Annals of Botany **132**: 293–318, 2023
https://doi.org/10.1093/aob/mcad091, available online at www.academic.oup.com/aob

ANNALS OF BOTANY
Nonprofit since 1887

A comprehensive genus-level phylogeny and biogeographical history of the Lythraceae based on whole plastome sequences

Peter W. Inglis[1,*], Taciana B. Cavalcanti[1], Marlon G. Facco[2], Freek T. Bakker[3] and Shirley A. Graham[4]

[1]*Embrapa Recursos Genéticos e Biotecnologia, Parque Estação Biológica, Av. W5 Norte (final), Caixa Postal 02372 – Brasília, DF – CEP 70770-917, Brazil,* [2]*Departamento de Botânica, Instituto de Biociências, Universidade de São Paulo, Brazil,* [3]*Biosystematics Group, Wageningen University & Research, Postbus 647, NL-6700 AP, Wageningen, The Netherlands, and* [4]*Missouri Botanical Garden, 4344 Shaw Boulevard, Saint Louis, MO 63110, USA*
[*]*For correspondence. E-mail peterwinglis@gmail.com*

Received: 9 May 2023 Editorial decision: 3 July 2023 Accepted: 5 July 2023

• **Background and Aims** The Lythraceae are a mainly subtropical to tropical family of the order Myrtales with 28 currently accepted genera and approximately 600 species. There is currently no well-supported phylogenetic and biogeographical hypothesis of the Lythraceae incorporating all currently accepted genera, which we sought to provide.
• **Methods** Plastomes of representative species of 18 distinct Lythraceae genera were sequenced and annotated. Together with existing sequences, plastomes of all 28 currently accepted genera in the Lythraceae were brought together for the first time. The plastomes were aligned and a Bayesian phylogenetic hypothesis was produced. We then conducted a time-calibrated Bayesian analysis and a biogeographical analysis.
• **Key Results** Plastome-based Bayesian and maximum-likelihood phylogenetic trees are generally congruent with recent nuclear phylogenomic data and resolve two deeply branching major clades in the Lythraceae. One major clade concentrates shrubby and arboreal South American and African genera that inhabit seasonally dry environments, with larger, often winged seeds, adapted to dispersal by the wind. The second major clade concentrates North American, Asian, African and several near-cosmopolitan herbaceous, shrubby and arboreal genera, often inhabiting humid or aquatic environments, with smaller seeds possessing structures that facilitate dispersal by water.
• **Conclusions** We hypothesize that the Lythraceae dispersed early in the Late Cretaceous from South American to North American continents, with subsequent expansion in the Late Cretaceous of a North American lineage through Laurasia to Africa via a boreotropical route. Two later expansions of South American clades to Africa in the Palaeocene and Eocene, respectively, are also hypothesized. Transoceanic dispersal in the family is possibly facilitated by adaptations to aquatic environments that are common to many extant genera of the Lythraceae, where long-distance dispersal and vicariance may be invoked to explain several remarkable disjunct distributions in Lythraceae clades.

Key words: Lythraceae, plastome, phylogeny, biogeography, dispersal.

INTRODUCTION

The Lythraceae are a moderate-sized, mainly subtropical to tropical, cosmopolitan family of the order Myrtales with 28 currently accepted genera and approximately 600 species (Cavalcanti *et al*., 2022; POWO, 2023). The genera are annual or perennial herbs, shrubs or small to large trees, often hydrophilic, occurring in diverse habitats, such as ponds, swamps, rivers, mangroves, sandy oceanic beaches, savannah, semi-desert, and tropical dry or wet forests (Graham, 2007). Diagnostic traits include opposite entire leaves, flowers generally actinomorphic, perigynous, 4–16-merous, cup-shaped or tubular, sepals valvate, petals crumpled and caducous, stamens in two whorls and of two lengths, style one, ovary superior, rarely semi-inferior, fruits mostly capsular, seeds numerous, typically very small, and without noticeable endosperm (Graham *et al*., 1993; Graham, 2007; Graham and Graham, 2014). Among the most well-known genera are the temperate heterostylous genus *Lythrum* L. (Fig. 1: Plate 3E); the Crape Myrtle, *Lagerstroemia* L. (Fig. 1: Plate 3C), a popular horticultural shrub of warm climates; *Lawsonia* L., the source of henna; and the pomegranate, *Punica* L. (Fig. 1: Plate 4E). The herbaceous perennial *Cuphea* P. Browne (Fig. 1: Plate 1F), with bilateral, spurred flowers, is the largest genus with ~256 accepted species (POWO, 2023).

Lythraceae are today found on all major continents except Antarctica and are predominantly subtropical to tropical in distribution. Fifteen genera occur in the Old World, of which eight are exclusive (*Capuronia* Lourteig, *Duabanga* Buch.-Ham, *Galpinia* N.E.Br., *Koehneria* S.A.Graham, Tobe & Baas, *Pemphis* J.R.Forst & G.Forst, *Sonneratia* L.f., *Tetrataxis* Hook.f. and *Woodfordia* Salisb.), three are native to this region but widely cultivated around the world (*Lagerstroemia*, *Lawsonia* and *Punica*), and one is invasive to North America (*Trapa* L.). In the New World, of the 16 genera recorded, 13 are exclusive (*Adenaria* Kunth, *Cuphea*, *Decodon* J.F.Gmel, *Didiplis* Jacq., *Diplusodon* Pohl, *Ginoria* Jacq., *Gyrosphragma*

