR. Five of these were restricted to the WIOR, namely *Croton* sections *Andrichnia* Baill., *Anisophyllum* Baill., *Decapetalon* Müll.Arg., *Klotzschiphytum* (Baill.) Baill. and *Monguia* Baill. The others, including *C.* sections *Cyclostigma* (Griseb.) Müll.Arg. and *Argyrocroton* (Müll.Arg.) G.L.Webster, included species from the New World or other parts of the Old World, with species from Madagascar. A detailed molecular and morphological study by Riina, Berry &

Van Ee (2009) showed that the Malagasy tree species *Croton goudotii* Baill. and *Croton mongue* Baill. do not belong to the largely New World arborescent group of *Croton* section *Cyclostigma* but are instead convergent in their habit and floral characters. Likewise, Van Ee *et al.* (2011) showed that the New World species that Webster (1993) had ascribed to *Croton* section *Argyrocroton* belong to different New World sections, leaving *C.* section *Argyrocroton* entirely Old World in distribution. A further drawback to Webster's system was that he left many Old World species unassigned to section in his treatment, due to his avowed lack of familiarity with *Croton* from Africa, Madagascar and Asia.

The late Alan Radcliffe-Smith, Old World Euphorbiaceae expert at the Royal Botanic Gardens, Kew, left behind in 2002 a manuscript on *Croton* from Madagascar and the Comoro Islands that included 150 native species and 146 varieties, with 26 new species, three new subspecies and 113 new varieties proposed. However, there were numerous unresolved issues with the manuscript, so it was left unpublished until Martin Cheek decided to deposit three copies of a slightly revised version at three botanical libraries in London. As evidenced by the listing of the new names in the IPNI online database (www.ipni.org), this qualifies as an effective publication, as of 23 December 2016 (Radcliffe-Smith, 2016). Based on our own more comprehensive review of *Croton* worldwide (Berry *et al.*, 2005; Van Ee *et al.*, 2011; Van Ee, Forster & Berry, 2015) and of the WIOR in particular (Berry & Van Ee, 2011; Berry, Van Ee & Haber, 2011; Berry *et al.*, 2016a, b; Kainulainen *et al.*, 2016), we anticipate that >90% of the new names published by Radcliffe-Smith will end up in synonymy due to oversplitting or misplacement under incorrect species.

All molecular phylogenetic studies to date that have included WIOR species have recovered them as a monophyletic group, although none of these has included >20 species from the region (Berry *et al.*, 2005; Riina *et al.*, 2009; Van Ee *et al.*, 2011, 2015). Our current project has involved a detailed review of existing herbarium specimens, several field trips to Madagascar and extensive DNA sequencing of freshly collected material and herbarium specimens. Our overall aim is threefold: (1) to revise the species-level taxonomy of WIOR *Croton* spp.; (2) to evaluate the phylogenetic relationships among the WIOR species and (3) to determine the biogeographic history of the group in relation to Africa and other parts of the world. This particular study focuses on the last two of these aims, attempting to determine whether the WIOR *Croton* spp. form a monophyletic group and if there is any evidence of dispersal of *Croton* out of Madagascar to other parts of the world.

MATERIAL AND METHODS

TAXON SAMPLING

To infer the evolutionary relationships of western Indian Ocean *Croton* spp. and their relationship with *Croton* spp. from other parts of the world, we included as many taxa from Madagascar and the surrounding islands as possible. During two collection trips to Madagascar in 2009, we collected nearly 200 voucher specimens and silica gel-dried samples of *Croton* for DNA analysis. We also sampled specimens from five herbaria (K, MO, P, TAN and WIS). One hundred accessions from 71 *Croton* spp. from Madagascar and the surrounding islands were included in this study, including 11 undescribed species. This sampling represents *c*. 50% of the total diversity of *Croton* in the WIOR. For the rest of the Old World, 15 representative species were sampled from Africa, Australia and Southeast Asia, out of 68 non-WIOR Old World species analysed by Van Ee *et al*. (2015). That study, which concluded among other things that Old World *Croton* spp. are monophyletic and suggested that the Madagascan species are part of a clade sister to a subset of African species, included 16 species from Madagascar, in an inverse proportion to the sampling here (16 species from Madagascar and 68 non-WIOR Old World species, compared to 71 species from Madagascar and 15 non-WIOR Old World species in this study). Seven New World *Croton* spp., representing the three exclusively New World subgenera (Van Ee *et al*., 2011), were selected as outgroup taxa to the Old World *Croton* spp. Following the findings of Berry *et al*. (2005) and Wurdack, Hoffmann & Chase (2005), *Astraea lobata* (L.) Klotzsch and *Brasiliocroton mamoninha* P.E.Berry & Cordeiro were used as outgroups to *Croton*. Details of the sampling and voucher specimens are given in Supporting Information Table S1. All of our cited specimens are available in the Tropicos database (www.tropicos.org), including photographs of many of the collections from Madagascar.

DNA EXTRACTION, AMPLIFICATION AND SEQUENCING

Total genomic DNA was extracted from silica gel-dried specimens and fragments sampled from existing herbarium specimens using the Qiagen DNeasy Plant Mini kit (Qiagen, Valencia, CA, USA) according to the manufacturer's protocol. For herbarium specimens, the warm incubation step was lengthened to 1 h while mixing at 1200 r.p.m. to ensure cell lysis, and the cold incubation step was lengthened to overnight to precipitate most of the polysaccharides and proteins. Seven DNA regions, the nuclear ribosomal internal transcribed spacer (ITS), exon 9 of the low-copy nuclear gene EMBRYO DEFECTIVE 2765 (EMB2765), the plastid regions *matK, trnL-F* (*trnL^{UAA}-trnL^{UAA}-trnF^{GAA}*),