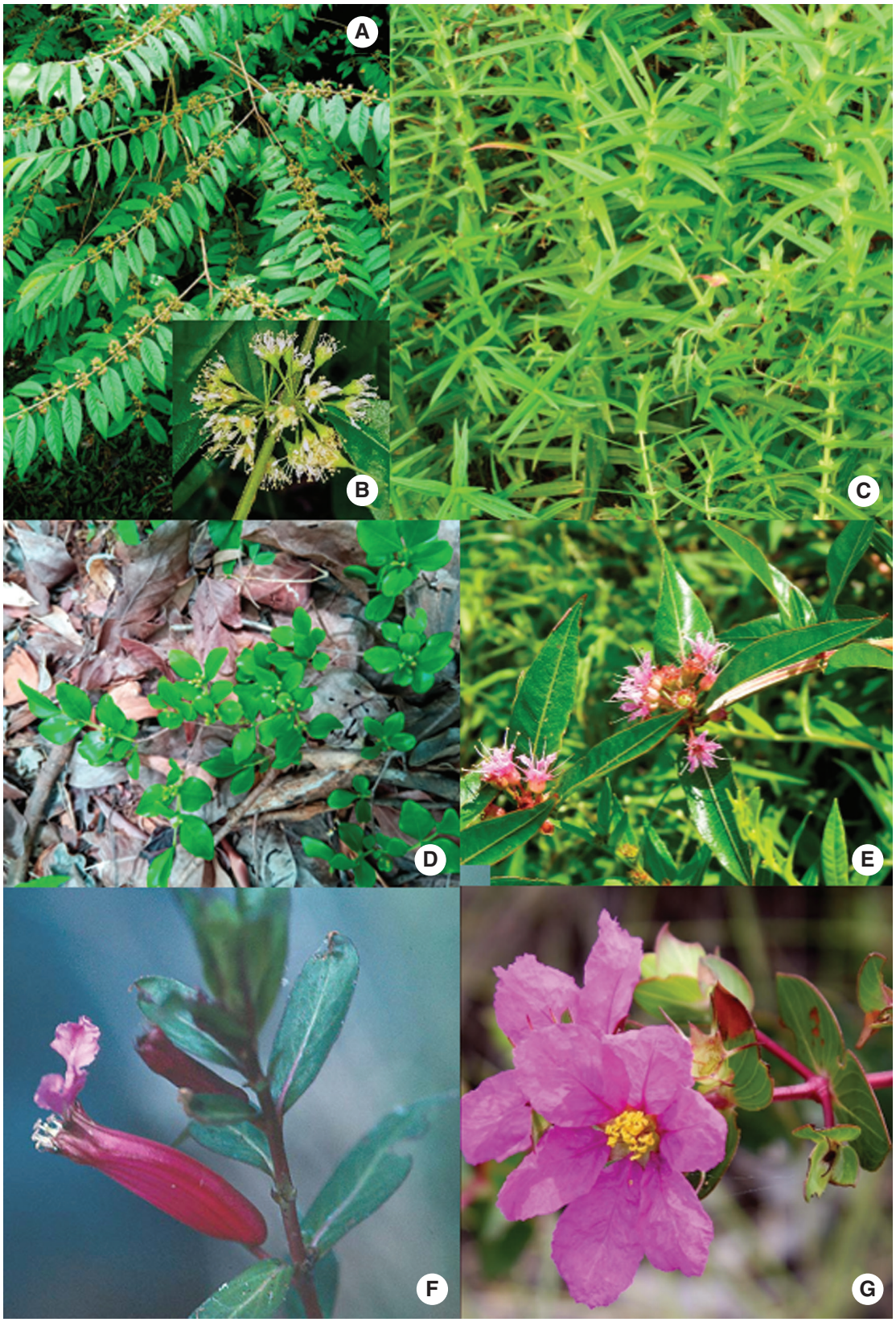

FIG. 1. Continued

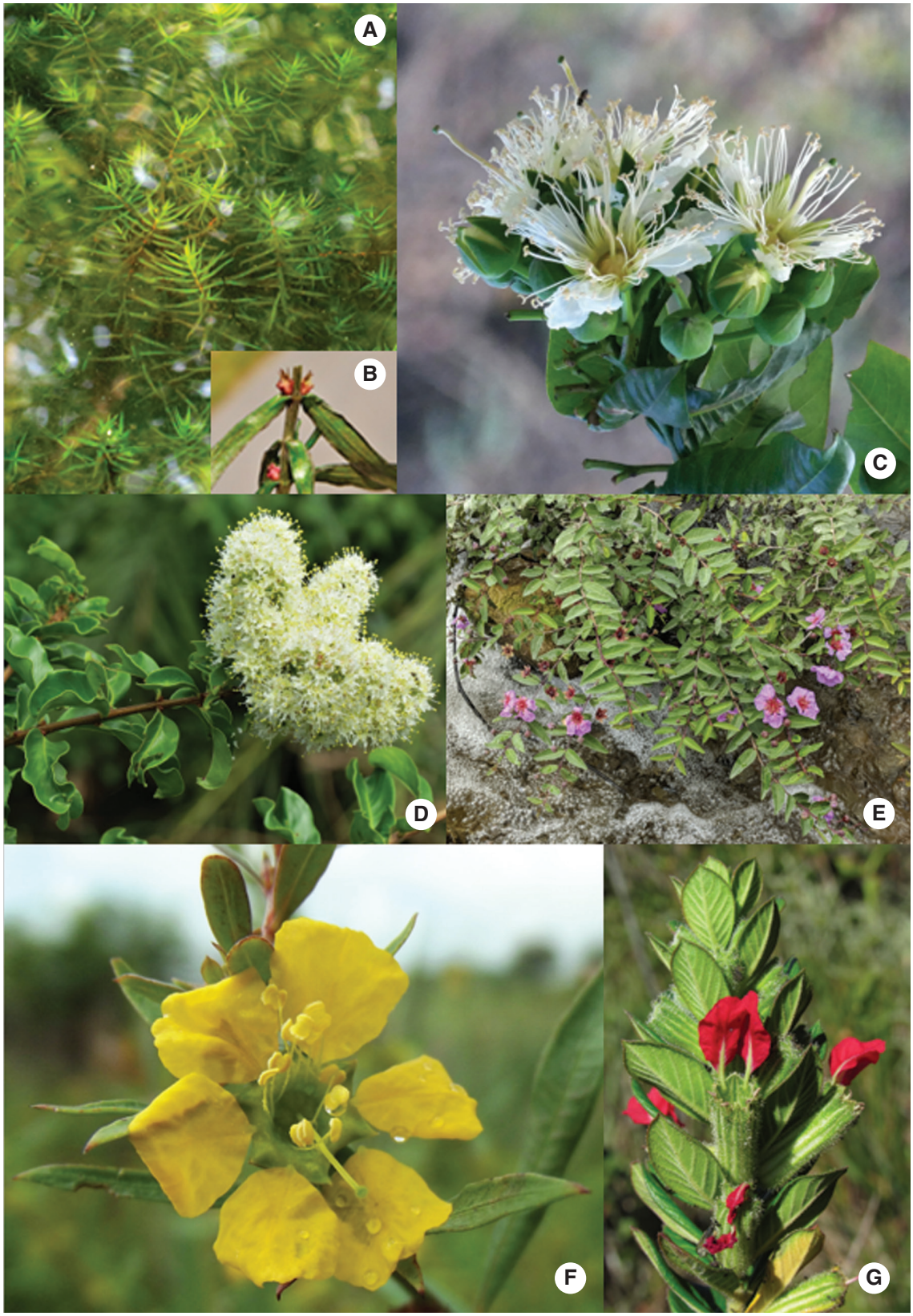

FIG. 1. Continued

FIG. 1. Continued

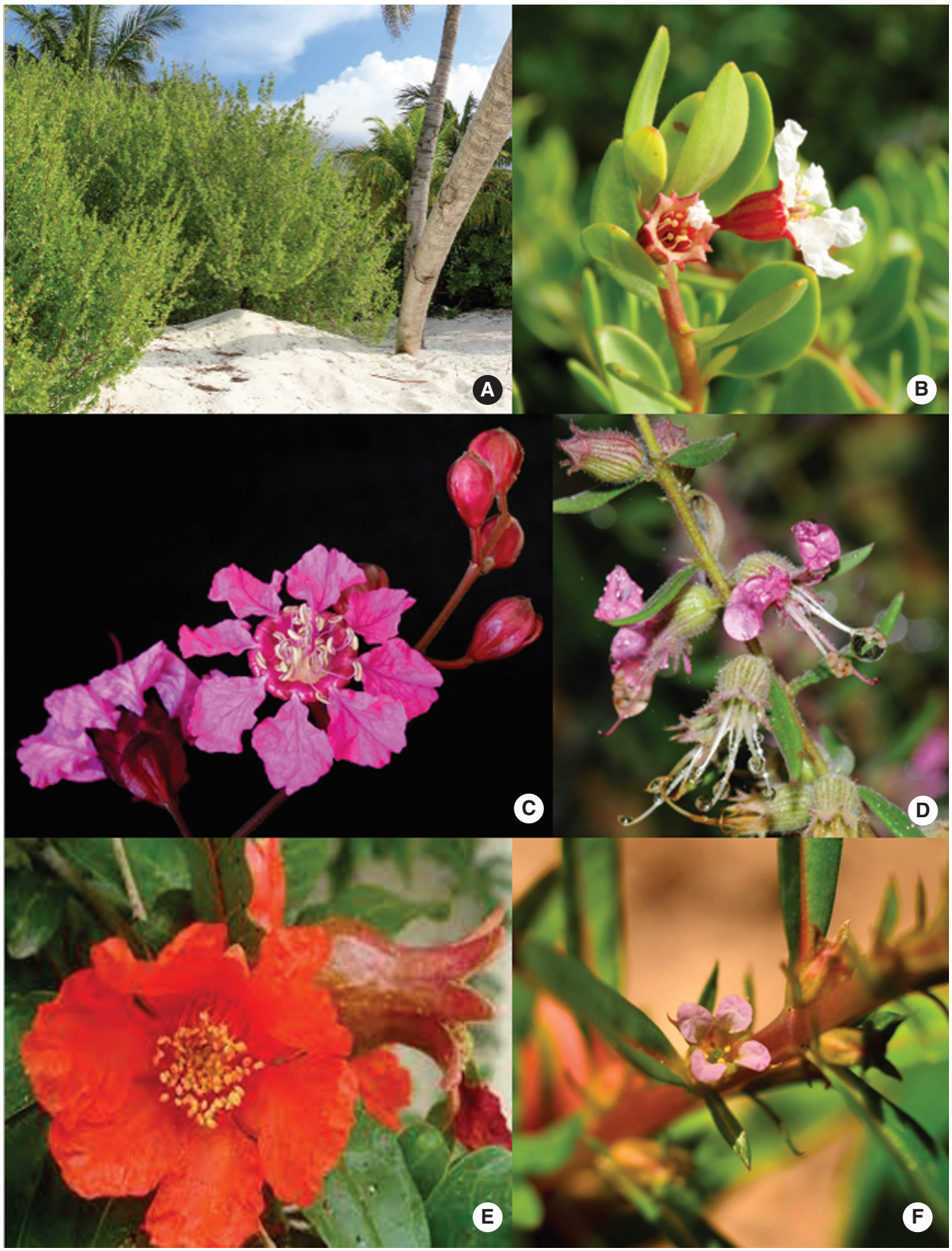

FIG. 1. Continued

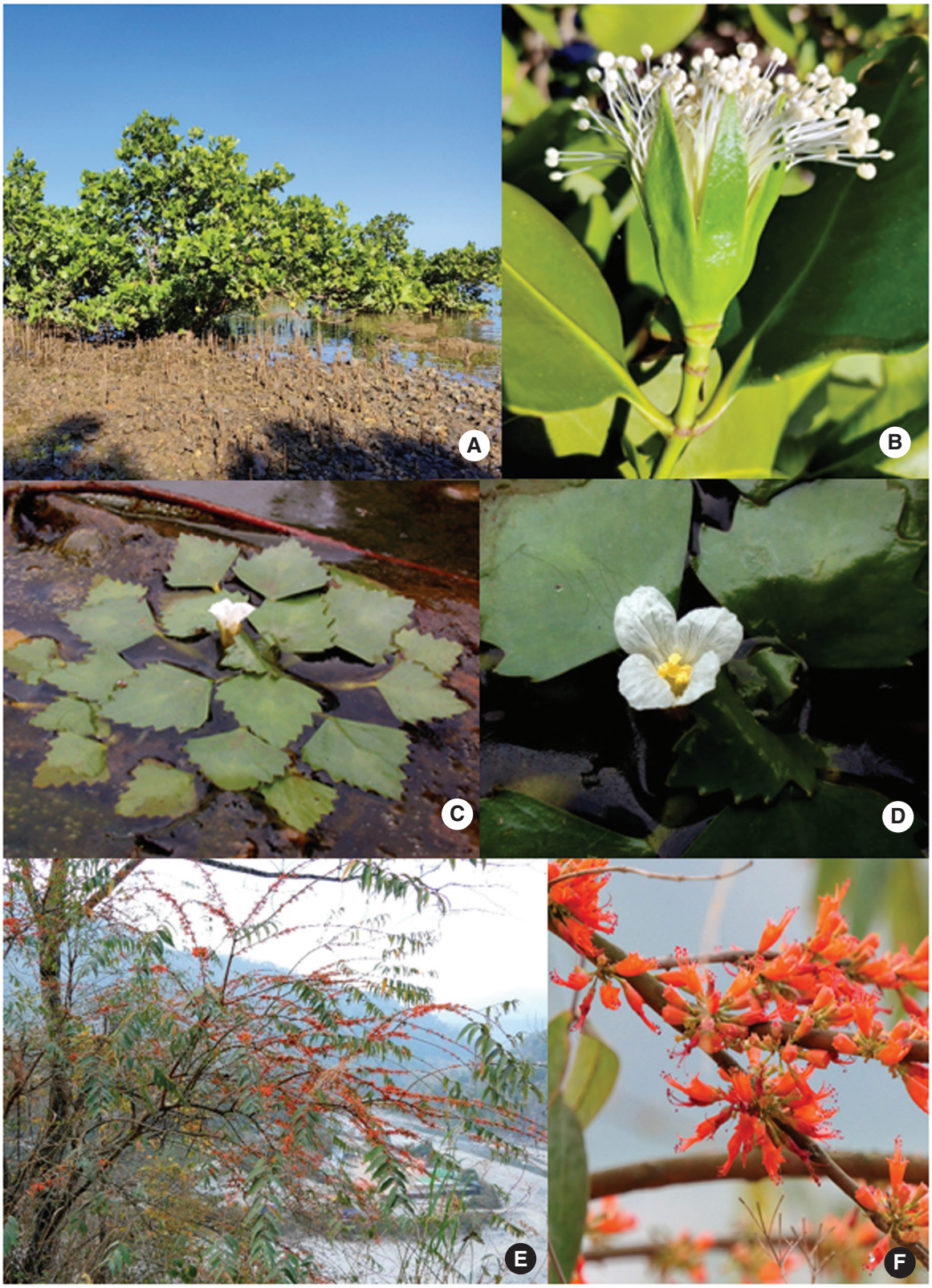

Fig. 1. Plate 1. Lythraceae genera. A-B, *Adenaria floribunda* Kunth, Acre, Brazil; C, *Ammannia latifolia* L., Paraíba, Brazil; D, *Capuronia benoistii* (Leandri) P.E.Berry, Madagascar; E, *Decodon verticillatus* (L.) Elliott, Pennsylvania, USA; F, *Cuphea rasilis* S.A.Graham, Mexico; G, *Diplusodon sordidus* Koehne, Goiás, Brazil. Photographs by: A, B, Marcos Silveira; C, Frederico Acaz Sonntag; D, Landy Rajaovelona, RBG Kew Madagascar; E, Josiah Townsend; F, Shirley A. Graham; G, Mauricio Mercadante. Plate 2. Lythraceae genera. A-B, *Didiplis diandra* (Nutt. ex DC.) Alph.Wood, Arkansas, USA; C, *Duabanga grandiflora* (Roxb. ex DC.) Walp., India; D. *Galpinia transvaalica* N.E.Br., Eswatini; E, *Ginoria americana* Jacq., Cuba; F, *Heimia apetala* (Spreng.) S.A.Graham & Gandhi, Rio Grande do Sul, Brazil; G, *Gyrosphragma latipetala* T.B.Cavalc. & Facco, Minas Gerais, Brazil. Photographs by: A, B, Jim Keesling; C, Dinesh Valke; D, Phil

T.B.Cavalc. & Facco, *Heimia* Link, *Lafoensia* Vand., *Lourtella* S.A.Graham, Baas & Tobe, *Pehria* Sprague, *Physocalymma* Pohl and *Pleurophora* D.Don). However, some *Cuphea* species are cultivated in several countries, and *Cuphea carthagenensis* (Jacq.) J.F.Macbr. is considered an invasive plant (Graham, 2017, 2019). Three genera, *Ammannia* L., *Lythrum* and *Rotala* L., are considered to have centres of diversity in the Old World, but have one to few species occurring in North or South America (Graham, 2007).

The family was first monographed by Koehne (1881, 1903), who recognized 22 genera and restricted the family to perigynous-flowered genera, distinguishing the Lythraceae from three closely related monogeneric families with semi-inferior and inferior ovaries: Duabangaceae, Punicaceae and Sonneratiaceae. As systematics underwent revolutionary change in the 20th century, first with advances in analytical techniques applied to traditional morphological characters, then with increasing access to molecular data and the accompanying development of interpretive software, the ability to clarify relationships and to devise natural, clade-based, classifications has become an integral part of botanical systematics (Stevens, 2000). The notion that a monophyletic Lythraceae might include *Duabanga* (Fig. 1: Plate 2C), *Punica* and *Sonneratia* (Fig. 1: Plate 5A-B) was supported by a cladistic analysis of morphological characters of the order Myrtales (Johnson and Briggs, 1984), also recognizing that the Onagraceae and Trapaceae could be considered sister groups. Dahlgren and Thorne (1984) suggested that *Sonneratia*, *Duabanga* and *Punica* be placed in separate subfamilies of the Lythraceae. The results of a cladistic analysis of anatomy, floral morphology, pollen and seed morphology of an expanded Lythraceae produced two weakly supported clades: one comprising *Duabanga*, *Punica* and *Sonneratia*, together with *Lagerstroemia* and *Lawsonia* (Fig. 1: Plate 3D) and another comprising the remainder of the traditional Lythraceae (Graham *et al*., 1993). More recently, availability of molecular data promoted several significant re-evaluations of Lythraceae systematics. Plastid *rbcL* sequences supported a monophyletic Lythraceae, including *Lawsonia*, *Nesaea* Comm. ex Kunth, *Duabanga*, *Trapa* (Fig. 1: Plate 5C-D), *Lythrum, Cuphea* and *Punica* (Conti *et al*., 1997) and analysis of nuclear rRNA internal transcribed spacer region (ITS) sequences confirmed that (former) Sonneratiaceae and Duabangaceae (*Sonneratia*, *Duabanga*) were nested within the Lythraceae (Shi *et al*., 2000). A later study using plastid *rbcL*, *trnL-F* and *psaA-ycf3* regions, as well as ITS and morphology, ultimately provided support for the union of the monogeneric families Duabangaceae, Punicaceae, Sonneratiaceae and Trapaceae into a monophyletic Lythraceae (Graham *et al*., 2005). Addressing more specific taxonomic questions, ITS and plastid *rbcL* and *trnL-F* regions were used to resolve relationships among *Ammannia* (Fig. 1: Plate 1A-B), *Hionanthera* A.Fern. & Deniz, *Nesaea* and *Rotala* (Fig. 1: Plate 4F), leading the former three genera to be subsumed in *Ammannia* (Graham *et al*., 2011). More recently, a fourth genus, *Crenea* Aubl., was also shown to be nested within *Ammannia*, using ITS, *rbcL*, *trnL*, *trnL-F* and *matK* regions, in a study that included 27 currently accepted Lythraceae genera (Graham *et al*., 2021).

Most recently, high-throughput nucleic acid sequencing technology has massively expanded the amount of data available for phylogenetic analysis, greatly improving resolution and support in trees and enabling deeper understanding and confidence in our evaluation of plant systematics in an evolutionary framework (Hörandl and Appelhans, 2015). A recent large-scale phylogenomic study produced exon data using the universal Angiosperms353 nuclear probe kit, with the aim of improving resolution of relationships among families and genera in the order Myrtales (Maurin *et al*., 2021). Lythraceae sampling included 26 of the 28 currently accepted genera in the family, missing *Lafoensia* and the recently described *Gyrosphragma* (Cavalcanti *et al*., 2022). Such massive data present an analytical challenge and some incongruencies were recorded between ASTRAL coalescence-based, and concatenated maximum-likelihood (ML) approaches in the topology of the backbone of the order (Maurin *et al*., 2021). However, the former approach is generally favoured in terms of phylogenetic accuracy (Mirarab and Warnow, 2015).

Genes amplified and sequenced from plastid DNA have long been the mainstay of plant molecular systematics (Shaw *et al*., 2005; Gitzendanner *et al*., 2018). High-throughput sequencing technologies and streamlined organelle genome assembly techniques now offer low-cost and time-effective generation of plastome data, not only from high-quality samples, but also from DNA extracted from historical herbarium specimens (Bakker *et al*., 2016; Bakker 2017). Maternal inheritance and evolutionary conservation of plastomes provides highly orthologous alignments of large genomic data sets that are valuable for phylogenetic analyses and calibrated divergence estimations (Wen *et al*., 2018; Stettler *et al*., 2021). In the Lythraceae, Gu *et al*. (2019) sequenced the plastomes of several *Lagerstroemia* species as well as *Duabanga grandiflora* (Roxb. ex DC.) Walp, *Trapa natans* L, *Lythrum salicaria* L., *Lawsonia inermis* L., *Woodfordia fruticosa* (L.) S.Kurz (Fig. 1: Plate 5E-F) and *Rotala rotundifolia* (Buch.-Ham.) Koehne. Together with *Pemphis acidula* J.R Forst. & G. Forst. (Fig. 1: Plate 4A-B), *Heimia myrtifolia* Cham & Schlect. (accepted name: *Heimia apetala* (Spreng.) S.A.Graham & Gandhi) (Fig. 1: Plate 2F) and *Punica granatum* L. plastomes, a phylogenetic tree was determined, congruent with current phylogenetic hypotheses in the family. Two subsequent studies, also focusing on *Lagerstroemia*, have used similarly limited taxon sampling

White; E, Todd Olson; F, Marlon G. Facco; G, Paulo M. Gonella. Plate 3. A, *Koehneria madagascariensis* (Baker) S.A.Graham, H.Tobe & Baas, Madagascar; B, *Lafoensia pacari* A.St.-Hil., Brasília, Brazil; C, *Lagerstroemia indica* L., Brazil, cultivated; D, *Lawsonia inermis* L., Anguilla, cultivated; E, *Lythrum maritimum* Kunth, Rio Grande do Sul, Brazil; F-G, *Pehria compacta* Sprague, Honduras. Photographs by: A, Landy Rajaovelona, RBG Kew Madagascar; B, C, Mauricio Mercadante; D, Karl Questel; E, Sergio Bordignon; F, G, Oliver Komar, Zamorano Biodiversity Center, Honduras. Plate 4. A-B, *Pemphis acidula* J.R.Forst. & G.Forst., Maldives; C, *Physocalymma scaberrimum* Pohl, Brasília, Brazil; D, *Pleurophora saccocarpa* Koehne, Mato Grosso do Sul, Brazil; E, *Punica granatum* L., Brasília, Brazil, cultivated; F, *Rotala ramosior* (L.) Koehne, Rio Grande do Norte, Brazil. Photographs by: A, B, Emanuele Santarelli; C, E, Mauricio Mercadante; D, J.A. Siqueira Filho; F, Gabriel Garcia. Plate 5. A-B, *Sonneratia alba* Sm., New Caledonia; C-D, *Trapa natans* L., France; E-F. *Woodfordia fruticosa* Kurz, India. Photographs by: A, B, Benoit Henry; C, D, François-Xavier Taxil; E, F, Dinesh Valke.

to determine plastome-based calibrated phylogenies of the Lythraceae (Dong *et al*., 2021; Wang *et al*., 2023).

Of the phylogenetic hypotheses available in the Lythraceae, the study of Graham *et al.* (2021), based on ITS and four plastid loci, is the most well-sampled to date, including 27 current genera. The presented Bayesian consensus phylogram was generally highly resolved, although support for several clades was lacking. The Angiosperms353 study is currently the most highly resolved and supported phylogenetic hypothesis in the Lythraceae (Maurin *et al*., 2021). Neither of these studies included a calibrated divergence dating, or biogeographical analysis, which would have contributed to a more complete understanding of the evolution of the Lythraceae and of the distribution of its genera and species.

An early intuitive phylogeny, incorporating historical biogeographical ideas, placed the African genus *Nesaea* (now a synonym of *Ammannia*) as the oldest extant genus in the Lythraceae, along with other small, primarily herbaceous, marsh genera, such as *Rotala*, *Didiplis* and *Lythrum* (Koehne, 1886; reproduced in Graham *et al*., 2021). Based on extant plant distributions and plate tectonics, Raven and Axelrod (1974) suggested a Gondwanan origin for the Lythraceae and Onagraceae. A later cladistic analysis of morphological characters in the Lythraceae considered Old World woody shrubs and trees, such as *Sonneratia*, *Duabanga*, *Punica*, *Lagerstroemia*, *Pemphis* and *Capuronia* as less derived taxa, along with the North American *Decodon*, by virtue of its extensive middle Eocene fossil history in northern Europe as well as North America (Graham *et al*., 1993). Further, these authors hypothesized an early extensive radiation in the Lythraceae from an origin in the Old World, followed by dispersal to the Western Hemisphere of genera such as *Lythrum* and later diversification of genera, such as *Lourtella* and *Cuphea*, in South America. Overcoming some of the obstacles raised by extensive character homoplasy in phenotypic analyses, *rbcL* sequence data supported an early divergence of the Myrtales into two major clades, one comprising Myrtaceae and Melastomataceae and a second comprising Combretaceae, Onagraceae and Lythraceae, with fossil evidence supporting a West Gondwanan rather than Australasian origin for the entire order in the Medial Cretaceous (Conti *et al*., 1997). Some multi-gene molecular phylogenetic studies suggested a Laurasian rather than Gondwanan origin of the Lythraceae (Graham *et al*., 2005; Morris, 2007), influenced by the rich and diverse early fossil flora from the Northern Hemisphere, compared to the paucity of Southern Hemisphere Lythraceae fossils (Graham, 2013). These studies were hampered by low backbone support, possibly reflecting an early, rapid radiation in the family (Morris, 2007). In a seminal study, Berger *et al*. (2016) used six loci (*rbcL*, *ndhF*, *matK*, *matR*, 18S and 26S) to study the phylogeny and historical biogeography of the Myrtales, where they included 14 Lythraceae genera and used four fossil Lythraceae among their extensive calibrations. These authors provided additional evidence supporting a western Gondwanan origin for the Myrtales, with divergence of the Lythraceae and its sister, the Onagraceae, occurring on the early South American continent by the end of the Albian (~105 Ma), pre-dating earlier estimates (~93 Ma; Sytsma *et al*., 2004). Later work using many of the same fossils in their calibrations and similarly incomplete sampling of Lythraceae genera (Inglis and Cavalcanti, 2018; Dong *et al*., 2021; Wang *et al*., 2023) reported biogeographical findings largely congruent with Berger *et al*. (2016).