trnH^{(GUG)}-psbA and *psbK-psbI* and the mitochondrial intron *rps3* were tested on a subset of specimens for ease of amplification and number of informative sites. Of these, ITS, *matK*, *trnL-F* and *trnH-psbA* were chosen for this study. Polymerase chain reaction (PCR) amplifications were performed using the following primers: ITSi (Urbatsch, Baldwin & Donoghue, 2000), ITS2, ITS3 and ITS4 (White *et al.*, 1990; ITS2 and ITS3 only used when ITSi–ITS4 failed to amplify) for ITS; trnK570F, matK1710R (Samuel *et al.*, 2005) and matK353F (5′ TTTGGAAAGGTCGGAGACAGT 3′, newly designed), for *matK*; c, d, e and f (Taberlet *et al.*, 1991; d and e only used when c and f failed to amplify) for *trnL-F* and psbAf (Sang, Crawford & Stuessy, 1997) and trnH2 (Tate & Simpson, 2003) for *trnH-psbA*. The plastid regions were amplified using Ex Taq polymerase (Takara Bio Inc., Otsu, Shiga, Japan) and standard PCR settings, whereas ITS was amplified using Qiagen Taq polymerase (Qiagen, Valencia, CA, USA). The following touchdown PCR protocol was used for ITS: an initial denaturing step at 95 °C for 2 min followed by 46 cycles of denaturing at 95 °C for 1 min, annealing of 1 min at 65 °C for 1 cycle, 64 °C for 2 cycles, 63 °C for 3 cycles, 62 °C for 4 cycles, 60 °C for 5 cycles, 57 °C for 6 cycles and 55 °C for 25 cycles, a 1-min elongation step at 72 °C for each of the 46 cycles and a final 15-min elongation step at 72 °C. PCR products were purified using ExoSap–IT (USB Corporation, Cleveland, OH, USA.) and sequenced on an ABI 3730 XL sequencer (Applied Biosystems, ABI, Foster City, CA, USA). Sequences were edited using Sequencher v.4.10.1 (Gene Codes, Ann Arbor, MI, USA).

ALIGNMENT AND PHYLOGENETIC ANALYSES

Edited sequences were aligned using MUSCLE v.3.8.31 (Edgar, 2004) with default parameters selected, and the alignment was adjusted manually in BioEdit v.7.1.3 (Hall, 1999). DNA substitution models were evaluated using MrAIC v.1.4.4 (Nylander, 2004; a script dependent on the program PHYML v.3.0; Guindon & Gascuel, 2003); the GTR+I+G model was indicated as best-fit to the data for ITS and the combined plastid data (*matK*, *trnL-F* and *trnH-psbA*). Phylogenetic reconstructions were done by Bayesian Markov chain Monte Carlo inference (MCMC; Yang & Rannala, 1997), using the program MrBayes v.3.2.6 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003) on the Cipres Science Gateway portal v.3.3 (Miller, Pfeiffer & Schwartz, 2010). Three data sets were analysed: (1) ITS alone; (2) the three plastid regions combined and (3) ITS together with the plastid data as two partitions with unlinked model parameter estimates (except for topology). The MCMC was run for 20 million generations (sampling every 1000th generation), of which the

first 25% were excluded as burn-in. Effective sample sizes of model parameters were verified using Tracer v.1.5 (Rambaut & Drummond, 2007), and the phylogenetic trees were visualized using FigTree v.1.4 (Rambaut, 2012). Corresponding analyses, using the same partitions and substitution models, were also done in a maximum likelihood (ML) framework using the program IQ-tree (Nguyen *et al.*, 2015; Chernomor, Von Haeseler & Minh, 2016). Branch support was assessed by 1000 standard bootstrap replicates.

Shifts in geographical distribution were analysed by ML ancestral state reconstruction using Mesquite v.3.1 (Maddison & Maddison, 2016). Seven geographical areas were recognized: (1) the Americas, (2) mainland Africa and Socotra, (3) Asia, (4) Australia, (5) Madagascar, (6) the Comoro Islands and (7) the Mascarene Islands. To account for phylogenetic uncertainty, the analysis was done across 1000 trees randomly selected from the post-burn-in tree sample of the MrBayes analysis of the combined data. The Mk1 model (equal probability of change between states) was used, and the results were plotted as average state frequencies (probability for each state at each node) on the ML topology inferred in the IQ-tree analysis.

RESULTS

In total, 366 DNA sequences were newly generated for this study and have been deposited in GenBank (see Supporting Information, Table S1). The aligned ITS data comprise 699 base pairs (291 of them variable, 226 of these potentially informative), and the combined plastid data comprise 3920 base pairs (780 variable, 289 potentially informative). The phylogenetic hypotheses inferred from the ITS data (Supporting Information, Figs S1, S3) and the combined plastid data (Supporting Information, Figs S2, S4) show few moderately or strongly supported clades, i.e. those with a posterior probability (pp) of ≥0.95 (Bayesian analyses) or a bootstrap support (bs) of ≥70% (ML analyses). The results of the Bayesian and ML methods of phylogenetic reconstruction are similar and produced no strongly supported incongruences. The nuclear and plastid phylogenetic trees are generally congruent, differing only at a few shallow and weakly supported levels. We therefore focused our discussion on the combined nuclear and plastid data analyses.
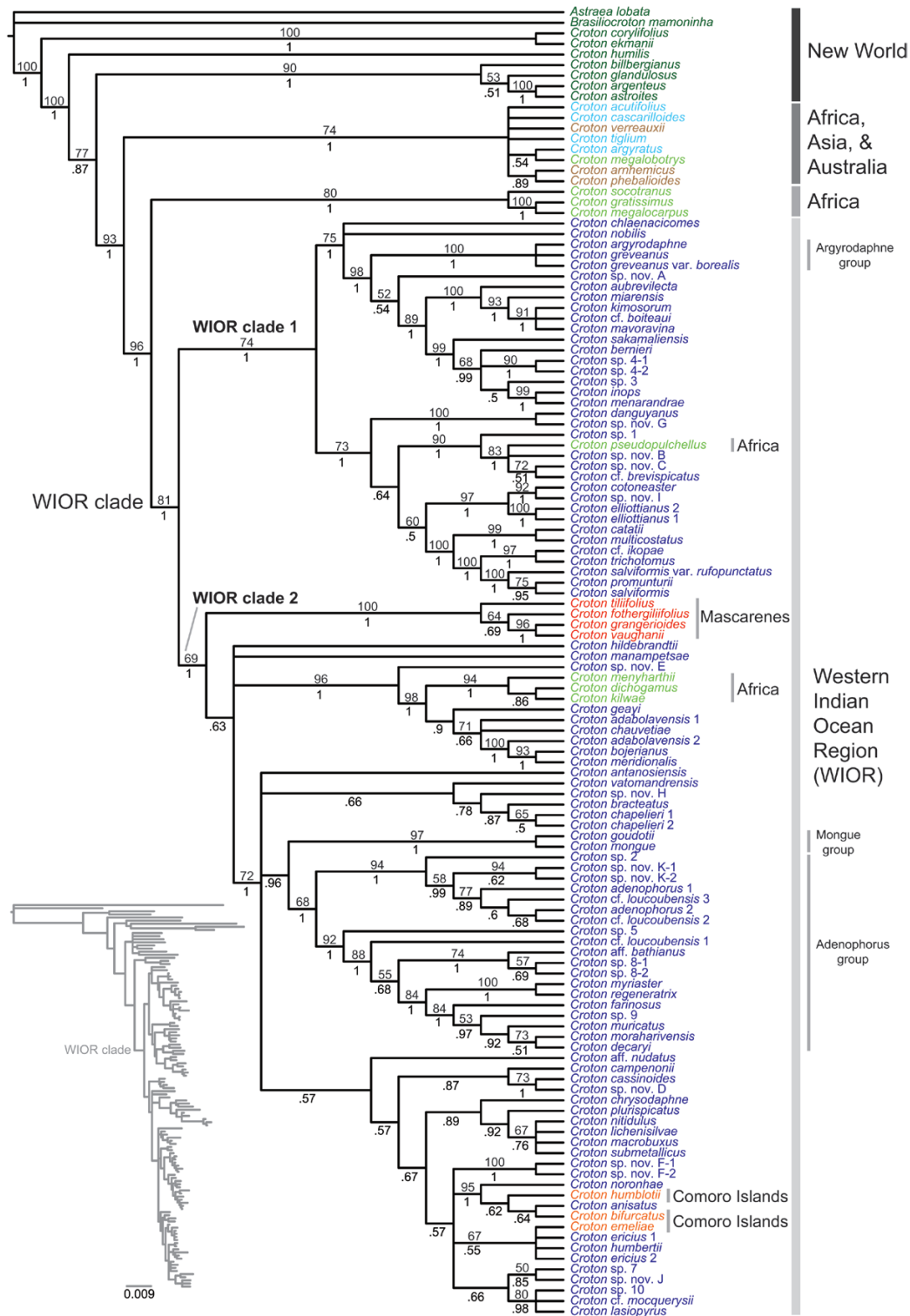
Analyses of the combined ITS and plastid DNA data resulted in an overall increase in resolution and clade posterior probabilities and bootstrap support (Figs 1, S5). *Croton* spp. from Madagascar and neighbouring islands were all part of a clade (WIOR clade; pp 1; bs 81%) that is sister to a clade of African and Socotran species (*Croton gratissimus* Burch., *Croton megalocarpus* Hutch. and *Croton socotranus* Balf.f.), and

this was in turn sister to a clade comprising African (*C. megalobotrys* Müll.Arg.), Asian (*Croton acutifolius* Esser, *Croton argyratus* Blume and *Croton cascarilloides* Raeusch.) and Australian (*Croton arnhemicus* Müll.Arg., *Croton phebalioides* Müll.Arg. and *Croton verreauxii* Baill.) species. A small clade of African species was found to be nested in the WIOR clade (*Croton dichogamus* Pax, *Croton kilwae* Radcl.-Sm. and *Croton menyhartii* Pax); another African species (*Croton pseudopulchellus* Pax) was also embedded in the WIOR clade. The species sampled from the Comoro Islands (*Croton bifurcatus* Baill., *Croton emeliae* Baill. and *Croton humblotii* Baill.) and the Mascarene Islands (*Croton fothergiliifolius* Baill., *Croton grangerioides* Bojer ex Baill., *Croton tiliifolius* Lam. and *Croton vaughanii* Croizat) were also nested in the WIOR clade, the Mascarene species being monophyletic (pp 1; bs 100%), whereas the accessions from the Comoro Islands are part of a poorly resolved group that includes *C. anisatus* Baill. from Madagascar.

Results of the ancestral area reconstruction are summarized in Figure 2. The ancestral area reconstruction unambiguously infers (100% probability) that the Old World species originated from the Americas. It also indicates a 41.2% probability that Madagascar was colonized from Africa, a 34.8% probability that Africa was colonized from Madagascar, a 12.9% probability that the WIOR clade originated from Asia and a 9.9% probability that it originated in the Americas (Fig. 2). In the WIOR clade, Madagascar is fully supported as the ancestral area for the Mascarene clade. Madagascar is also the most probable origin for the African clade that includes *C. dichogamus*, *C. kilwae* and *C. menyhartii* and for the African species *C. pseudopulchellus*. Similarly, irrespective of the low phylogenetic resolution of the relationships of the species sampled from the Comoro Islands, Madagascar is inferred as the only probable ancestral area for the species present in the Comoro Islands.

DISCUSSION

In the Old World clade of *Croton* (*Croton* subgenus *Croton*), we found that WIOR species are all part of a monophyletic group that forms a sister group to a clade of two African and one Socotran species. In the WIOR clade, one widespread African species, *C. pseudopulchellus*, is nested in one of the two main branches of the clade (WIOR clade 1, see Fig. 1) and another small clade of three African species (*C. dichogamus*, *C. kilwae* and *C. menyhartii*) is nested in WIOR clade 2 (Fig. 1). *Croton pseudopulchellus* is morphologically similar to a number of shrubby, Malagasy species with medium-sized, subverticillate leaves (glabrous above and silvery lepidote below) and congested, clustered