To summarize the main deficiencies in current understanding of evolutionary relationships in the Lythraceae, we emphasize incomplete sampling of genera and low backbone resolution and support in conventional single- or multi-gene phylogenetic hypotheses. While the latter deficiency has been largely addressed by recent nuclear phylogenomics results (Maurin *et al*., 2021), a comprehensive historical biogeographical model based on a well-supported and fully resolved phylogenetic hypothesis of Lythraceae genera is currently lacking. Moreover, the evolution of the plastid genome in the family is intrinsically of interest (Gitzendanner *et al*., 2018). One of the principal factors that determined our choice of methodologies to be employed in the current study was that much of the material we would work with was from older vouchered herbarium specimens. Complete plastome assembly from such degraded DNA has become commonplace, largely due to the relative overrepresentation of the plastid genome in preserved leaf tissue (Straub *et al*., 2012; Bakker *et al*., 2016). In contrast, baited single- or low-copy number nuclear gene datasets derived from herbarium specimens may suffer from much missing data (Brewer *et al*., 2019). Secondly, compared to analyses of whole plastomes, large and complex nuclear genomic datasets remain computationally challenging for Bayesian relaxed molecular clock methods used for divergence dating (Ho, 2014); we selected this analytical methodology for consistency with previous studies in the Lythraceae (Berger *et al*., 2016; Inglis and Cavalcanti, 2018; Dong *et al*., 2021; Wang *et al*., 2023). To address some of the deficiencies in current knowledge of Lythraceae evolution, we first compiled a database of occurrence records for the family, based on online records, to accurately map the genera and species and to record their global densities. We then sequenced, assembled and annotated plastid genomes from a representative species of all remaining currently accepted Lythraceae genera. We then used the aligned whole plastome sequences to calculate a robust molecular phylogenetic tree and conducted a Bayesian fossil-calibrated analysis to construct a model of the historical biogeography of the Lythraceae, fully sampled at the genus level, in an absolute time-frame.

MATERIALS AND METHODS

Taxon sampling and library preparation

Vouchered specimens of a representative species of each currently recognized genus in the Lythraceae were obtained (unless otherwise stated) from CEN or MO herbaria (Supplementary Data Table S1). Already complete and annotated plastome sequences were obtained from GenBank (Table S2). Genomic DNA was extracted from approximately one thumbnail-sized piece of herbarium preserved or silica-dried leaf tissue. We used a CTAB-based DNA extraction protocol supplemented with a sorbitol solution prewash of tissue macerate, as described previously (Inglis *et al*., 2018). In several cases, multiple miniprep-scale extractions from fragments of the same herbarium accession were pooled to accumulate sufficient DNA for successful library preparation of sufficient complexity for complete plastome assembly (>100 ng). Highly degraded

DNA samples derived from herbarium material were sent to a service provider (Daicel Arbor Biosciences, Ann Arbor, MI, USA), where libraries were prepared using single-stranded chemistry to improve library complexity (Gansauge and Meyer, 2013). Libraries were sequenced using the Illumina paired-end (2 × 150 bp) platform with projected data yield of 3.0 G per sample. High-quality DNA samples derived from silica-dried or fresh material were sent to another service provider (Novogene UK Company Limited, Cambridge, UK) for standard Illumina low-coverage whole genome shotgun library preparation and 2 × 150-bp paired-end sequencing with a mean data yield of 2.0 G and Q30 Phred scores of > 90 % in all libraries. Raw sequence data were trimmed for adapter and other artefact sequences with the Fastp program (Chen *et al*., 2018), which was also used for filtering of possible end-of-strand poly-G signals from Illumina two-colour chemistry. These were commonly encountered in the shorter reads from the single-stranded libraries used for highly degraded herbarium material, which were thereafter treated as single-ended during assembly.

A previously published plastome sequence attributed to *Rotala rotundifolia* (GenBank accession MK881626), included in earlier plastome-based phylogenic studies (Gu *et al*., 2019; Dong *et al*., 2021), was rejected because it unexpectedly clustered identically with the plastome from *Lawsonia inermis*, in stark contrast to several earlier molecular phylogenetic studies which clearly distinguished these genera (see Graham *et al*., 2005). We therefore independently sequenced the *R. rotundifolia* plastome and rDNA ITS sequence from new source material (Supplementary Data Table S1). We confirmed the expected phylogenetic placement of the new ITS sequence by a BLAST search of NCBI GenBank, where all 20 top hits were with *Rotala* spp. ITS sequences and the top hit was with an *R. rotundifolia* sequence from a vouchered source (CALI 132932; GenBank accession MH071602) with 100 % identity.

Geographical mapping of Lythraceae genera and species

A database of geographical occurrence records for Lythraceae (downloaded at the family level) was compiled from public sources GBIF (GBIF.org, 2023) and speciesLink (https://specieslink.net/). Using scripts from the analyses by Zizka *et al*. (2019) in R (R Core Team, 2023), the two databases were merged and spatial errors and problematic records were removed from the occurrences (records without geographical coordinates; records older than 1940; records based on fossils, tissue samples and living collections; records in the sea, on country or province centroids; individual localities with many records). Employing the same scripts, species-level names were resolved using an up-to-date taxonomic list of the family (accepted names, synonyms and native distribution), assembled from data from Plants of the World Online (POWO, 2023) and Flora e Funga do Brasil (2023). This list was also used to eliminate records outside the country in which a species was recorded in POWO when these data were available (Zizka *et al*., 2019). In QGIS (2023), the cleaned occurrence records were plotted on a world political division map (Natural Earth, 2023), and Lythraceae species and genus richness patterns were visualized using the FCS Biological Records Tool plugin, based on a 1° × 1° grid.

Plastome assembly, annotation and descriptive analysis

Plastomes were assembled using either GetOrganelle (v.1.7.6.1; Jin *et al*., 2020) with default settings or Geneious Prime (v.2022.1.1; Biomatters Ltd, Auckland, New Zealand). The latter assemblies employed the 'map to reference' function and the Geneious assembler module with default settings, using the phylogenetically closest available Lythraceae plastome sequence as reference. Newly sequenced plastome assemblies as well as plastomes obtained from GenBank were verified and circularized with NOVOWRAP (v.1.20; Wu *et al*., 2021), using the *Punica granatum* plastome (NC_035240) as reference. Library coverage statistics for the assembled plastomes were calculated using the BBMap tool (Bushnell, 2014). Newly completed plastomes were annotated using GeSeq (Tillich *et al*., 2017), part of the CHLOROBOX online suite of tools (https://chlorobox.mpimp-golm.mpg.de/index.html), supported by Chloë (www.chloe.plastid.org). Annotations were verified and prepared for database submission using GB2sequin (Lehwark and Greiner, 2019) and Geneious Prime. Plastome physical maps were drawn using OGDRAW (Greiner *et al*., 2019). Plastomes were screened for the presence of microsatellites (simple sequence repeats, SSRs) using the MIcroSAtellite identification tool (MISA-web; Beier *et al*., 2017), using the default minimum number of repetitions for each SSR motif between one and six bases. Searches were also made for long dispersed forward, reverse, palindromic and complementary repeats, using REPuter (Kurtz *et al*., 2001), with Hamming distance of three and minimum repeat size of 30. Comparative analysis of the higher-order organization among the plastomes employed the Progressive Mauve module of Geneious Prime, with default settings (v,2.4.0; Darling *et al*., 2010).

Plastome alignment and phylogenetic analysis

The plastomes were collected into a matrix using Geneious Prime and aligned with MAFFT (v.7.505, FFT-NS-2 option; Katoh and Standley, 2013). Phylogenetic analyses were conducted on a derivative of the full matrix with one of the two major inverted repeats (IRs) removed to avoid data duplication. The matrix was partitioned into large single copy (LSC), inverted repeat-a (IRa) and small single copy (SSC) blocks. We then tested the merger of partition models using the PartitionFinder algorithm (Lanfear *et al*., 2012), as implemented in IQ-TREE software (Nguyen *et al*., 2015), which favoured maintenance of the tripartite scheme. Optimal base substitution models for each partition, according to the Bayesian information criterion (BIC), were determined using the ModelFinder module of IQ-TREE (Kalyaanamoorthy *et al*., 2017). The plastome sequence used for rooting trees in all subsequent phylogenetic analyses was from *Ludwigia octovalvis* (Jacq) P.H. Raven (Onagraceae), which has been shown in multiple studies to be the deepest branching genus in its family and a suitable sister to the Lythraceae (see Berger *et al*., 2016; Dong *et al*., 2021). A phylogenetic analysis based on the partitioned plastome matrix was conducted using Bayesian Inference (BI) in MrBayes (v.3.2.7; Ronquist *et al*., 2012). Models for each partition were unlinked and optimized during the runtime using reversible jump Markov chain Monte Carlo (rjMCMC; Huelsenbeck *et*

al., 2004; nst = mixed, rates = invgamma). One cold and three heated MCMC chains were run for five million generations, sampling every 1000th generation and discarding the first 25% of the trees (the burn-in) before calculation of the 50% majority rule consensus phylogram. To account for possible uncertainty in indel-rich portions of the alignment, we repeated the BI analysis on a derivative of the plastome matrix with gap-containing columns removed. We also conducted BI analysis on a matrix comprising only genic regions extracted from each plastome. For comparison, an ML phylogenetic analysis based on the partitioned matrix of plastome sequences (LSC + IRa + SSC) was also conducted using RAxMLGUI 2.0 (Edler *et al*., 2020; Stamatakis, 2014). The optimal model for each partition according to BIC, among those available in RAxML, was determined to be GTR+I+G4 using ModelTest-NG (Darriba *et al*., 2020). Branch support was provided by 1000 rapid bootstrap replicates.

For comparison with previously published Lythraceae molecular phylogenetic trees and to confirm taxonomic identity of the accessions used in our libraries, nuclear ribosomal ITS sequences (rDNA ITS) were assembled from each next-generation sequencing library using GetOrganelle and organized into a matrix using Geneious Prime. In case of previously published Lythraceae plastomes, ITS sequences of the same species as the representative plastomes were obtained from GenBank. The new ITS sequences from our Illumina libraries were verified for taxonomic identity by BLAST-X (Altschul *et al*., 1990) searches against the NCBI nucleotide database. A BI phylogenetic analysis of the unpartitioned ITS matrix, rooted using *Ludwigia octovalvis*, was conducted as above. The ITS rjMCMC BI analysis was run for 20 million generations, sampling every 1000th generation.

Divergence time and historical biogeographical analyses

Because node age is key to interpretation of historical biogeography, reliable fossil selection for calibration is critical (Vasconcelos *et al*., 2017). Fourteen of the 28 extant Lythraceae genera have fossil representatives: *Adenaria*, *Ammannia* (*Crenea*), *Cuphea*, *Decodon*, *Duabanga*, *Lafoensia*, *Lagerstroemia*, *Lawsonia*, *Lythrum* (*Peplis* L.), *Pemphis*, *Punica*, *Sonneratia*, *Trapa* and *Woodfordia* and ten extinct genera are recognized, where the most common kinds of fossil remains are seeds and pollen (reviewed in Graham, 2013). We calibrated the minimum ages of five nodes across the phylogeny using a combination of five fossils strongly attributable to Lythraceae (Supplementary Data Table S3; reviewed in Graham, 2013) and secondary date priors for the estimated dates of the Lythraceae–Onagraceae divergence and the Lythraceae crown node, based on a published large-scale study of the Myrtales, including Lythraceae (Berger *et al*., 2016). Our three oldest calibrations were all nodes in clade 2A, based on fossilized pollen from Wyoming, USA, attributed to *Lythrum elkensis* (81 Ma; Grímsson *et al*., 2011), the fossil wood, *Sonneratiaoxylon preapetalum* Awasthi (*Sonneratia* – 63.8 Ma; Awasthi, 1968) and the leaf impression from India, *Lagerstroemia patelii* Lakhanpal & Guleria (56 Ma; Lakhanpal and Guleria, 1981). We also used two calibrations in clade 1, based on the fossil wood from France, *Punicoxylon eocenicum* Privé-Gill (*Punica* – 40.4 Ma; Privé-Gill, 1981) and fossilized pollen from Chiapas, Mexico, attributed to *Cuphea* (14 Ma; Palacios and Rzedowski, 1993). The former four fossil calibrations were also used by Berger *et al*. (2016) and shared by a plastome-based phylogenetic study of *Lagerstroemia* which included ten other Lythraceae genera (Dong *et al*., 2021). The latter study also used a fossil seed, *Lawsonia lawsonioides* Menzel (Mai, 1996; 16 Ma), which we rejected, since we found the date to be strongly incompatible with estimates of the *Ammannia–Lawsonia* node made without this prior, reducing effective sample sizes (ESS) during trial analyses. We also favoured the *Cuphea* fossil prior to better represent clade 1 taxa.

A divergence time analysis of the partitioned LSC + IRa + SSC plastome matrix was conducted using calibrated BI in BEAST2 (v.2.6.7; Bouckaert *et al*., 2014), using the Optimized Relaxed Clock (OCR) model (Douglas *et al*., 2021) and the Birth–Death speciation process (FBD; Heath *et al*., 2014). The OCR model is based on the Uncorrelated Relaxed Clock (Drummond *et al*., 2006), but includes operators especially efficient on long alignments. The Standard TVM site substitution model with four gamma categories and estimated frequencies, gamma shape and invariant proportion was used for each unlinked partition, where the TVM+F+I+I+R3 model was favoured according to BIC in independent ModelFinder analyses. The clock and tree models were linked during the analysis. The date offsets of all fossil priors were placed under a lognormal distribution, with the mean of the log-transformed distribution (M) set to 1.5 and the standard deviation of the log-transformed distribution (S) set to 1.0, allowing for the possibility that these nodes are considerably older than the fossils themselves. Two secondary priors informing previously published estimates for the mean ages of the Lythraceae crown and stem nodes were also used and placed under a normal distribution with a sigma (standard deviation of the normal distribution) set to 1.0. The analysis was run for 500 000 000 generations, with sampling of trees and logs every 50 000 generations. Convergence of independent chains and adequate ESS for all parameters was verified using Tracer (v.1.7.2; Rambaut *et al*., 2014). The first 25 % of sampled trees were excluded as burn-in prior to calculation of a maximum clade-credibility (MCC) tree using TreeAnnotator v.1.8.0 (Drummond *et al*., 2012).

To estimate the ancestral areas occupied by Lythraceae genera at nodes throughout the phylogenetic reconstruction, we conducted a historical biogeographical analysis in BioGeoBEARS (Matzke, 2014), implemented in the RASP package (Yu *et al*., 2020). The distribution matrix used in the analysis was coded using the aforementioned occurrence records for extant Lythraceae genera, overlaid with six 'biogeographical realms' of the world (Udvardy, 1975) as: A, Neotropic (South America, southern Mexico, Central America and the West Indies); B, Nearctic (central to northern Mexico and North America); C, Afrotropic (Sub-Saharan Africa including Madagascar and Mauritius); D, Palearctic (Eurasia, including western Europe, Mediterranean Africa and temperate Asia); E, Indomalaya (Southeast Asia, including India, Indo-China, the Malaysian Peninsula, the Philippines, Sumatra, Borneo and the Inner Banda Arc); and F, Australasia (Australia, including New Guinea, New Caledonia and New Zealand, as well as the Pacific Islands) (map shown as an inset in Fig. 7). These areas define large areas of the world, within which terrestrial organisms have evolved in relative isolation by features that

have constituted natural barriers to migration (Udvardy, 1975). Similar biogeographical coding schemes, considered important in the distributions of Myrtales (Berger *et al.*, 2016), have also been used in Vochysiaceae (Gonçalves *et al.*, 2020) and Melastomataceae (Reginato *et al.*, 2022). A BioGeoBEARS Model Test indicated that a DIVALIKE model with jump speciation (i.e. dispersal between non-adjacent areas) best describes distribution patterns of Lythraceae genera, according to AIC (DIVALIKE+j model; full results given in Supplementary Data Table S4). For the BioGeoBEARS ancestral range reconstruction, we used a random selection of 100 of the post-burnin trees from the BEAST analysis and the BEAST consensus chronogram as input. Because of the wide distributions and possibility of long-distance dispersal (LDD) of several extant Lythraceae genera, such as *Lythrum*, *Rotala*, *Ammannia* and the outgroup, *Ludwigia*, we placed no a priori limit on the number of occupied ancestral areas (max areas = 5) nor time-stratified constraints, and the probabilities for movement between all biogeographical areas was set to equal.

RESULTS

World distribution of the Lythraceae

A total of 976 585 occurrence records were obtained from public sources, and after geographical and taxonomic filtering, 293 239 records were retained. All 28 genera of Lythraceae were sampled and of the 679 currently accepted species according to POWO (2023), 546 were represented. Our results show two main centres of Lythraceae species richness (Fig. 2): one in the mountains of western and southern Mexico and the other in eastern Brazil, in the Espinhaço mountain range and the central plateau, in the Cerrado biome. Lower Lythraceae species richness is present in the other tropical and subtropical regions of the world. At the genus level (Fig. 3), the Americas also concentrate diversity of the Lythraceae, but the pattern is somewhat different, where central and eastern Brazil, south-eastern USA and mountainous portions of Colombia are centres of diversity. Tropical South Asia, Japan and northern Australia also possess high Lythraceae genus richness, in addition to some regions of southern Africa and Madagascar.