Figure 1. Majority-rule consensus tree from the Bayesian MCMC inference of the combined ITS and plastid data for *Croton* WIOR spp. Bootstrap values from ML analyses (1000 replicates) are indicated above the branches, and posterior probabilities are given below the branches. The geographical distribution of clades is indicated by the vertical bars to the right, as are the species groups that are discussed in the article. Taxa in dark green are from the New World, in light blue from Asia, in light brown from Australia, in light green from Africa and Socotra, in dark blue from Madagascar, in red from the Mascarene Islands and in orange from the Comoro Islands. The inset to the lower left shows the corresponding phylogram.
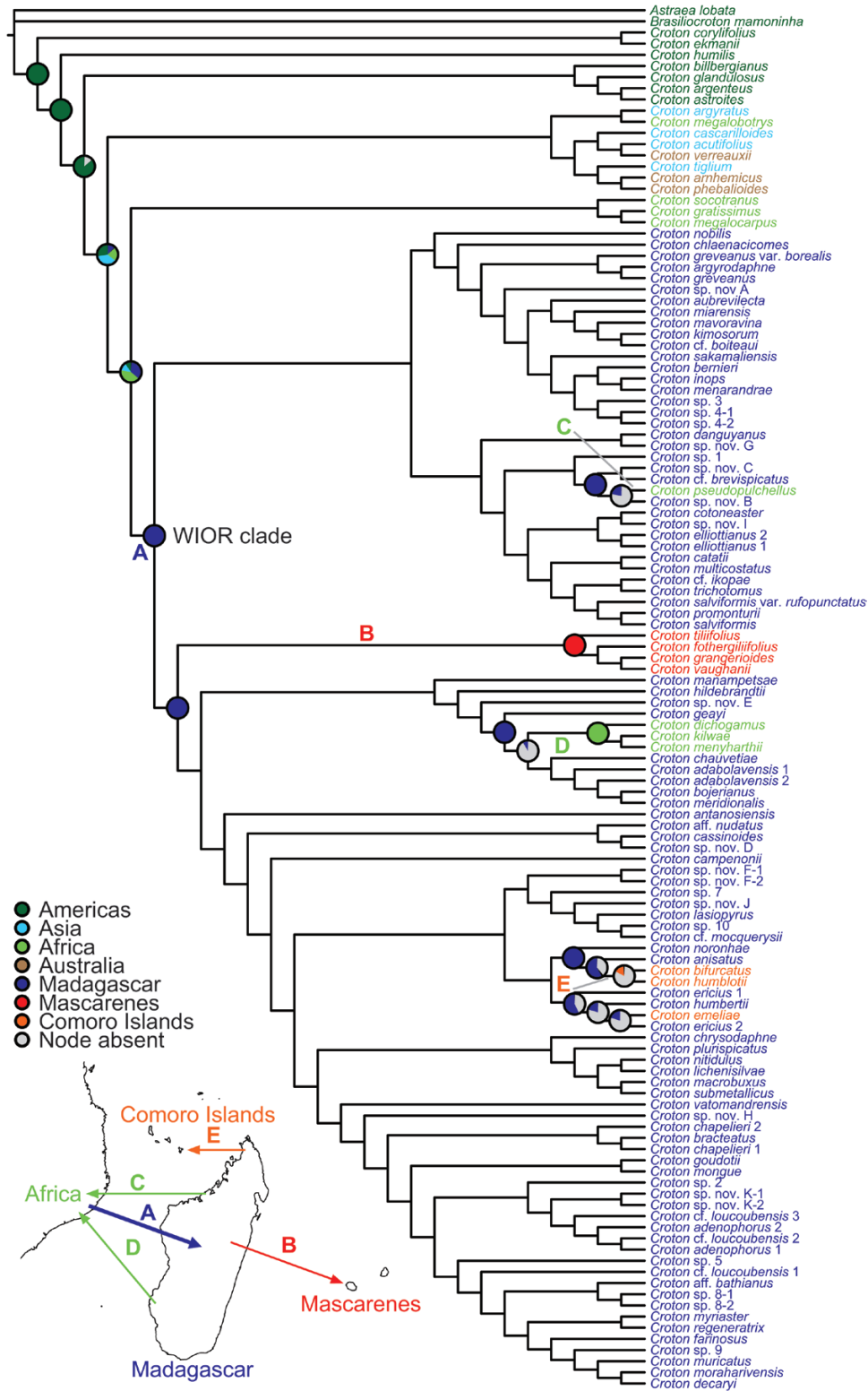
Figure 2. Ancestral state reconstruction of the geographical distribution of *Croton* in the WIOR for the most relevant nodes of the ML tree. Pie charts summarize the probability for an area as being ancestral at a given node and the prevalence of that node in the tree sample. The inset map summarizes the most probable biogeographic scenarios of dispersal.

inflorescences, such as *C. brevispicatus* Baill. (Fig. 3A). The other clade differs in that all of its members have some stellate pubescence on the upper side of the leaves, but they are also shrubs with silvery lepidote indumentum on the lower leaf surface along with some scattered brown scales, and they have longer inflorescences. This is also a common syndrome among Malagasy *Croton* spp., e.g. in species such as *Croton bojerianus* Baill. (Fig. 3D). The majority of the other *c*. 50 African *Croton* spp. are trees, whereas shrubs are the predominant life form of *Croton* in Madagascar. Both *C. dichogamus* and *C. kilwae* are restricted to East Africa from Tanzania to Mozambique, whereas *C. menyhartii* also occurs there, but extends farther south into South Africa, west to Angola and north to

Ethiopia. Van Ee *et al.* (2015) also found the same species to be nested in a largely Madagascan clade, but with their different taxon sampling (only 16 species sampled from Madagascar) they did not distinguish them as belonging to distinct subclades. The study of Van Ee *et al.* (2015) focused mainly on the Australasian *Croton* spp., covering 22 of the 29 species from Australia and 20 of the 29 species native to Thailand, with an additional 11 Asian species and 15 from Africa and Arabia sampled beyond the four species that are embedded in the WIOR clade. The geographical origin of the taxon sampling in that study was an inverse proportion of WIOR to non-WIOR Old World species compared to this study. Nonetheless, with 19 of the 54 native African and Arabian species and 49 of the *c*.

Figure 3. *Croton* spp. in the WIOR clade. (A) *C. brevispicatus* – this Malagasy species resembles the African *C. pseudo-pulchellus*; (B, C) *C. argyrodaphne*; (B) branch with staminate flowers with nine to 11 stamens; (C) pistillate flower – note the valvate sepals and stipitate style with highly divided stigmatic tips and sparsely lepidote ovary; (D) *C. bojerianus* – this Malagasy species resembles the African *C. dichogamus*; (E, F) *C. nobilis*; (E) underside of leaves (silvery-lepidote with scattered brown scales) and staminate flowers with *c*. 30 stamens; (F) pistillate flowers – note the reduplicate-valvate sepals and non-stipitate styles with *c*. 12 stigmatic tips and densely stellate-pubescent ovary (Image credits: A–B: K. Kainulainen; D: B. Van Ee; C, E–F: P. Berry).

185 native Asian-Pacific *Croton* spp. sampled to date, there is still ample possibility that more extra-regional species could turn up nested in the WIOR clade. It is also conceivable that some of the remaining unsampled Madagascan taxa could emerge outside the WIOR clade, but our sampling already includes most of the morphological, ecological and geographical diversity of the genus in the WIOR. In the near future, we will be working to obtain a more thorough sampling of African and Asian-Pacific species so that we can more rigorously examine the biogeographic history of the genus in the Old World.