Plastome organization in the Lythraceae

Complete circular plastomes were successfully assembled from all accessions, including samples processed from both standard double-stranded and single-stranded libraries. Mean base coverage, estimated from BBMap analysis, ranged from 142-fold (*Koehneria*) to 982-fold (*Ginoria*) (Supplementary Data Table S5). All of the complete Lythraceae plastid genomes (and outgroup, *Ludwigia octovalvis*) display the quadripartite structure of two single-copy regions (LSC and SSC) separated by a pair of IRs, typical of flowering plants (Wicke *et al.*, 2011). Additionally, alignment and analysis using Progressive Mauve showed the plastomes having a high degree of synteny (Fig. S1). A typical physical map, exemplified by the plastome of *Lourtella resinosa* S.A.Graham, Baas & Tobe, is illustrated (Fig. 4). The size of Lythraceae plastomes was found to vary from 152 205 bp (*Lagerstroemia indica* L.) to 165 558 bp (*Rotala rotundifolia*), but most clustered closely around a mean of 158 kb (Table S6). GC content of the plastomes was found to vary slightly from 36.4 % (*Trapa natans*) to 37.6 % (*Lagerstroemia indica*). As well as their conservation in length, Lythraceae plastomes are generally highly conserved in gene content and gene order (Fig. S2). All newly sequenced and annotated plastomes, regardless of assembled size, possess 112 genes, as called by the ChloroBox annotator.

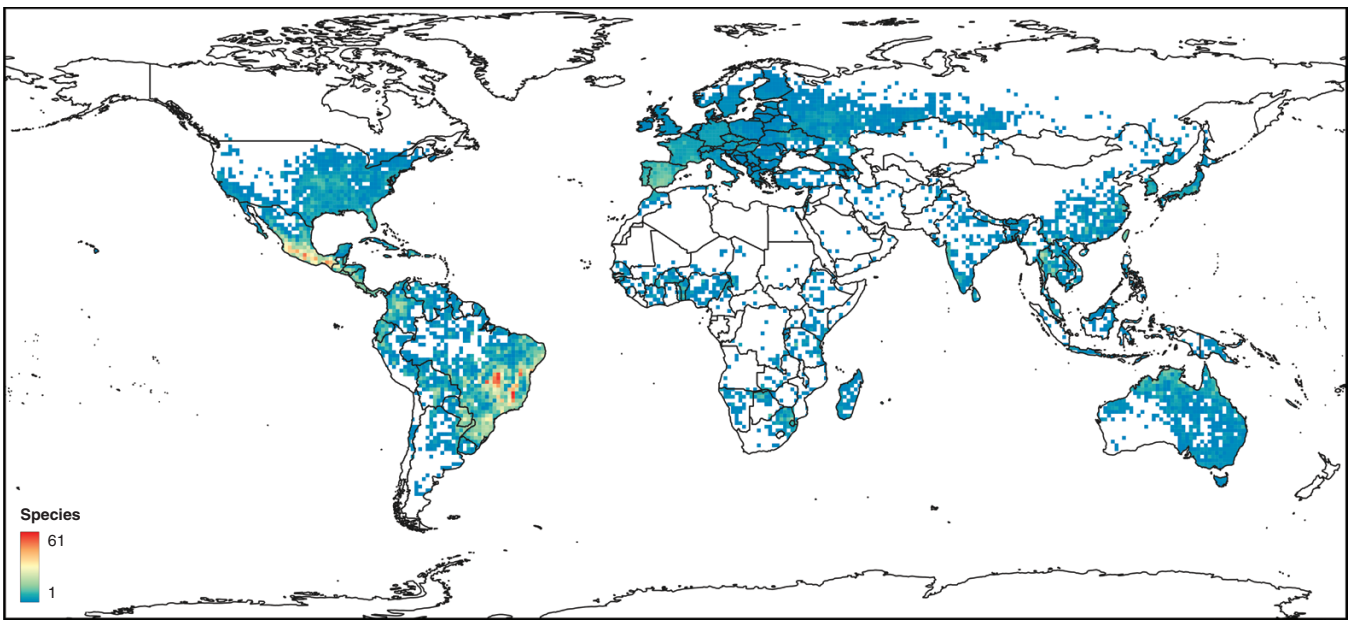

FIG. 2. Lythraceae species richness. The colour of each square indicates the number of different Lythraceae species recorded in a 1° × 1° world grid, according to the indicated scale.

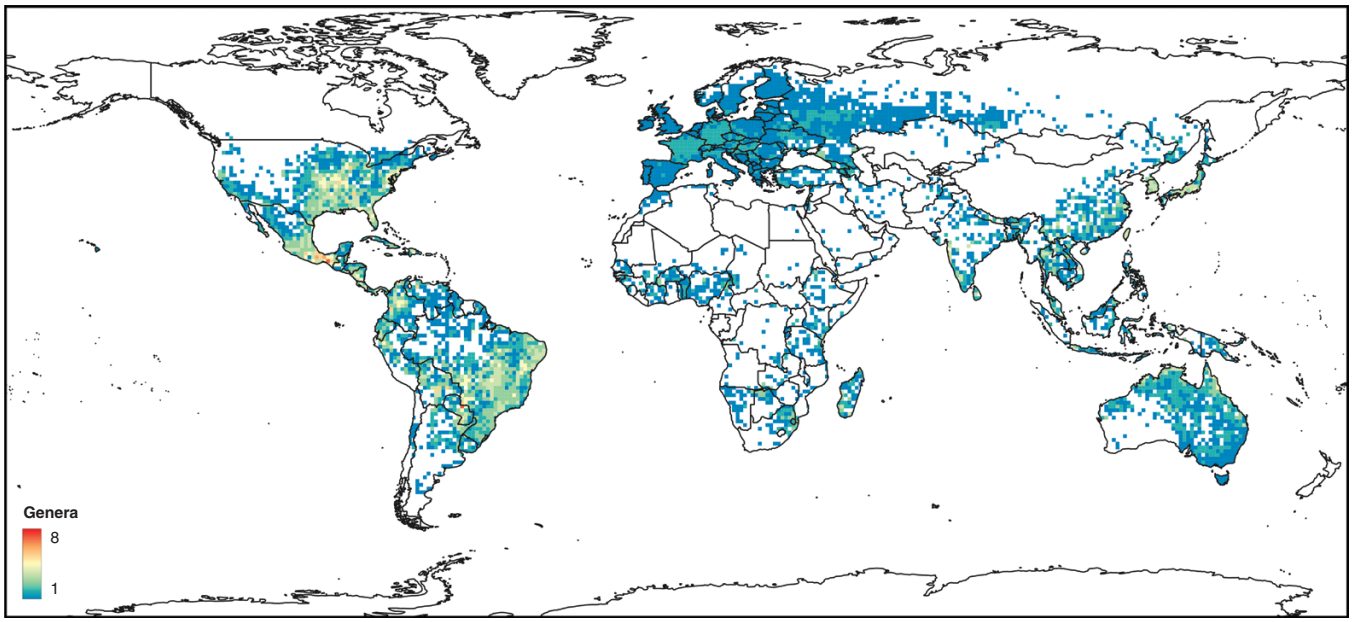

FIG. 3. Lythraceae genus richness. The colour of each square indicates the number of different Lythraceae genera recorded in a 1° × 1° world grid, according to the indicated scale.

The most common class of chloroplast microsatellites, or tandem SSRs (cpSSR), among the Lythraceae plastomes are A/T mononucleotides ($n \geq 10$), being most abundant in *Trapa natans* (89 distinct tracts) and least abundant in *Duabanga grandiflora* (30 tracts) (Supplementary Data Tables S7 and S8). Nineteen Lythraceae genera were found to possess G/C mononucleotide repeats of $n \geq 10$, though this cpSSR was rare in each plastome, with a maximum number of five distinct poly-G/C tracts being found in the *Pleurophora saccocarpa* Koehne plastome. Higher order cpSSRs were infrequent among the plastomes. The dinucleotide class of cpSSR exclusively comprised AT repeats and the longest tract ($n = 9$) was found in *Woodfordia fruticosa*. Short trinucleotide cpSSRs ($n = 5$ only) were found in 12 out of 29 genera and only *Duabanga grandiflora* was found to possess a single tetranucleotide cpSSR (AAAT; $n = 5$). Non-A/T cpSSRs were most frequent in *Sonneratia alba* Sm. (12) and none were found in the plastome of *Gyrosphragma latipetala* T.B.Cavalc. & Facco. The total number of long dispersed repeats (LDRs; Table S7) varied from 27 in *Pleurophora saccocarpa* to 102 in *Ammannia robusta* Heer & Regel, where roughly equal numbers of the forward and palindromic (reverse-complement) classes of LDR were found in each genus. The complement and reverse classes of LDR were comparatively rare and neither cpSSR nor LDR content was proportional to overall plastome size in the Lythraceae.

In the majority of the Lythraceae, IRb is inserted into either the 3′ end of the *rps19* gene or the nearby *rpl2-rps19* intergenic region of the LSC and into a pseudogenized copy of *ψycf1* of the SSC, which in a few cases overlaps the reading frame of *ndhF*. IRa is almost exclusively inserted into the large copy of *ycf1* of the SSC and into the *rpl2–trnH-*^{GUG} intergenic region. In some cases, the distal end of IRa carries a fragment of *ψrps19*. The somewhat smaller *Ginoria nudiflora* (Hemsl.) Koehne and *Tetrataxis salicifolia* (Thouars) Baker plastomes have IRs retracted from the SSC, inserting into *trnR-*^{ACG}*–trnN-*^{GUU} intergenic spacer copies. The notably larger size of the *R. rotundifolia* plastome is due to expansion of the IRs into the LSC, where the element extends into *rpoA* (Supplementary Data Table S6; Fig. S2). The *Didiplis diandra* (Nutt. ex DC.) Alph.Wood (Fig. 1: Plate 2A-B) plastome is also larger than most Lythraceae plastomes, where the IRs have expanded as far as the *rpl22–rps3* intergenic region. The smaller sizes of some of the Lythraceae plastomes, such as of *Ginoria nudiflora*, *Lagerstroemia indica*, *Sonneratia alba*, *Tetrataxis salicifolia* and *Trapa natans*, are more convincingly accounted for by reduced intergenic DNA (Supplementary Data Table S7).

Phylogenetic analysis

The phylogram based on the alignment of plastome data (LSU, IRa and SSU) was fully resolved and fully supported in the BI analysis (Fig. 5). Although identical in topology, six particularly short branches have lower bootstrap support in the ML analysis. The runtime of the BI analysis was sufficient for the MCMC chains to converge satisfactorily, which was confirmed in Tracer plots (Rambaut *et al.*, 2014) and where the average standard deviation of split frequencies fell to 0.000029 and, for all model parameters, values for the potential scale reduction factor (PSRF) and ESS were 1.000 and > 1200 respectively. The consensus tree based on BI analysis of the matrix filtered for gapped positions and the tree based on only genic regions were closely similar in general topology and support to the full matrix (not shown). In the ITS BI analysis, the standard deviation of split frequencies fell to 0.0077 and, for all model parameters, PRSF and ESS values were 1.00 and > 200 respectively. The topology of the ITS BI tree is largely congruent with the plastome-based tree, but is much less resolved and more poorly supported (Supplementary Data Fig. S3).

The Lythraceae crown node in the plastome-based BI phylogram is resolved into two principal clades: clade 1

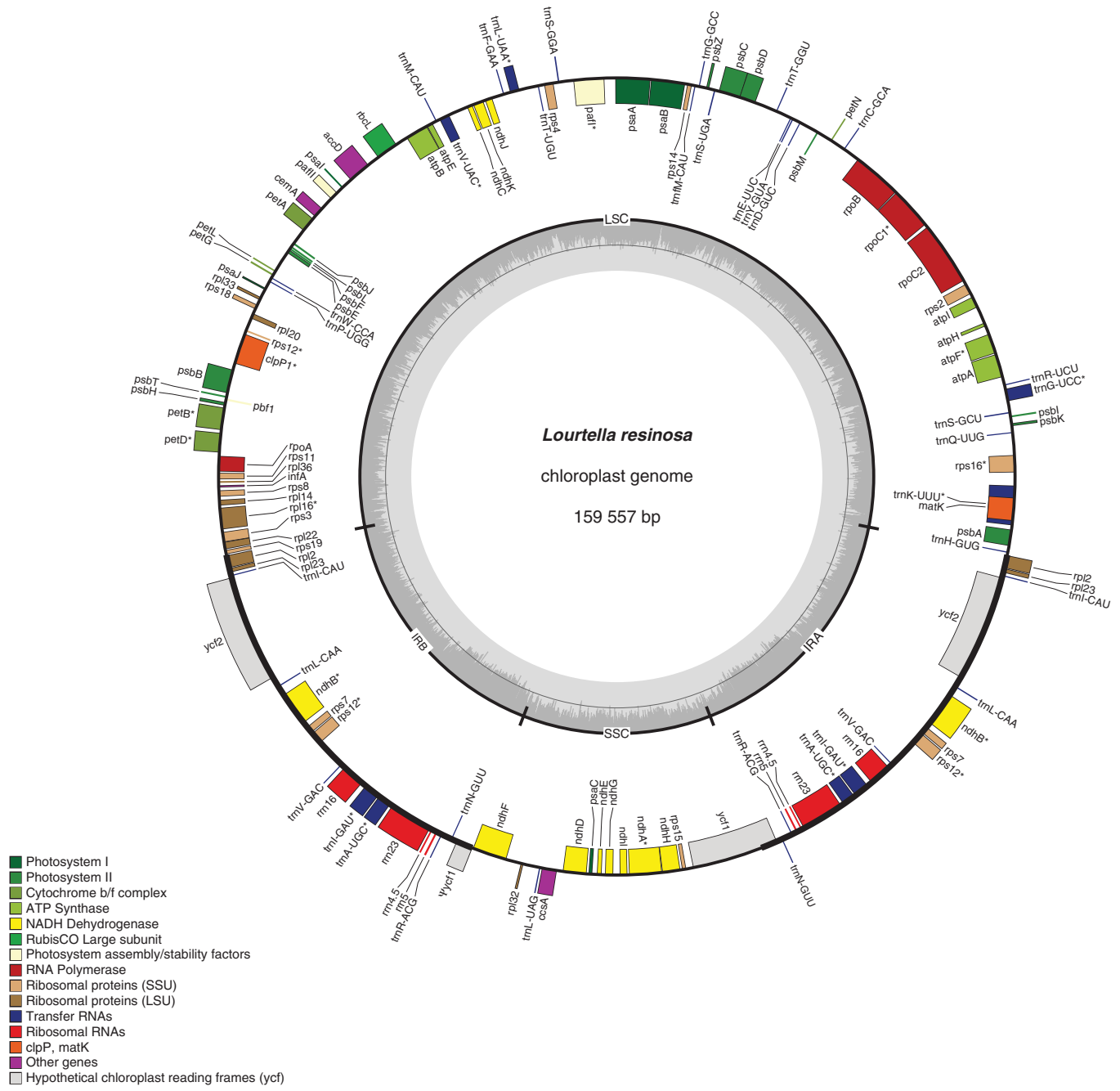

FIG. 4. Physical map of the annotated *Lourtella resinosa* plastome. The central circle (in grey) represents local GC content and genes marked with an asterisk contain introns.

(*Cuphea–Pemphis*) and clade 2 (*Ammannia–Heimia*) (Fig. 5). Clade 1 is divided into subclade 1A, with a short branch leading to a group comprising *Lourtella* and *Physocalymma*, then a long branch leading to *Diplusodon*, and two further subclades, one comprising *Adenaria*, *Pehria*, *Koehneria* and *Woodfordia*, and another comprising *Cuphea*, the recently discovered *Gyrosphragma* and *Pleurophora*. A short branch at the base of clade 2 leads to two subclades: 2A and 2B. The node uniting these subclades is fully supported in the Bayesian analysis and supported by a bootstrap value of 97 % in the ML analysis. Subclade 2A comprises three closely spaced divergences leading to *Decodon*, then *Lythrum*, and then a bifurcation leading to two monophyletic groups, one comprising *Sonneratia*, *Trapa*, *Lagerstroemia* and *Duabanga* (subclade 2A-2), and another comprising *Ammannia*, *Lawsonia*, *Tetrataxis* and *Ginoria* (subclade 2A-1). Subclade 2B comprises *Heimia*, *Rotala* and *Didiplis.*

Historical biogeography of the Lythraceae

The plastome-based Bayesian chronogram (Fig. 6) provides estimates for the dates of the Onagraceae–Lythraceae divergence

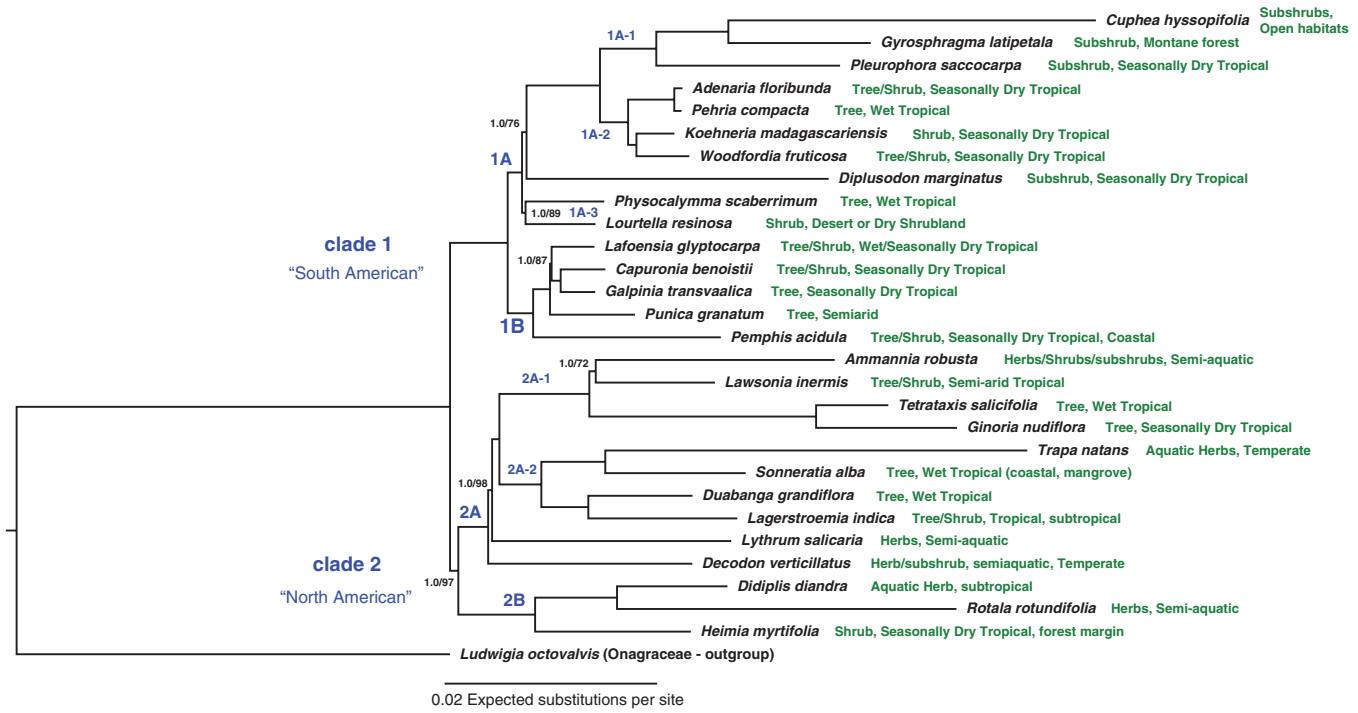

FIG. 5. Bayesian inference (BI) consensus phylogram of the Lythraceae, based on plastome (LSU, IRa, SSU) data. The topology was fully supported (PP = 1.0) in the BI analysis, but branches with < 100% bootstrap support in the ML analysis are indicated, preceded by the PP value. The tree was rooted with *Ludwigia octovalvis* (Onagraceae). Habit and ecological niche are indicated in green and clade designations, as discussed in the text, are in blue.

and the Lythraceae crown node of ~104.6 Ma (95% highest posterior density (HPD) = 102.7–106.6 Ma) and ~95.7 Ma (95% HPD = 93.9–97.7 Ma), in the late Early Cretaceous and early Late Cretaceous respectively. Ancestral range estimates (AREs) using BioGeoBEARS are shown, where area probability is represented by pie charts for each node (Fig. 7). Probabilities for predicted dispersal and vicariance events are given in the Supplementary Data.

'South American' lineage (clade 1)

The South American lineage is estimated to have diverged by ~71 Ma (95 % HPD = 55.4–87.3 Ma), with one subclade remaining largely *in situ* (subclade 1A – except for *Woodfordia–Koehneria*) and another dispersing to Africa, producing disjunct populations (subclade 1B). The 1B node, although at low probability, is estimated to be African in the BioGeoBEARS analysis, and to have led to the divergence of *Pemphis* by ~55 Ma (95 % HPD = 43.3–68.0 Ma), with further dispersal to Asian, Eurasian and Austro-Pacific regions. The clade sister to *Pemphis*, with disjunct distribution in South America, Africa and Eurasia, appears to have been followed by the vicariant divergence of *Punica*, which today has a native distribution from Iran to northern India, by ~44 Ma (95 % HPD = 40.6–48.7 Ma), and of *Lafoensia* (Fig. 1: Plate 3B) by ~42 Ma (95 % HPD = 35.5–50.0 Ma); representing the remaining South American part of the lineage. The remaining African part of this lineage led to the divergence of two monospecific genera, *Capuronia* (Fig. 1: Plate 1D) and *Galpinia* (Fig. 1: Plate 2D), by ~30 Ma (95 % HPD = 11.9–44.2 Ma).

Subclade 1A is confidently estimated to be South American, diverging into two lineages by ~62 Ma (95 % HPD = 47.0–79.4 Ma). By ~59 Ma (95 % HPD = 47.0–79.4 Ma), one lineage is estimated to have led to the divergence of *Physocalymma* and *Lourtella* and another to *Diplusodon* and its sister lineage, which our results indicate later diverged into two clades by ~45 Ma (95 % HPD = 30.4–59.5 Ma), where one branch is predicted to have dispersed to Africa and another remained wholly South American. The South American branch is estimated to have led to the divergence of *Pleurophora* by ~33 Ma (95 % HPD = 21.6–46.5 Ma) and to the divergence of *Gyrosphragma* and *Cuphea* by ~20 Ma (95 % HPD = 14.3–28.6 Ma). *Cuphea* is predicted to have later expanded to North America, subsequently speciating extensively in the Americas, especially in Brazil and Mexico, to become the most speciose genus in the Lythraceae. The remaining branch with disjunct South American–African distribution is predicted to have diverged vicariously by ~25 Ma (95 % HPD = 10.6–40.2 Ma), with the African component leading to the divergence of *Koehneria* and *Woodfordia* by ~19 Ma (95 % HPD = 5.9–33.3 Ma), with the latter genus expanding to South Asia. The remaining South American component of the clade led to the divergence of *Pehria* and *Adenaria* by ~4 Ma (95 % HPD = 0.6–8.6 Ma), predicted to be the most recently diverged genera in the Lythraceae.

'North American' lineage (clade 2)

The North American lineage is estimated to have begun to diverge by ~92 Ma (95 % HPD = 88.4–96.2 Ma) into subclade 2A and 2B lineages (Figs 6 and 7). In subclade 2A, our results suggest that closely spaced events in the North American region led to the divergence, by ~86 Ma (95 % HPD = 81.6–90.3 Ma), of *Decodon*, and by ~84 Ma (95 % HPD = 78.8–89.0

Ma), of *Lythrum*. Although the former genus is today restricted to eastern North America, *Lythrum* today possesses a near-cosmopolitan modern distribution. The lineage sister to *Lythrum* is estimated have dispersed to Asia and diverged into two clades by ~81 Ma (95 % HPD = 75.6–86.9 Ma), one remaining in Asia and a second clade expanding to Africa. Our results suggest that the Asian clade diverged *in situ* by ~74 Ma (95 % HPD = 68.6–80.1 Ma) into two lineages, with one lineage leading to the divergence of *Duabanga* and *Lagerstroemia* by ~60 Ma (95 % HPD = 56.2–65.7 Ma), with eventual expansion of the range of the latter genus to the Austro-Pacific region. The second lineage is estimated to have led to the divergence of *Trapa* and *Sonneratia* by ~66 Ma (95 % HPD = 64.0–69.5 Ma), with subsequent expansion of the range of *Trapa* from Asia to Eurasia and Africa, and of *Sonneratia* to East Africa and the Austro-Pacific region. The African lineage (*Ammannia–Ginoria*) is estimated to have diverged by ~56 Ma (95 % HPD = 37.1–74.0 Ma) and expanded to Asia. One lineage is estimated to have led to the divergence of *Ammannia* and *Lawsonia* by ~51 Ma (95 % HPD = 31.0–70.4 Ma) with subsequent wide dispersal of the former genus throughout wet regions of the global tropics and subtropics and of the latter genus to East Africa, the Middle East and India. The second 'African' lineage is estimated to have possessed a vicariant African and South American distribution, leading to the divergence of *Tetrataxis* and *Ginoria* by ~25 Ma (95 % HPD = 9.2–42.2 Ma).

The first genus to diverge in subclade 2B is shown to be *Heimia*, by ~65 Ma (95 % HPD = 41.2–85.8 Ma) in North America. By ~40 Ma (95 % HPD = 17.1–63.2 Ma), also in North America, the remainder of subclade 2B is estimated to have diverged into two genera of aquatic or semi-aquatic herbs: the monospecific *Didiplis*, today native to south-eastern to central North America, and *Rotala*, a genus of ~72 species distributed worldwide throughout the tropics and subtropics.

DISCUSSION

Plastome variation and evolution in the Lythraceae

The Lythraceae plastomes examined in the present study demonstrate a high degree of synteny and structural conservation. Small variations were observed, particularly in the precise site of insertion of the large inverted repeats, but no clear phylogenetic correlation was discernible, apart from a small trend for inverted repeat expansion into the LSU seen in two taxa of aquatic habitats in clade 2B: *Didiplis diandra* and, most markedly, in *Rotala rotundifolia*. Further plastome sequences in other species in Lythraceae genera will be required for deeper generalizations to be made, in terms of plastome length and organization. However, *Lagerstroemia* is one of the Lythraceae with the largest number of sequenced plastomes to date, where closely similar plastome lengths in different species, between 151 968 and 152 629 bp, have been reported (Dong *et al*., 2021). Fifteen plastomes for species of *Trapa* have also been reported, with lengths from 155 453 to 155 559 bp (Fan *et al*., 2022). Differences in annotated gene content compared to several previously published Lythraceae plastomes (Supplementary Data Table S6) are expected with advances in sensitivity of annotation software (Guyeux *et al*., 2019). However, the total length of annotated genes appears to be uniform among Lythraceae plastomes, as is the combined length of introns, especially when annotation variance is taken into account. Chloroplast translation initiation factor 1, *infA*, is a pseudogene in all Lythraceae (Table S6), where the smallest open reading frame with an AUG initiation codon is seven codons in *R. rotundifolia* (two copies in inverted repeats) and the largest is 68 in *Capuronia benoistii* P.E.Berry. The Lythraceae *ψinfA* is variably frameshifted or truncated at the 3′ or 5′ end compared to, for example, *infA* from *Vitis vinifera* L. (77 AAs) which, uncommonly among rosids, has been reported to be functional (Millen *et al*., 2001).

A plastome-based Lythraceae phylogeny

A bipartite model of early Lythraceae evolution into two major lineages is congruent with a recent phylogenomic study of the Myrtales, using the Angiosperm353 nuclear probe set (Maurin *et al*., 2021), as is the early divergence of the *Heimia–Didiplis–Rotala* clade. This well-supported and comprehensive backbone topology for the Lythraceae is also largely congruent with some more recent well-sampled multigene phylogenetic hypotheses in the family that have suggested a similar topology (Graham *et al*., 2021; Cavalcanti *et al*., 2022), but differ significantly from earlier, more poorly supported topologies (see Graham *et al*., 2005). In particular, the positions of *Decodon* and *Lythrum* in earlier analyses have been unstable, occurring either at or near the base of the phylogeny (Graham *et al*., 2005, 2011) or on shallower branches (Conti *et al*., 1997; Huang and Shi, 2002; Morris, 2007). Sister relationships and tree topology throughout our plastome-based tree and the ML tree based on the 353 nuclear loci supermatrix (hereafter referred to as the *353* tree; Maurin *et al*., 2021) are generally highly congruent, except for three details. First, in the plastome tree, closely spaced divergences lead to *Decodon* and then *Lythrum* in clade 2A. The *353* ML tree, however, resolves *Decodon* and *Lythrum* as a short-stemmed monophyletic group, as does our ITS tree (Supplementary Data Fig. S3). Second, plastome data place *Lawsonia* as sister to *Ammannia* in a short-stemmed clade, sister to a clade containing *Tetrataxis* and *Ginoria.* The *353* tree and our ITS tree, however, place *Lawsonia* as sister to a clade uniting *Ammannia* with a *Tetrataxis–Ginoria* clade. Earlier chloroplast *rbcL* and *trnL-F* trees place *Ginoria* and *Lawsonia* in a clade, sister to *Ammannia*, though with low support (Graham *et al*., 2005). Third, plastome data place *Physocalymma* and *Lourtella* in a short-stemmed clade sister to *Diplusodon* and then a monophyletic group comprising the remainder of clade 1A genera, in agreement with the ITS tree. The *353* tree, however, places *Diplusodon*, then *Physocalymma* and then *Lourtella* in ascending steps at the base of the equivalent of clade 1A. While the same topology was recovered in both BI and ML analyses, the three discordances involve nodes with < 100 % rapid bootstrap support in the ML analysis (ML-BP), while all nodes were fully supported in the Bayesian analysis (BI-PP). Simulation studies have shown that BI-PP and ML-BP are often strongly correlated but can provide substantially different estimates of support on short internodes or topologies supported by few characters (Alfaro *et al*., 2003). ASTRAL analysis of the *353* data placed low PP values around the *Diplusodon*, *Physocalymma* and *Lourtella* relationships, while

quartet scores of many internal nodes of the tree, including the three nodes discordant with plastome data, were only moderate (Maurin *et al*., 2021).

There are several potential sources of conflict among gene trees, and discordance between nuclear and organellar phylogenies, including hidden paralogy, hybridization, incomplete lineage sorting (ILS) due to rapid radiation, lack of signal due to saturation, recombination and horizontal gene transfer (Smith *et al*., 2015, 2020; Folk *et al*., 2017; Stull *et al*., 2020). For tree reconstruction, these problems have been circumvented by using uniparental markers and nuclear markers that undergo rapid concerted evolution (Linder and Rieseberg, 2004). Hybridization is an important evolutionary process in plants, often accompanied by polyploidy or whole-genome duplication, and is a major source of observed cytonuclear discordance due to chloroplast capture (Stull *et al*., 2020). Considering the Lythraceae plastome – *353* discordances: the *Ammannia–Ginoria* clade is notable in the Lythraceae for its concentration of polyploids, where *Ammannia* is one of the most chromosomally diverse genera in the family and variants of *Lawsonia inermis*, deliberately selected and maintained over centuries of cultivation, have yielded diverse counts ($n = 5$, 12, 15, 16, 17 and 18) (Graham and Cavalcanti, 2001). Additionally, *Ginoria* and *Tetrataxis*, both $n = 28$, are believed to be palaeopolyploids. *Lythrum* is another of the most chromosomally diverse genera in the *Lythraceae* (Graham and Cavalcanti, 2001). In this genus, counts of $n = 5$, 10, 15, 20 and 30 are recorded, and within the heterostylous *Lythrum salicaria* alone, counts of $n = 15$, 25 and 30 occur. Though no chromosome counts have been reported to date in *Lourtella*, the chromosome number in the palaeopolyploid *Diplusodon* ($n = 15$) appears to be a result of descending dysploidy from $n = 16$ (Cavalcanti, 2022). *Physocalymma* has been reported as diploid ($n = 8$) (Graham and Cavalcanti, 2001).

Earliest divergences in the Lythraceae

The Lythraceae presents a fascinating model of evolution of a plant family full of contrasts, with its collection of species-rich and monospecific genera and distributions ranging from highly endemic to near-global. Several Lythracean genera, as exemplified by *Decodon* and *Lythrum*, are estimated herein to be extremely ancient. The monospecific *Decodon*, today endemic to eastern North America (POWO, 2023), has an earlier extensive fossil presence in northern latitudes of the Western Hemisphere and wide representation throughout the Miocene and Pliocene in Europe before the onset of colder climates in the Pliocene and Pleistocene restricted it to its present narrow range (Grímsson *et al*., 2012; Graham, 2013). *Lythrum* with about 35 species, some widely distributed, also has an extensive fossil record in Europe, mainly dating from the middle Miocene to the present. However, remarkable fossil discoveries in North America extend records for both genera to the Campanian of the Late Cretaceous (Grímsson *et al*., 2011). They would have diverged within a Late Cretaceous time frame as predicted by our results (see section 'Agreement of plastome-based divergence estimates with the fossil record').