TAXONOMIC GROUPINGS OF MADAGASCAN TAXA

The phylogenetic hypothesis inferred from the combined data recovers several well-supported internal clades of Malagasy *Croton*. However, characterizing these clades morphologically will require more detailed taxonomic work and denser molecular sampling. Leandri (1939) relied on characters such as high vs. low stamen number, lepidote vs. stellate pubescence and leaf size and shape to delineate 18 species groups that he recognized for Malagasy *Croton* spp., but we are largely not able to match his morphologically defined species groups to the molecular phylogenetic trees. There are a few exceptions, however. *Croton goudotii* and *C. mongue* form a strongly supported clade and they both belong to Leandri's 'Groupe Mongue', in which he also included two other species, one of which (*C. oreades* Baill.) we now consider to be a synonym of *C. mongue* and the other, *Croton tsaratananae* Leandri, may prove to be a synonym of *C. goudotii*. This is a distinctive group of species that consists of large trees with alternate, palmately veined leaves, stellate pubescence and pistillate flowers with well-developed petals (Fig. 4A–C). Another of Leandri's groups that is well supported by our data is his 'Groupe Adenophorum' (Fig. 1, as Adenophorus Group), which corresponds closely to Webster's (1993) concept of *C.* section *Anisophyllum*. Species of this group are characterized by having opposite leaves, stellate pubescence and glands in the axils of the secondary veins on the abaxial surface of the leaves (Fig. 4D–F). They also often have conspicuous stipules and pseudoterminal inflorescences of tightly packed flowers. This is a much larger group (16 species recognized by Leandri, 1939) and, as shown by the number of still unidentified species, one much in need of taxonomic revision.

One of Leandri's groups that is partly supported by our results is his 'Groupe Argyrodaphne', which he characterized as having large, penninerved leaves with shiny, lepidote undersides and staminate flowers with <12 stamens. We obtained high support (pp 1, bs 100) for a clade of three taxa placed in this group, namely *C. argyrodaphne* Baill., *Croton greveanus* Baill. and *C. greveanus* var. *borealis* Leandri (Figs 1, 3), but other species that Leandri placed in this group, such as *Croton catatii* Baill. and *Croton antanosiensis* Leandri, emerge in entirely different clades in the WIOR (*C. catatii* in WIOR clade 1 and *C. antanosiensis* in WIOR clade 2).

An example of one of Leandri's groups that is not at all supported by the molecular data is his 'Groupe Nobile', which like 'Groupe Argyrodaphne' discussed above are usually trees with large, penninerved leaves with lepidote undersides, but with staminate flowers that have >12 stamens. Species from this group that we sampled include *Croton nobilis* Baill. (Fig. 3E–F), *Croton chrysodaphne* Baill. (synonym = *C. lepidotus* A.DC.), *Croton chlaenacicomes* Leandri, *Croton multicostatus* Müll.Arg. (synonym = *Croton vernicosus* Baker), *Croton trichotomus* Lam. and *C. hildebrandtii* Baill. None of these species was sister to any of the others in this group in our results, with four of them belonging to WIOR clade 1 and two belonging to WIOR clade 2 (Fig. 1). Notably, *C. multicostatus* forms a strongly supported clade (pp 1, bs 99) with *C. catatii*, which was placed by Leandri (1939) in Groupe Argyrodaphne. We have provided evidence in other papers (Berry & Van Ee, 2011; Berry *et al.*, 2011, 2016a) that these large, lepidote-leaved *Croton* spp. from Madagascar only superficially resemble each other in their vegetative characters; when details, particularly of the pistillate flowers and inflorescences, are examined, they are surprisingly diverse (Fig. 3C, F). Similarly, the WIOR-endemic sections of lepidote-leaved species in the classification scheme of Webster (1993), namely *C.* sections *Monguia* and *Decapetalon*, show no correspondence to any of the clades obtained in our results.

BIOGEOGRAPHIC SCENARIOS

In a review of the historic biogeography of Madagascar, Yoder & Nowak (2006) suggested that a large majority of taxa on Madagascar originated by dispersal out of Africa during the Cenozoic. There is at least partial support for this hypothesis for *Croton* in our results, which show the WIOR clade as sister to a mainly East African clade of *Croton*. Since Madagascar has been isolated from Africa and other continents for >80 million years (Rabinowitz, Coffin & Falvey, 1983; Storey *et al.*, 1995) and the estimated age of *Croton* is *c*. 55 million years (Van Ee *et al.*, 2008, with no error margins reported), a biogeographic scenario of vicariance is improbable for this group. Instead, we hypothesize a dispersal event from Africa to Madagascar, e.g. via vegetation rafts transported by historical ocean currents to Madagascar. Ali & Huber (2010) inferred the ocean surface currents as travelling from north-eastern
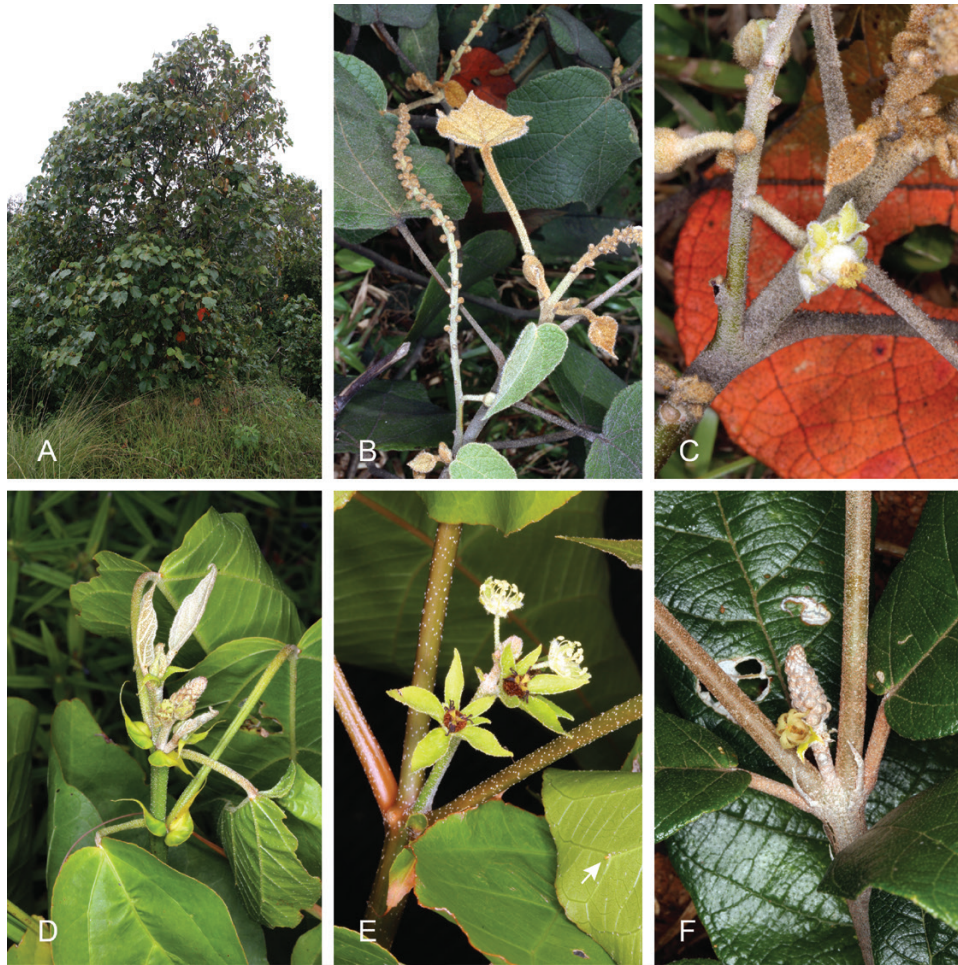
Figure 4. Madagascan *Croton* spp. from the Mongue and Adenophorus groups. (A–C) *C. goudotii*; (A) tree habit; (B) palmate leaf venation with lobed blade and elongate inflorescence; (C) pistillate flower with well-developed petals; (D–E) *C. adeno-phorus*; (D) branch showing opposite leaves, conspicuous stipules and tightly packed axillary inflorescence; (E) flowering branch with open pistillate flowers (the lower two with unequal green sepals) and staminate flowers (the distal two) – note the stalked gland at the junction of two secondary veins on the lower surface of the leaf blade (arrow); (F) *Croton farinosus*; flowering branch – note the opposite leaves, conspicuous stipules and tightly packed inflorescence (Image credits: P. Berry).

Mozambique and Tanzania towards Madagascar during the Palaeogene (66–23 million years ago), which is consistent with the putative age of *Croton*.

Several studies have found Madagascar to be the ancestral area of plant lineages in the neighbouring Comoro Islands, Mascarene Islands, mainland Africa, India and Southeast Asia (Yuan *et al.*, 2005; Janssen *et al.*, 2008; Janssens *et al.*, 2009; Le Péchon *et al.*, 2010; Wikström *et al.*, 2010; Bacon *et al.*, 2016). In this study, we find multiple instances of *Croton* spp. with distributions outside of Madagascar nested in Malagasy clades (Figs 1, 2), indicating that several independent out-of-Madagascar dispersal events have occurred to the Comoro Islands, the Mascarene Islands and to mainland Africa (supported by the ancestral state reconstruction in Fig. 2). Contemporary ocean currents,

which generally flow from Madagascar south-east-wards towards Africa, with several large gyres (Schott, Xie & McCreary, 2009; Ali & Huber, 2010), could have facilitated these dispersals from Madagascar to mainland Africa and the Comoro Islands.

The Comoro Islands are a volcanic archipelago located 350–600 km west of the north-western tip of Madagascar, roughly halfway between Madagascar and mainland Africa. The oldest island there is Mayotte, which emerged above sea level *c*. 8 Mya (Nougier, Cantagrel & Karche, 1986). Five *Croton* spp. are known from these islands, four of which are endemic: *C. emeliae*, *C. bifurcatus*, *C. humblotii* and *C. tulasnei* Baill. (the first three are sampled in this study). The fifth species, *C. adenophorus* Baill., also occurs on Madagascar. The inferred positions of these

species in the WIOR clade are inconclusive as to the number of colonizations of the Comoro Islands from Madagascar. *Croton bifurcatus* and *C. humblotii* probably represent one introduction and *C. emeliae* could also be a result of that event. *Croton adenophorus* is reported from the Comoro Islands, but so far only accessions from Madagascar have been sampled for molecular analyses. If the Comoros populations of this species are monophyletic with the Madagascan populations, this would probably represent a distinct dispersal event from Madagascar, since *C. adenophorus* is part of the strongly supported Adenophorus Group. *Croton tulasnei* (unsampled) is morphologically similar to *C. adenophorus*, and we hypothesize they may be sister species, so it may have originated from a common ancestor with *C. adenophorus*, either on the Comoros or on Madagascar.

The Mascarene Islands are another group of volcanic islands located 800–1500 km east of Madagascar. Mauritius emerged between 7 and 8 Mya (Duncan, 1990), whereas Réunion is younger, at *c*. 2 Mya (McDougall, 1971). Our sampling included four of the six *Croton* spp. found in this archipelago, all of which are endemic to the islands (five on Mauritius and one on Réunion). The four species sampled, *C. fothergiliifolius*, *C. grangerioides*, *C. tiliifolius* and *C. vaughanii*, form a well-supported clade nested in the WIOR clade (Fig. 1), which the ancestral state reconstruction supports as a single dispersal event from Madagascar to the Mascarene Islands and subsequent speciation there (Fig. 2). A similar pattern of a single, well-supported Mascarene clade nested in a larger Madagascan clade has also been shown in *Gastonia* Comm. ex Lam and *Polyscias* J.R.Forst. & G.Forst. in Araliaceae (Plunkett, Lowry & Vu, 2004), *Pyrostria* Comm. ex Juss. in Rubiaceae (Wikström *et al.*, 2010), *Badula* Juss. in Primulaceae (Bone *et al.*, 2012) and *Sideroxylon* L. in Sapotaceae (Stride, Nylinder & Swenson, 2014).

Overall, our results indicate four out-of-Madagascar dispersal events in *Croton* (Fig. 2), with additional similar dispersal events possible. This shows the importance of Madagascar as a source area for colonizations, not only to other islands in the western Indian Ocean but also back to Eastern Africa.

The large number of unidentified or undescribed species in our sampling reflects both the difficulty of identifying specimens of *Croton* from Madagascar and the high number of taxonomic novelties that remain to be published. We are confident that at least 11 of the unnamed specimens are undescribed species and they will be published shortly. At that time, the new species will be cross-referenced to their number and position in the phylogenetic trees presented here. Our ongoing molecular studies on Old World *Croton* aim to increase the density of species sampling substantially and to increase the number of molecular markers. This will

hopefully allow us to better evaluate species relationships among the Madagascan species and determine if we can produce a more realistic classification of the >150 *Croton* spp. that have proliferated on the island.