Wide dispersal and diversity in some genera, together with deep divergences of some clades, present a challenge for estimation of historical biogeographical patterns, possibly resulting in the null estimations for the ancestral ranges of some of the deepest nodes in our analysis. Based on fossil evidence and later DNA sequence evidence, the Myrtales has been frequently given a Gondwanan origin (Raven and Axelrod, 1974; Conti *et al*., 1997; Muller, 1981; Johnson and Briggs, 1984; Sytsma *et al*., 2004). The historical biogeographical analysis of Berger *et al*. (2016), based on a concatenation of six loci *(rbcL*, *ndhF*, *matK*, *matR*, 18S, and 26S), supported a West Gondwanan (South America + Africa) origin of the Myrtales, with a stem age of ~124 Ma. An Early Cretaceous date (~115 Ma) was estimated for the divergence of the Onagraceae and Lythraceae in South America and expansion of the Lythraceae into North America by ~90 Ma. Our plastome-based chronogram is in close agreement with the results of Berger *et al*. (2016), but estimates the expansion of the Lythraceae from the South American to North American continents to have occurred earlier by at least ~95 Ma. Another Myrtales chronogram based on whole plastome data (Zhang *et al*., 2021), estimated a much younger Lythraceae–Onagraceae divergence date and internal Lythraceae divergences, inconsistent with the accepted ages of several Lythraceae fossils. The geographical origin of the Onagraceae–Lythraceae divergence has a majority null probability in our analysis (Fig. 7), but minority probability components are predicted to have wide biogeographical distribution, probably influenced in the analysis by the near-global distributions of the outgroup *Ludwigia*, and several deeply branching genera in the Lythraceae (e.g. *Rotala* and *Lythrum*). The most likely distribution of the Lythraceae crown node is a joint North and South American range (AB), vicariantly leading to disjunct North and South American lineages (Fig. 6; Supplementary Data Fig. S4).

Correlation of plastome-based divergence estimates with the fossil record

There is abundant fossil evidence for the early presence of Lythraceae in North America (reviewed by Graham, 2013), particularly in clade 2A, contributing to the narrower HPD intervals obtained for the date estimates in this clade (Fig. 6). Seeds from the Late Cretaceous Cerro del Pueblo Formation, Coahuila, northern Mexico (~74 Ma; Estrada-Ruiz *et al*., 2009) compare closely with modern *Decodon verticillatus* (L.) Elliott and pollen assigned to *Lythrum* (and the subsumed *Peplis*) has been described from the Late Cretaceous Eagle Formation at Elk Basin, Wyoming (82–81 Ma) and from western Central Siberia (72–69 Ma) (Grímsson *et al*., 2011). Few fossils ascribed to genera of clade 2B have been verified. Pollen from the Eocene to Oligocene of Cameroon under the name *Heterocolpites verrucatus* would fit our date estimates if correctly attributed to *Crenea* (now a synonym of *Ammannia*) (Graham, 2013). Fossil evidence does support our findings on the dates and possible Asian origins of the deeply branching *Trapa–Lagerstroemia* clade, where a number of specimens of the fossilized wood *Sonneratioxylon* Hofmann, attributable to *Sonneratia*, have been described from Danian levels (67–64 Ma) of the Deccan Intertrappean flora of India (Awasthi, 1968; Srivastava and Guleria, 2006). Fossil woods of *Duabangoxylon* Prakash &

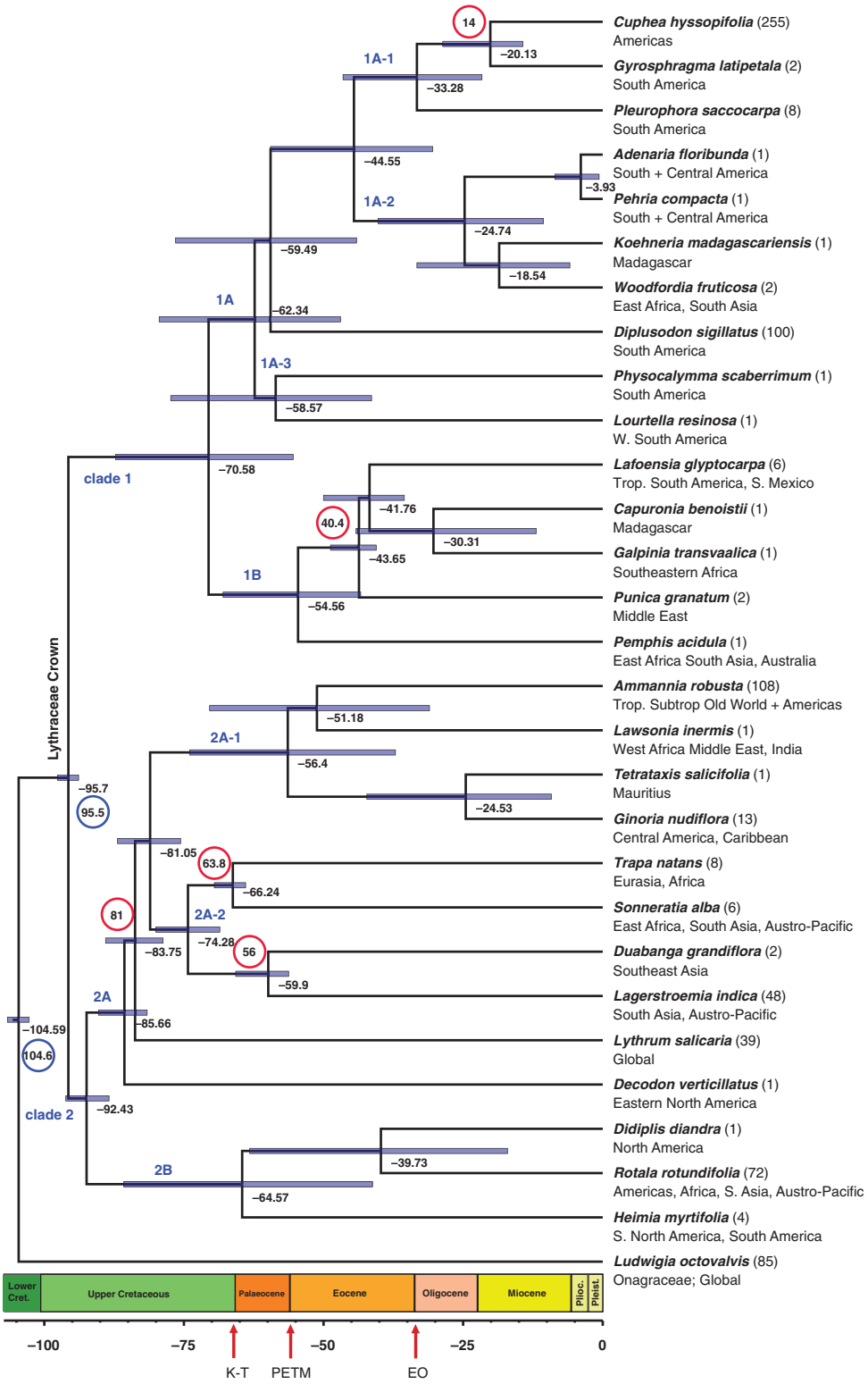

FIG. 6. BEAST chronogram based on plastome (LSU, IRa, SSU) data. Mean divergence times are given for each node, where the 95% HPD is indicated by bars. The scale is in millions of years before the present (Ma). Full posterior support was obtained for all branches. Sampled species names are given, followed by the number of extant species for that genus in parentheses as well as the distribution of that genus (data taken from Kew Plants of the world online – POWO, 2023). Five log-normal calibrated nodes are circled in red and two nodes with normal secondary calibrations are circled in blue.

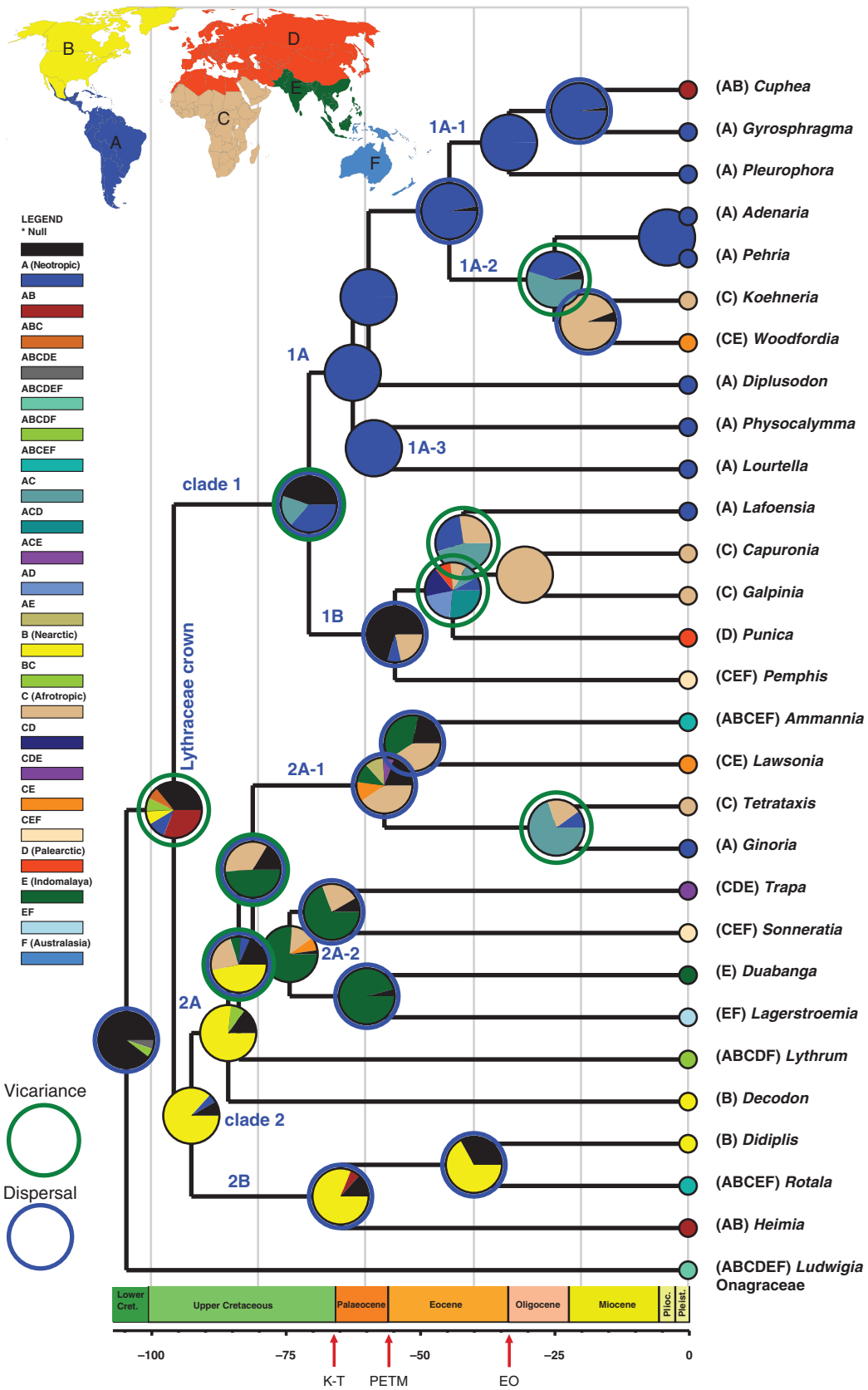

FIG. 7. Ancestral range estimate on the Lythraceae chronogram using BioGeoBEARS (DIVALIKE+J model) in RASP. The scale is in millions of years before the present (Ma) and the six biogeographical areas coded in the analysis are shown on the map inset. Each pie chart contains the likelihood percentage for each estimated area per genus or clade.

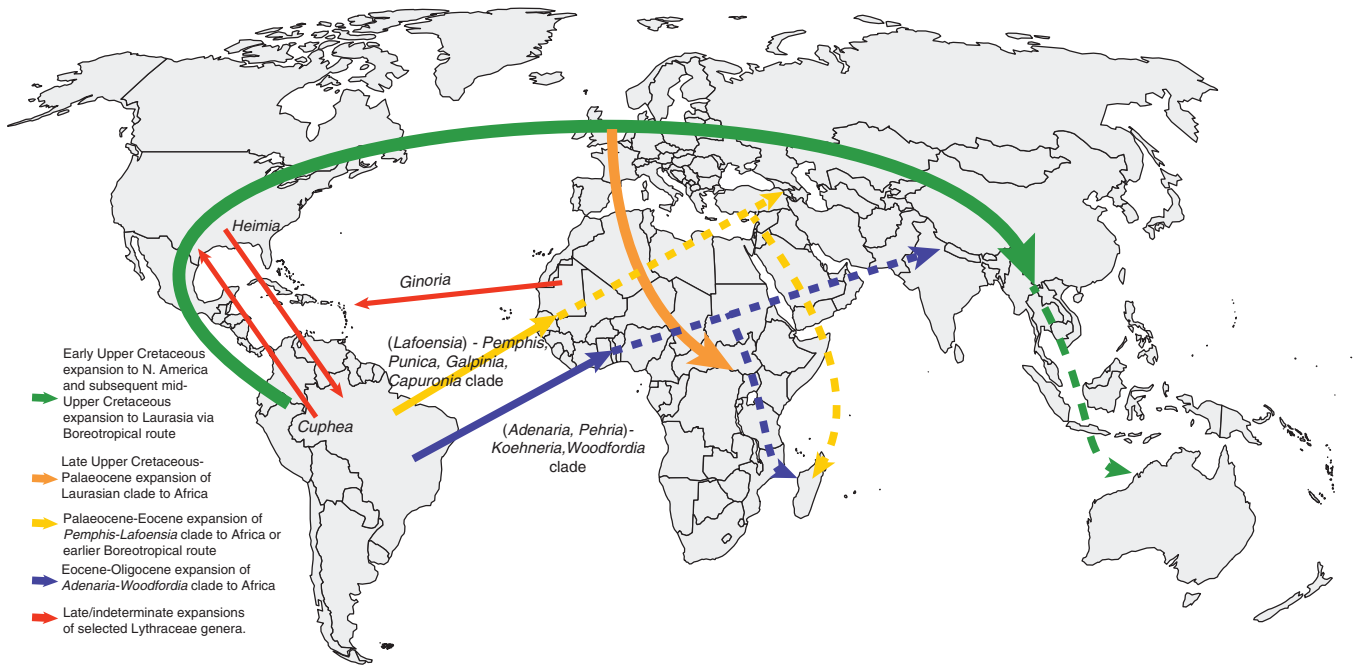

FIG. 8. Hypothesized long-distance expansion routes of Lythraceae clades and selected genera shown on present-day map (excluding Antarctica).

Awasthi, attributed to *Duabanga*, have been described from the Middle Miocene of India (Prakash and Awasthi 1970) and the Miocene of Myanmar (Prakash and Bande, 1980); and fossils attributable to *Lagerstroemia* have been described from the Early Eocene and Palaeocene of India (Lakhanpal and Guleria, 1981). *Trapa* and the extinct genus *Hemitrapa* Miki are aquatic plants with extensive fossil fruit and pollen records in Eurasia and North America and *Trapa* is frequent in Miocene swamp deposits in Europe and Asia (Graham, 2013). Older fossils attributable to *Trapa* and *Hemitrapa* have been identified from the Palaeocene of southernmost Alberta and Saskatchewan and from Alaska, which may represent a disjunct and ancient population of the genus from eastern Siberia (Graham, 2013). Elsewhere in clade 2B, the oldest putative occurrence of *Ammannia* is an unverifiable single fossil seed, *Ammannia lakensis* M.E.J. Chandler, from the early Eocene Pipe-Clay series, southern England (Chandler, 1962), again in agreement with our calibrated reconstruction.

The fossil record attributable to clade 1 Lythraceae is generally sparse and late (Graham, 2013). Rather than palaeobotanical fact, this observation may be affected by regional biases, where the Tertiary plant macrofossil record is strongly biased to North America, Europe and northern Asia, where studies of angiosperm leaf and seed fossils have a long tradition (Morley and Dick, 2013). On the other hand, clade 1 contains a concentration of dry habitat specialist genera, which may have provided fewer opportunities for fossilization compared to the many wetland taxa in clade 2. In geographically diverse subclade 1B (*Pemphis* to *Lafoensia*; Fig. 4), the oldest fossil, attributed to *Punica*, is silicified wood of *Punicoxylon eocenicum* Privé-Gill from the middle Eocene of the Paris Basin (48.6–40.4 Ma; Privé-Gill, 1981), in agreement with our estimated date of divergence of this genus. *Pemphis*-like pollen from the late Eocene (~37–34 Ma) lacustrine beds in the Ebro Basin, Spain, has also been reported (Cavagnetto and Anadón, 1996). The oldest convincing *Cuphea* fossils are early to middle Miocene (~14 Ma) pollen grains from Chiapas, Mexico (Palacios and Rzedowski, 1993), where our biogeographical evidence strongly points to a South American origin of this genus. *Woodfordia* is tentatively represented by a leaf impression from the middle Miocene of lower Siwalik sediments in western Nepal (Prasad, 1993), although our biogeographical evidence suggests an African location for the divergence of this genus and *Koehneria* in the early Miocene. Fossil reports of *Adenaria* are limited to seeds from the Pliocene (~3 Ma) of the high plain of Bogota, Colombia (Wijninga, 1996), close to our estimated date of divergence of this genus in South America.

Routes for dispersal in the Lythraceae

Geological and tectonic models suggest that an arc of volcanic islands between North and South America may have provided a corridor for floristic interchange between the two continents in the Late Cretaceous (Iturralde-Vinent and MacPhee, 1999). However, marine rocks of virtually every stage from the Early Cretaceous to Late Eocene are well developed in the volcanic arc portions of present-day Puerto Rico, Hispaniola, Cuba and Jamaica; older terrestrial indicators were everywhere succeeded by younger marine beds, indicating that any islands which existed were transitory (Macphee and Iturralde-Vinent, 2005). It has been suggested that some plant groups with strongly differentiated components in both tropical North and South America, such as the Cactaceae, might reflect Late Cretaceous or Early Tertiary migration by such island-hopping routes, with subsequent isolation until the closure of the Isthmus of Panama (Gentry, 1982; reviewed in: Pennington and Dick, 2004; Graham, 2010). Seeds washed from South

American rivers may eventually reach Caribbean or North American shores as flotsam, where wind is thought to be much more important for the motion of objects on the surface than sea currents (Iturralde-Vinent and MacPhee 1999). However, transit times are likely to be long, making such sea journeys an unusual method of successful dispersal for many kinds of organisms (Macphee and Iturralde-Vinent, 2005). Since our hypothesized range expansion of the early Lythraceae from South to North American continents and subsequent vicariant establishment of clade 1 and clade 2 lineages (Fig. 7) is estimated to have occurred by ~95 Ma, assumptions of stepping-stone, or transoceanic, LDD are speculative. Several extant Lythraceae genera, such as *Rotala*, *Lythrum*, *Ammannia*, *Sonneratia* and *Pemphis*, include species associated with wetland or coastal habitats, brackish waters and salt-water mangrove habitats, which is likely to have facilitated their dispersal throughout the tropics and temperate zones (Graham *et al.*, 1993). Other species with less salt-tolerant propagules growing in or in proximity to rivers might conceivably also survive marine transit by rafting on detritus, such as driftwood (Thiel and Gutow, 2004).

The Lythraceae chronogram (Figs 6–8) is marked by an estimated Late Cretaceous expansion of clade 2 lineages from North America to Asia, with secondary expansion to Africa. Dispersal of plants and animals between Siberia and north-east Asia to the west and Alaska and the Yukon Territories to the east via the developing Aleutian Islands chain has been widely accepted, supported by fossil remains and the shared presence of the same or similar species (Brikiatis, 2014; Graham, 2018). For questions of when the geographical and climatic conditions would have allowed successful dispersal and establishment and whether by water, land or air, the answer is probably unique to each taxon. The Bering area is thought to have served as a northern dispersal route between the continents since its formation in the Late Cretaceous (~100 Ma), with intermittent periods of marine or land connection between the Arctic and Pacific oceans through the Palaeocene and early Eocene (Sanmartín *et al.*, 2001; Brikiatis, 2014; Graham, 2018). The high latitude (between 69° and 75°N) of the Beringian route, however, may have restricted passage of megathermal lineages for most of this time (Morley, 2003). Two southern routes, the De Geer and Thulean routes, are thought to have provided alternative connections between eastern North America and West Eurasia (reviewed in Brikiatis, 2014). The De Geer route is thought to have existed from the latest Cretaceous to the early Palaeocene, ~71 Ma, while the Thulean route, probably established long after the interruption of the De Geer route, possibly offered a southerly connection between North America and western Europe at least twice, at ~57 and ~56 Ma (Brikiatis, 2014). These times post-date our estimates for earliest dispersals of Lythraceae. North America and Eurasia, however, are thought to have been in tectonic contact via Greenland throughout the Late Cretaceous and Early Tertiary, separated, except at very high latitudes, by a narrow seaway (Morley, 2003). Virtually frost-free climates might have periodically extended well into the mid-latitudes during that time, allowing exchange of terrestrial biota between the North American and Eurasian Boreotropical provinces, as evidenced by the simultaneous appearance of the same pollen types in both regions (Morley, 2003). A southern expansion route through Antarctica for the Lythraceae is not envisaged in our analysis, given the late estimate for occupation of the Australian region (Fig. 7). The early expansion of the Lythraceae mirrors a model used to explain disjunctions in the distribution of Malpighiaceae, where from a South American origin, members of several distinct clades are hypothesized to have dispersed into North America and subsequently moved via North Atlantic land connections into the Old World during Boreotropical episodes starting in the Eocene (Davis *et al.*, 2002). In the Lythraceae, however, we estimate Laurasian expansion to have occurred much earlier in the Late Cretaceous (Fig. 7). Expansion from the Neotropics to Indo-Malaya via North America, facilitated by a Late Cretaceous/Palaeocene South American–North American land bridge, has also been proposed to explain the distribution of the Melastomataceae (Reginato *et al.*, 2022).

The stem of the *Ammannia–Ginoria* subclade in clade 2 is long, and range expansion to Africa from a Laurasian ancestor could date from the Late Cretaceous to as late as the Eocene. However, the node representing divergence of the *Ammannia–Ginoria* clade (2A-1) from the *Trapa–Lagerstroemia* clade (2A-2) also contains a significant African component to its estimated range, so that expansion could have occurred in the earlier portion of this timescale. Both *Sonneratia* and *Trapa* are adapted to aquatic environments: *Sonneratia* is native to mangroves in large portions of East Africa and Madagascar, as well as South Asia, the Indo-Pacific and Northern Australia, whereas *Trapa* is native to freshwater habitats in Eurasia and Africa (Graham, 2007; POWO, 2023). Possible late Cretaceous trans-Tethyan dispersals from Europe to Africa are supported by the appearance of the Laurasian *Normapolles* pollen group in North Africa by this time (Herngreen *et al.*, 1996). This dispersal route was probably severed by the beginning of the Palaeocene, and it was not until the Late Eocene that dispersals of megathermal taxa between Europe and Africa are again recorded (Cavagnetto and Anadón, 1996; Morley, 2003). The *Ammannia–Ginoria* clade is estimated to have begun to diverge around the time of the Palaeocene–Eocene thermal maximum (PETM), probably in Africa (Fig. 7). Soon after, in the early Eocene, one lineage is predicted to have led to the divergence of *Ammannia* and *Lawsonia* in Africa. *Ammannia* seeds are produced in great quantity, often 200–300 or more per capsule (Graham and Graham, 2014). Buoyed by a large aerenchymatous float (Graham, 1985), they are well adapted to dispersal in aquatic environments, facilitating possible transoceanic LDD. Today, certain *Ammannia* species are distributed globally throughout warm latitudes by commercial trade in seed rice contaminated by seed from species of *Ammannia* adventive in rice fields (Caton *et al.*, 1997).

The estimated divergence of *Tetrataxis* and *Ginoria* in the Oligocene constitutes one of the most remarkable vicariant disjunctions in the Lythraceae phylogeny. Monotypic *Tetrataxis* is endemic to Mauritius in the southern Indian Ocean, whereas the 13 species of *Ginoria* are native to southern Mexico and the Antilles (POWO, 2023). A reciprocal Oligocene LDD possibly explains the divergence of the South American *Adenaria–Pehria* clade from the African *Koehneria–Woodfordia* clade (Fig. 7). Complete separation of the South American and African continents is thought to have occurred sometime between ~106 and ~84 Mya, with contact between tropical West Africa with north-eastern South America persisting longest (Pitman *et al.*, 1993; Pletsch *et al.*, 2001). The simultaneous

appearance of (non-Lythraceous) novel pollen types in Africa and South America suggests that putative island chains from the Rio Grande Rise to the Walvis Ridge and the Ceará to Sierra Leone Rises could have facilitated transatlantic dispersal of plants as late as ~76 Ma (Morley, 2003). Dispersal across very wide physical barriers, such as the Atlantic Ocean, has been used to explain disjunctions at species, generic and higher taxonomic levels in multiple taxa (Renner, 2004; Pennington and Dick, 2004). Inferred directions and modes of dispersal can be related to sea currents between Africa and South America and to winds blowing from north-eastern Brazil north-west to Africa (Waters, 2008). More generally, seed morphology and anatomy suggest that dispersal by water (in both directions) appears more common than dispersal by wind or by birds (Renner, 2004). Sweepstakes dispersal by sea, rather than via a northern land bridge, might better explain both wide Eocene/Oligocene disjunctions in the Lythraceae, since these clades have no closely related sister taxa native to North America or Europe, aside from the aforementioned widespread *Ammannia*. The much earlier expansion of the *Lafoensia–Pemphis* clade in the Late Cretaceous or Palaeocene could represent transit via hypothesized mid-Atlantic South American–African island chains (Muellner-Riehl and Rojas-Andrés, 2022), but direct sea crossing seems equally possible. In our analysis, the first extant genus to diverge from this clade in the early Eocene is *Pemphis* (Fig. 6), today native to a coastal arc from West Africa through southern Asia, to northern Australia. A boreotropical route for dispersal of this clade between South America and Africa is not supported by extant North American members of the clade 1B lineage, except for populations of *Lafoensia punicifolia* DC. in southern Mexico and Central America. However, fossil evidence (Graham, 2013) places *Pemphis*-like pollen in the Ebro Basin (modern eastern Spain) from the late Eocene and *Punica*-like wood in the middle Eocene of the Paris Basin, indicating that a northern route for expansion of clade 1B cannot be excluded if the climate in higher latitudes in the Eocene/Oligocene was permissible, followed by range contraction and extinction in those areas. The middle Eocene node representing an estimate for the divergence of *Punica* is complex in our biogeographical analysis, containing disjunct elements from South America and the Old World. The modern south-western Eurasian distribution of *Punica* could therefore either represent a vicariant population that spread north from Africa, or be an *in situ* relict of a Boreotropical/Laurasian population of the 1B lineage that spread to the Old World from South America, as early as the latter part of the Late Cretaceous. Considering the African–European expansion route, the western-most portions of the Tethyan Sill, the Alboran and Apulian routes possibly provided earlier crossing opportunities, but by the Middle Eocene, an eastern Iranian route is thought to have opened, leading to progressive closing of the Tethys Sea in the Middle East area (Gheerbrant and Rage, 2006). The natural range of *Punica* today is likely to have been altered by the ancient to modern cultivation of pomegranate fruits.

Lafoensia represents another remarkable biogeographical disjunction in the Lythraceae. Although Eocene reciprocal LDD from Africa to South America could be invoked to explain the distribution of *Lafoensia*, biogeographical analysis suggests that this genus possibly represents a vicariant lineage in the clade, emerging from a geographically diverse node that includes components of both the ancestral South American and African gene pools. In the same manner, the more recent divergence of the African/Madagascan genera *Capuronia* and *Galpinia* could also result from the same diverse node. *Lafoensia* is an American genus (POWO, 2023) and based on preliminary results, we hypothesize that the genus originated in eastern South America, expanding through the Cerrado into Central America after closure of the Panama isthmus (T. B. Cavalcanti, unpubl. data). *Capuronia* and *Koehneria* are monotypic endemic genera of Madagascar (POWO, 2023) and probably evolved *in situ* in the Oligocene and early Miocene, respectively, from introductions of two distinct East African lineages, possibly dispersing via the Davie Ridge that is thought to have connected, at least partially, Africa and Madagascar across the Mozambique Channel (McCall, 1997; reviewed in Ali and Hedges, 2022). Additionally, during the early development of Madagascar's rain forests in the Oligocene, predominant marine currents are thought to have been eastward, favouring colonization from African biota (Ali and Huber, 2010). A comparable eastward transoceanic Oligocene LDD from yet another Afrotropical lineage may also explain the current isolation of the monotypic *Tetrataxis* on the island of Mauritius.

Seed biology and dispersal in the Lythraceae

Lythraceae have adaptive structures for seed dispersal by water and wind. The seeds of species of several Lythraceae genera possess a single straight or helical trichome, unique among angiosperms, that produces mucilage (Graham and Graham, 2014). Mucilage possibly promotes dispersal by securing the seed to the soil and by maintaining moisture (Teixeira *et al*., 2020; Tsai *et al*., 2021). Of the 13 Lythraceae genera that possess mucilage-secreting seeds, eight occur in hot and dry regions (Graham and Graham, 2014). Mucilage may also protect the seed in the digestive tract after ingestion by animals or by affecting the ability of seeds to sink in water (Tsai *et al*., 2021). Genera that inhabit humid or aquatic environments, such as *Ammannia* and *Rotala*, possess large-celled aerenchyma on the raphal side of the boat-shaped seeds. The seeds of others such as *Decodon*, *Lawsonia*, *Pemphis*, *Sonneratia* and *Tetrataxis* are anatomically adapted for water dispersal by the possession of extensive internal float tissue, while species of *Lythrum* possess mucilaginous seed coat trichomes, also associated with enhanced dispersal (Graham and Graham, 2014). Independent of land bridges, water-mediated LDD of seeds is probably important for range expansion of the Lythraceae, particularly in clade 2 lineages, but possibly also for dispersal of the clade containing *Pemphis*, a coastal inhabitant, in clade 1. Water-mediated dispersal has ptobably contributed to clade 2 genera demonstrating the widest geographical distributions in the Lythraceae (Fig. 6).

For Lythraceae genera forming shrubby and arboreal elements of seasonally dry forest or grassland environments in the tropics, seed dispersal is usually by wind, where the seeds are commonly larger and winged, such as in *Diplusodon*, *Lafoensia* and *Physocalymma* (Graham, 2007; Graham and Graham 2014). Fleshy fruits and seeds are found in *Punica*, *Capuronia* and *Sonneratia*, which are dispersed by animals (Graham and Graham, 2014). Fruit- or seed-eating birds might provide a

mechanism for transoceanic LDD beyond local and regional scales. Although there is no direct evidence in the Lythraceae, by sampling birds caught while in migratory flight, transport of seeds over hundreds of kilometres from mainland to oceanic islands has been demonstrated (Viana *et al*., 2016). Although animal dispersal may be more efficient, niche choice may result in transport to sites ecologically similar to the point of origin, providing less opportunity for niche shifts over macroevolutionary time, compared to random abiotic dispersal (Vasconcelos *et al*., 2021).

Species richness in the Lythraceae

As previously mentioned, the Lythraceae form a remarkable mosaic, containing several highly species-rich genera as well as many particularly species-poor lineages, or *depauperons* (reviewed in Donoghue and Sanderson, 2015), several of which are estimated to be ancient, *Decodon* being an extreme example. In the Melastomataceae, *depauperons* have been associated with restricted niche, small size and limited abiotic dispersal (Bastos *et al*., 2022). No similar generalizations can be made in the Lythraceae, where once more exemplified by *Decodon*, fossil evidence suggests a much wider historical distribution and possible species richness (Grímsson *et al*., 2012; Graham, 2013). Further, several depauperate Lythraceous genera, such as *Decodon*, *Galpinia*, *Pemphis*, *Physocalymma*, *Lourtella* and *Didiplis* are today endemic to significant swathes of continental territory (POWO, 2023). Perhaps more easily understood, the depauperate state of the monospecific island genera *Tetrataxis* (Mauritius), *Koehneria* (Madagascar) and *Capuronia* (Madagascar) may be attractively explained by niche restriction and genetic bottlenecking (Barrett, 1996). Two other monospecific genera, *Adenaria* and *Pehria*, are predicted to have diverged only about 4 million years ago. At the other end of the species-richness spectrum, the BI chronogram suggests that *Diplusodon* is an old genus (Fig. 6), but radiations resulting in around 100 extant species have been demonstrated to be concentrated only in the last 5 million years, during the climatically unstable Pliocene and Pleistocene, when glaciation-related fluctuations are thought to have brought about cycles of expansion and contraction of the Cerrado Biome (Inglis and Cavalcanti, 2018). Proneness to fire and a landscape punctuated by sky island refugia in the form of the *campos rupestres* of eastern Brazil are thought to have stimulated rapid diversification and species richness in diverse lineages occupying the biome (Simon *et al*., 2009; Simon and Pennington, 2012; Vasconcelos *et al*., 2020). *Diplusodon* is sister to a clade containing two similarly old neotropical monospecific genera, *Physocalymma* and *Lourtella*, native to a wet tropical belt from Ecuador to Bolivia and Brazil and to desert or dry shrubland of Peru and Bolivia, respectively, where relative niche age and stability may have contributed to their limited diversification. Sister to *Gyrosphragma*, which has just two species native to the Brazilian Atlantic Rainforest (Cavalcanti *et al*., 2022), *Cuphea* is a younger genus, with a stem dated to ~20 Ma (Fig. 6), but with around 255 extant species found throughout the neotropics and southern North America (POWO, 2023). Such abrupt shifts in divergence rates in the Lythraceae are probably not solely associated with niche expansion (or contraction), but by innovations in breeding systems, life history, polyploidy and dispersal efficiency; factors which merit deeper study across the family.