REFERENCES

Ali JR, Huber M. 2010. Mammalian biodiversity on Madagascar controlled by ocean currents. *Nature* **463:** 653–656.

Bacon CD, Simmons MP, Archer RH, Zhao LC, Andriantiana J. 2016. Biogeography of the Malagasy Celastraceae: multiple independent origins followed by widespread dispersal of genera from Madagascar. *Molecular Phylogenetics and Evolution* **94:** 365–382.

Baillon H. 1858. *Étude générale du groupe des Euphorbiacées*. Paris: Librarie de Victor Masson.

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.

Berry PE, Van Ee BW. 2011. Recognition of *Croton multicostatus* Müll.Arg. (Euphorbiaceae) as native to Madagascar. *Candollea* **66:** 109–112.

Berry PE, Van Ee BW, Haber EA. 2011. Resolving the tangled identity of *Croton chrysodaphne* Baill. (Euphorbiaceae) in Madagascar. *Candollea* **66:** 113–118.

Berry PE, Van Ee BW, Kainulainen K, Achtemeier L. 2016a. *Croton cupreolepis* (Euphorbiaceae), a new coppery-lepidote tree species from eastern Madagascar. *Systematic Botany* **41:** 977–982.

Berry PE, Van Ee BW, Kainulainen K, Razafindraibe H. 2016b. *Croton aleuritoides* P.E. Berry (Euphorbiaceae), a

distinctive new tree species from Montagne des Français in northern Madagascar. *Candollea* **71:** 181–188.

Bone RE, Strijk JS, Fritsch PW, Buerki S, Strasberg D, Thébaud C, Hodkinson TR. 2012. Phylogenetic inference of *Badula* (Primulaceae), a rare and threatened genus endemic to the Mascarene Archipelago. *Botanical Journal of the Linnean Society* **169:** 284–296.

Chernomor O, von Haeseler A, Minh BQ. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. *Systematic Biology* **65:** 997–1008.

Dumetz N. 1999. High plant diversity of lowland rainforest vestiges in eastern Madagascar. *Biodiversity and Conservation* **8:** 273–315.

Duncan RA. 1990. The volcanic record of the Réunion hotspot. *Proceedings of the Ocean Drilling Program, Scientific Results* **115:** 3–10.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32:** 1792–1797.

Frodin DG. 2004. History and concepts of big plant genera. *Taxon* **53:** 753–776.

Ganzhorn JU, Lowry PPI, Schatz GE, Sommer S. 2001. The biodiversity of Madagascar: one of the world's hottest hotspots on its way out. *Oryx* **35:** 346–348.

Govaerts R, Frodin DG, Radcliffe-Smith A. 2000. *World checklist and bibliography of Euphorbiaceae*, 4 vols. Kew: Royal Botanic Garden.

Grubb PJ. 2003. Interpreting some outstanding features of the flora and vegetation of Madagascar. *Perspectives in Plant Ecology, Evolution and Systematics* **6:** 125–146.

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

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41:** 95–98.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17:** 754–755.

Janssen T, Bystriakova N, Rakotondrainibe F, Coomes D, Labat JN, Schneider H. 2008. Neoendemism in Madagascan scaly tree ferns results from recent, coincident diversification bursts. *Evolution* **62:** 1876–1889.

Janssens SB, Knox EB, Huysmans S, Smets EF, Merckx VS. 2009. Rapid radiation of *Impatiens* (Balsaminaceae) during Pliocene and Pleistocene: result of a global climate change. *Molecular Phylogenetics and Evolution* **52:** 806–824.

Kainulainen K, van Ee B, Antilahimena P, Razafindraibe H, Berry PE. 2016. New species and species reports of *Croton* L. (Euphorbiaceae) from the eastern forest corridor of Madagascar. *Candollea* **71:** 327–356.

Klanderud K, Mbolatiana HZH, Vololomboahangy MN, Radimbison MA, Roger E, Totland Ø, Rajeriarison C. 2010. Recovery of plant species richness and composition after slash-and-burn agriculture in a tropical rainforest in Madagascar. *Biodiversity and Conservation* **19:** 187–204.

Le Péchon T, Dubuisson JY, Haevermans T, Cruaud C, Couloux A, Gigord LD. 2010. Multiple colonizations from Madagascar and converged acquisition of dioecy in the Mascarene Dombeyoideae (Malvaceae) as inferred from chloroplast and nuclear DNA sequence analyses. *Annals of Botany* **106:** 343–357.

Leandri MJ. 1939. Les *Croton* de Madagascar et des îles voisines. *Annales du Musée Colonial de Marseille.* **7:** 5–100.

Leandri MJ. 1948. Euphorbiacées récoltées à Madagascar et aux Mascareignes par MM. Lam et Meeuse en 1938. *Notulae Systematicae* **13:** 182–187.

Leandri MJ. 1969. Sur quelques espèces malgaches de *Croton* (Euphorbiacées). *Adansonia* **9:** 497–510.

Leandri MJ. 1970a. Quatre *Croton* nouveaux du sud de Madagascar (Euphorbiacées). *Adansonia* **10:** 183–190.

Leandri MJ. 1970b. *Croton* malgaches. *Adansonia* **10:** 309–314.

Leandri MJ. 1972a. *Croton* nouveax de l'ouest de Madagascar (Euphorbiacées). *Adansonia* **12:** 65–71.

Leandri MJ. 1972b. Contribution a l'étude des *Croton* malgaches à grandes feuilles argentées. *Adansonia* **12:** 403–408.

Leandri MJ. 1973a. Quelques *Croton* malgaches nouveaux. *Adansonia* **13:** 173–176.

Leandri MJ. 1973b. Deux espèces et une varieté malgache de *Croton*. *Adansonia* **13:** 295–297.

Leandri MJ. 1973c. Un *Croton* nouveau de l'est de Madagascar. *Adansonia* **13:** 423–424.

Madagascar Catalogue. 2013. *Catalogue of the vascular plants of Madagascar*. St. Louis: Missouri Botanical Garden; Antananarivo: Missouri Botanical Garden, Madagascar Research and Conservation Program. Available at: http://www.efloras.org/madagascar

Maddison WP, Maddison DR. 2016. *Mesquite: a modular system for evolutionary analysis*. *Version 3.10*. Available at: http://mesquiteproject.org

McDougall I. 1971. The geochronology and evolution of the young volcanic island of Réunion, Indian Ocean. *Geochimica et Cosmochimica Acta* **35:** 261–270.

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

Müller Argovensis J. 1866. Euphorbiaceae. In: Candolle AP de, ed. *Prodromus systematis naturalis regni vegetabilis*, vol. 15. Paris, Strasbourg, London: Treuttel & Würtz, 189–1261.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32:** 268–274.

Nougier J, Cantagrel JM, Karche JP. 1986. The Comores archipelago in the western Indian Ocean: volcanology, geochronology, and geodynamic setting. *Journal of African Earth Sciences* **5:** 135–145.

Nylander JAA. 2004. *MrAIC.pl. Version 1.4.4. Program distributed by the author, Evolutionary Biology Centre, Uppsala University. Software.* Available at: https://github.com/nylander/MrAIC

Plunkett GM, Lowry PPI, Vu NV. 2004. Phylogenetic relationships among *Polyscias* (Araliaceae) and close relatives

from the western Indian Ocean Basin. *International Journal of Plant Sciences* **165**: 861–873.

Rabinowitz PD, Coffin MF, Falvey D. 1983. The separation of Madagascar and Africa. *Science* **220**: 67–69.

Radcliffe-Smith A. 2016. *The Genus* Croton *in Madagascar and the Comoro Is*. 213 pp. 'Preprint' deposited in the libraries of Royal Botanic Gardens, Kew; Natural History Museum, London; and Royal Horticultural Society, Wisley on December 23, 2016 by Martin Cheek.

Rambaut A. 2012. *FigTree. Version 1.4. Institute of Evolutionary Biology, University of Edinburgh. Software*. Available at: http://tree.bio.ed.ac.uk/software/figtree/

Rambaut A, Drummond A. 2007. *Tracer. Version 1.5. Institute of Evolutionary Biology, University of Edinburgh. Software*. Available at: http://tree.bio.ed.ac.uk/software/tracer/

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.

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

Samuel R, Kathriarachchi H, Hoffmann P, Barfuss MH, Wurdack KJ, Davis CC, Chase MW. 2005. Molecular phylogenetics of Phyllanthaceae: evidence from plastid *matK* and nuclear PHYC sequences. *American Journal of Botany* **92**: 132–141.

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

Schatz GE. 2001. *Generic tree flora of Madagascar*. Kew: Royal Botanic Gardens; St Louis: Missouri Botanical Garden.

Schott FA, Xie S-P, McCreary JP Jr. 2009. Indian Ocean circulation and climate variability. *Reviews of Geophysics* **47**: RG1002.

Storey M, Mahoney JJ, Saunders AD, Duncan RA, Kelley SP, Coffin MF. 1995. Timing of hot spot-related volcanism and the breakup of Madagascar and India. *Advancement of Science* **267**: 852–855.

Stride G, Nylinder S, Swenson U. 2014. Revisiting the biogeography of *Sideroxylon* (Sapotaceae) and an evaluation of the taxonomic status of *Argania* and *Spiniluma*. *Australian Systematic Botany* **27**: 104–118.

Sussman RW, Rakotozafy A. 1994. Plant diversity and structural analysis of a tropical dry forest in southwestern Madagascar. *Biotropica* **26**: 241–254.

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.

Tate JA, Simpson BB. 2003. Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. *Systematic Botany* **28**: 723–737.

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 DNA. *Systematic Botany* **25**: 539–565.

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

Van Ee BW, Forster PI, Berry PE. 2015. Phylogenetic relationships and a new sectional classification of *Croton* (Euphorbiaceae) in Australia. *Australian Systematic Botany* **28**: 219–233.

Van Ee BW, Jelinski N, Berry PE, Hipp AL. 2006. Phylogeny and biogeography of *Croton alabamensis* (Euphorbiaceae), a rare shrub from Texas and Alabama, using DNA sequence and AFLP data. *Molecular Ecology* **15**: 2735–2751.

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

Webster GL. 1993. A provisional synopsis of the sections of the genus *Croton* (Euphorbiaceae). *Taxon* **42**: 793–823.

White TJ, Bruyns T, Lee S, Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, eds. *PCR protocols: a guide to methods and applications*. San Diego: Academic Press, 315–324.

Wikström N, Avino M, Razafimandimbison SG, Bremer B. 2010. Historical biogeography of the coffee family (Rubiaceae, Gentianales) in Madagascar: case studies from the tribes Knoxieae, Naucleeae, Paederieae and Vanguerieae. *Journal of Biogeography* **37**: 1094–1113.

Wurdack KJ, Hoffmann P, Chase MW. 2005. Molecular phylogenetic analysis of uniovulate Euphorbiaceae (Euphorbiaceae sensu stricto) using plastid *rbcL* and *trnL-F* DNA sequences. *American Journal of Botany* **92**: 1397–1420.

Yang Z, Rannala B. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. *Molecular Biology and Evolution* **14**: 717–724.

Yoder AD, Nowak MD. 2006. Has vicariance or dispersal been the predominant biogeographic force in Madagascar? Only time will tell. *Annual Review of Ecology, Evolution, and Systematics* **37**: 405–431.

Yuan YM, Wohlhauser S, Möller M, Klackenberg J, Callmander M, Küpfer P. 2005. Phylogeny and biogeography of *Exacum* (Gentianaceae): a disjunctive distribution in the Indian ocean basin resulted from long distance dispersal and extensive radiation. *Systematic Biology* **54**: 21–34.

SUPPORTING INFORMATION

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

Figure S1. Majority-rule consensus tree from the Bayesian MCMC inference of the ITS data. Numbers below the branches represent Bayesian posterior probabilities support values. Corresponding bootstrap support values from the ML analyses (1000 replicates) are given above the branches.

Figure S2. Majority-rule consensus tree from the Bayesian MCMC inference of the combined plastid data. Numbers below the branches represent Bayesian posterior probabilities support values. Corresponding bootstrap support values from the ML analyses (1000 replicates) are given above the branches.

Figure S3. ML tree from the ML analysis of the ITS data with bootstrap support (1000 replicates) for nodes indicated.

Figure S4. ML tree from the ML analysis of the combined plastid data with bootstrap support (1000 replicates) for nodes indicated.

Figure S5. ML tree from the ML analysis of the combined ITS and plastid data with bootstrap support (1000 replicates) for nodes indicated.

Table S1. List of the taxa and accessions sampled, including GenBank accession numbers and voucher specimens. Sequences newly generated for this study begin with KY-. All herbarium specimens have been entered into the Tropicos database (www.tropicos.org), and updates to their determinations will be made there.