CONCLUSIONS AND PERSPECTIVES

Our plastome-based phylogenetic reconstruction resolves the Lythraceae as two deeply branching major clades that diverged vicariantly ~10 Ma after the Lythraceae–Onagraceae split, following range expansion from South American to North American continents. Strong backbone support provides a framework for understanding the evolution of the Lythraceae and should form the basis for a much-needed stable taxonomy. Refinement of the model will require additional dated phylogenetic trees, especially among the more species-rich and geographically dispersed Lythraceae genera. Such data are currently available only in *Diplusodon* (Inglis and Cavalcanti, 2018) and *Lagerstroemia* (Dong *et al*., 2021), whose crown ages are reported to be late Pliocene and early Oligocene, respectively, reflecting contrasting biogeographical histories. We hypothesize that current distributions and diversification in the family were shaped not only by continental movements and climate change, but by LDD and vicariance. Transoceanic dispersal in the Lythraceae has possibly been facilitated in certain cases by adaptations to aquatic environments, common to many extant genera, possibly explaining several remarkable disjunctions in the family. Our model of a Late Cretaceous expansion of the North American lineage through Laurasia to Africa via a boreotropical route is afforded much support by the fossil record (Graham, 2013). However, the fossil history of the South American lineage is poor, including the black box that is the period around the Lythraceae–Onagraceae split ~105 Ma and our hypothesized divergence of South American and North American clades, ~95 Ma. Regional biases in the fossil record should be addressed, if possible, in future studies to provide physical evidence for the biogeographical history of the Lythraceae.

SUPPLEMENTARY DATA

Supplementary data are available online at https://academic.oup.com/aob and consist of the following.

Table S1. Voucher information of newly sequenced samples and GenBank numbers of plastomes and ITS sequences. **Table S2.** GenBank numbers of published plastomes and ITS sequences included in analysis. **Table S3.** Fossil and secondary calibrations used in BEAST analysis. **Table S4.** Results of BioGeoBEARS Modeltest. **Table S5.** BBMap library coverage statistics for Lythraceae plastome assemblies. **Table S6.** Physical characteristics and inverted repeat insertion points of Lythraceae plastomes. **Table S7.** Direct and dispersed repeats, and gene content of Lythraceae plastomes. **Table S8.** Full results of MISA scan of Lythraceae plastomes. **File S1.** Additional information for Fig. 6 BioGeoBears/RASP. **Fig. S1.** Progressive Mauve alignment of Lythraceae plastomes. **Fig. S2.** Linear physical maps of annotated Lythraceae plastomes, prepared using OGDRAW. **Fig. S3.** BI consensus phylogram of the Lythraceae, based on ITS data. **Fig. S4.** Most likely ancestral range estimate on the Lythraceae chronogram using BioGeoBEARS (DIVALIKE+J model) in RASP.

FUNDING

P.W.I. was the recipient of a Fundação de Apoio à Pesquisa Agrícola – FUNDAG postdoctoral fellowship during the period of this research.

LITERATURE CITED

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. doi:10.1093/molbev/msg028.

Ali JR, **Hedges SB. 2022**. A review of geological evidence bearing on proposed Cenozoic land connections between Madagascar and Africa and its relevance to biogeography. *Earth-Science Reviews* **232**: 104103. doi:10.1016/j.earscirev.2022.104103.

Ali JR, **Huber M. 2010**. Mammalian biodiversity on Madagascar controlled by ocean currents. *Nature* **463**: 653–656. doi:10.1038/nature08706.

Altschul SF, **Gish W**, **Miller W**, **Myers EW**, **Lipman DJ. 1990**. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–410. doi:10.1016/S0022-2836(05)80360-2.

Awasthi N. 1968. A fossil wood of *Sonneratia* from the Tertiary of South India. *Palaeobotanist* **17**: 254–257. doi:10.54991/jop.1968.800.

Bakker FT, **Lei D**, **Yu J**, ***et al***. **2016**. Herbarium genomics: plastome sequence assembly from a range of herbarium specimens using an iterative organelle genome assembly (IOGA) pipeline. *Biological Journal of the Linnean Society* **117**: 33–43. doi:10.1111/bij.12642.

Bakker FT. 2017. Herbarium genomics: skimming and plastomics from archival specimens. *Webbia* **72**: 35–45. doi:10.1080/00837792.2017.1313383.

Barrett SCH. 1996. The reproductive biology and genetics of island plants. *Philosophical Transactions: Biological Sciences* **351**: 725–733. https://www.jstor.org/stable/56423.

Bastos B, **Bacci LF**, **Reginato M**, **Bochorny T**, **Goldenberg R. 2022**. Limited dispersal ability and restricted niche characterize 'depauperons' in Melastomataceae. *Perspectives in Plant Ecology, Evolution and Systematics* **57**: 125701. doi:10.1016/j.ppees.2022.125701.

Beier S, **Thiel T**, **Münch T**, **Scholz U**, **Mascher M. 2017**. MISA-web: a web server for microsatellite prediction. *Bioinformatics* **33**: 2583–2585. doi:10.1093/bioinformatics/btx198.

Berger BA, **Kriebel R**, **Spalink D**, **Sytsma KJ. 2016**. Divergence times, historical biogeography, and shifts in speciation rates of Myrtales. *Molecular Phylogenetics and Evolution* **95**: 116–136. doi:10.1016/j.ympev.2015.10.001.

Bouckaert R, **Heled J**, **Kühnert D**, ***et al***. **2014**. BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10**: e1003537. doi:10.1371/journal.pcbi.1003537.

Brewer GE, **Clarkson JJ**, **Maurin O**, ***et al***. **2019**. Factors affecting targeted sequencing of 353 nuclear genes from herbarium specimens spanning the diversity of angiosperms. *Frontiers in Plant Science* **10**: 1102. doi:10.3389/fpls.2019.01102.

Brikiatis L. 2014. The De Geer, Thulean and Beringia routes: key concepts for understanding early Cenozoic biogeography. *Journal of Biogeography* **41**: 1036–1054. doi:10.1111/jbi.12310.

Bushnell, B., 2014. *BBMap: a fast, accurate, splice-aware aligner*. Berkeley: Lawrence Berkeley National Lab. (LBNL).

Caton B, **Foin T**, **Hill J. 1997**. Mechanisms of competition for light between rice (*Oryza sativa*) and redstem (*Ammannia* spp.). *Weed Science* **45**: 269–275. doi:10.1017/S0043174500092821.

Cavagnetto C, **Anadón P. 1996**. Preliminary palynological data on floristic and climatic changes during the Middle Eocene-Early Oligocene of the eastern Ebro Basin, northeast Spain. *Review of Palaeobotany and Palynology* **92**: 281–305. doi:10.1016/0034-6667(95)00096-8.

Cavalcanti TB. 2022. The genus *Diplusodon* (Lythraceae). *Flora Neotropica*, Vol. **122**. Cham: Springer, 516. doi:10.1007/978-3-030-65875-5.

Cavalcanti TB, **Facco MG**, **Inglis PW**, **Graham SA**, **Gonella PM. 2022**. A new genus of Lythraceae emerges from rocky outcrops of the Atlantic Forest in Brazil. *Botanical Journal of the Linnean Society* **200**: 85–103. doi:10.1093/botlinnean/boac004.

Chandler MEJ. 1962. The Lower Tertiary floras of southern England. II. *Flora of the pipe-clay series of Dorset (Lower Bagshot)*. London: British Museum (Natural History). x + 176 pp + 29 pl. doi:10.5962/bhl.title.110079

Chen S, **Zhou Y**, **Chen Y**, **Gu J. 2018**. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**: i884–i890. doi:10.1093/bioinformatics/bty560.

Conti E, **Litt A**, **Wilson PG**, ***et al***. **1997**. Interfamilial relationships in Myrtales: molecular phylogeny and patterns of morphological evolution. *Systematic Botany* **22**: 629. doi:10.2307/2419432.

Dahlgren R, **Thorne RF. 1984**. The order Myrtales: circumscription, variation, and relationships. *Annals of the Missouri Botanical Garden* **71**: 633. doi:10.2307/2399158.

Darling AE, **Mau B**, **Perna NT. 2010**. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* **5**: e11147. doi:10.1371/journal.pone.0011147.

Darriba D, **Posada D**, **Kozlov AM**, **Stamatakis A**, **Morel B**, **Flouri T. 2020**. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Molecular Biology and Evolution* **37**: 291–294. doi:10.1093/molbev/msz189.

Davis CC, **Bell CD**, **Mathews S**, **Donoghue MJ. 2002**. Laurasian migration explains Gondwanan disjunctions: Evidence from Malpighiaceae. *Proceedings of the National Academy of Sciences* **99**: 6833–6837. doi:10.1073/pnas.102175899.

Dong W, **Xu C**, **Liu Y**, **Shi J**, **Li W**, **Suo Z. 2021**. Chloroplast phylogenomics and divergence times of Lagerstroemia (Lythraceae). *BMC Genomics* **22**: 434. doi:10.1186/s12864-021-07769-x.

Donoghue MJ, **Sanderson MJ. 2015**. Confluence, synnovation, and depauperons in plant diversification. *New Phytologist* **207**: 260–274. doi:10.1111/nph.13367.

Douglas J, **Zhang R**, **Bouckaert R. 2021**. Adaptive dating and fast proposals: Revisiting the phylogenetic relaxed clock model. *PLoS Computational Biology* **17**: e1008322. doi:10.1371/journal.pcbi.1008322.

Drummond AJ, **Ho SYW**, **Phillips MJ**, **Rambaut A. 2006**. Relaxed phylogenetics and dating with confidence. *PLoS Biology* **4**: e88. doi:10.1371/journal.pbio.0040088.

Drummond AJ, **Suchard MA**, **Xie D**, **Rambaut A. 2012**. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**: 1969–1973. doi:10.1093/molbev/mss075.

Edler D, **Klein J**, **Antonelli A**, **Silvestro D. 2020**. raxmlGUI 2.0: A graphical interface and toolkit for phylogenetic analyses using RAxML. *Methods in Ecology and Evolution* **12**: 373–377. doi:10.1111/2041-210X.13512.

Estrada-Ruiz E, **Calvillo-Canadell L**, **Cevallos-Ferriz SRS. 2009**. Upper Cretaceous aquatic plants from Northern Mexico. *Aquatic Botany* **90**: 282–288. doi:10.1016/j.aquabot.2008.11.004.

Fan X, **Wang W**, **Wagutu GK**, **Li W**, **Li X**, **Chen Y. 2022**. Fifteen complete chloroplast genomes of *Trapa* species (Trapaceae): insight into genome structure, comparative analysis and phylogenetic relationships. *BMC Plant Biology* **22**: 230. doi:10.1186/s12870-022-03608-7.

Flora e Funga do Brasil. **2023**. *Jardim Botânico do Rio de Janeiro*. Available at: http://floradobrasil.jbrj.gov.br/ (Accessed: 10 January 2023).

Folk RA, **Mandel JR**, **Freudenstein JV. 2017**. Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of angiosperms. *Systematic Biology* **66**: 320. doi:10.1093/sysbio/syw083.

Gansauge M-T, **Meyer M. 2013**. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols* **8**: 737–748. doi:10.1038/nprot.2013.038.

GBIF.org. **2023**. *GBIF Occurrence Download* (accessed 7 January 2023). doi:10.15468/dl.zmndfq

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. doi:10.2307/2399084.

Gheerbrant E, **Rage J-C. 2006**. Paleobiogeography of Africa: how distinct from Gondwana and Laurasia? *Palaeogeography, Palaeoclimatology, Palaeoecology* **241**: 224–246. doi:10.1016/j.palaeo.2006.03.016.

Gitzendanner MA, **Soltis PS**, **Yi T-S**, **Li D-Z**, **Soltis DE. 2018**. Plastome phylogenetics: 30 years of inferences into plant evolution. *Advances in Botanical Research* **85**: 293–313. doi:10.1016/bs.abr.2017.11.016.

Gonçalves DJP, **Shimizu GH**, **Ortiz EM**, **Jansen RK**, **Simpson BB. 2020**. Historical biogeography of Vochysiaceae reveals an unexpected

perspective of plant evolution in the Neotropics. *American Journal of Botany* **107**: 1004–1020. doi:10.1002/ajb2.1502.

Graham A. 2010. Late Cretaceous and Cenozoic history of Latin American vegetation and terrestrial environments. *Monographs in Systematic Botany from the Missouri Botanical Garden* **113**. St. Louis, MO: Missouri Botanical Garden Press. ISBN 10: 1930723687/ ISBN 13: 9781930723689

Graham A. 2018. *Land Bridges. Ancient environments, plant migrations, and New World connections*. Cambridge: University of Chicago Press. doi:10.7208/chicago/9780226544328.001.0001

Graham SA. 1985. A revision of Ammannia (Lythraceae) in the Western Hemisphere. *Journal of the Arnold Arboretum* **66**: 395–420. doi:10.5962/bhl.part.13185.

Graham SA. 2007. Lythraceae. In: **Kubitzki K**, ed. *Flowering Plants · Eudicots. The families and genera of vascular plants*, **vol 9**. Berlin: Springer. doi:10.1007/978-3-540-32219-1_30

Graham SA. 2013. Fossil records in the Lythraceae. *The Botanical Review* **79**: 48–145. doi:10.1007/s12229-012-9116-1.

Graham SA. 2017. A revision of *Cuphea* section Brachyandra s. s. (Lythraceae). *Systematic Botany* **42**: 859–919. doi:10.1600/036364417x696528.

Graham SA. 2019. A revision of *Cuphea* Section Melvilla (Lythraceae). *Systematic Botany Monographs* **105**: 1–146. ISBN: 9781943751051

Graham SA, **Cavalcanti TB. 2001**. New chromosome counts in the Lythraceae and a review of chromosome numbers in the family. *Systematic Botany* **26**: 445–458. doi:10.1043/0363-6445-26.3.445.

Graham SA, **Crisci JV**, **Hoch PC. 1993**. Cladistic analysis of the Lythraceae *sensu lato* based on morphological characters. *Botanical Journal of the Linnean Society* **113**: 1–33. doi:10.1111/j.1095-8339.1993.tb00326.x.

Graham SA, **Graham A. 2014**. Ovary, fruit, and seed morphology of the Lythraceae. *International Journal of Plant Sciences* **175**: 202–240. doi:10.1086/674316.

Graham SA, **Diazgranados M**, **Barber JC. 2011**. Relationships among the confounding genera *Ammannia*, *Hionanthera*, *Nesaea* and *Rotala* (Lythraceae). *Botanical Journal of the Linnean Society* **166**: 1–19. doi:10.1111/j.1095-8339.2011.01126.x. doi:10.1111/j.1095-8339.2011.01126.x.

Graham SA, **Hall J**, **Sytsma K**, **Shi S. 2005**. Phylogenetic analysis of the Lythraceae based on four gene regions and morphology. *International Journal of Plant Sciences* **166**: 995–1017. doi:10.1086/432631.

Graham SA, **Inglis PW**, **Cavalcanti TB. 2021**. The phylogenetic position of *Crenea* in the Lythraceae based on molecular evidence, and the transfer of its two species to *Ammannia*. *Annals of the Missouri Botanical Garden* **106**: 325–339. doi:10.3417/2021629.

Greiner S, **Lehwark P**, **Bock R. 2019**. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research* **47**: W59–W64. doi:10.1093/nar/gkz238.

Grímsson F, **Zetter R**, **Hofmann C-C. 2011**. *Lythrum* and *Peplis* from the Late Cretaceous and Cenozoic of North America and Eurasia: New evidence suggesting early diversification within the Lythraceae. *American Journal of Botany* **98**: 1801–1815. doi:10.3732/ajb.1100204.

Grímsson F, **Ferguson DK**, **Zetter R. 2012**. Morphological trends in the fossil pollen of *Decodon* and the paleobiogeographic history of the genus. *International Journal of Plant Sciences* **173**: 297–317. doi:10.1086/663968.

Gu C, **Ma L**, **Wu Z**, **Chen K**, **Wang Y. 2019**. Comparative analyses of chloroplast genomes from 22 Lythraceae species: inferences for phylogenetic relationships and genome evolution within Myrtales. *BMC Plant Biology* **19**: 218. doi:10.1186/s12870-019-1870-3.

Guyeux C, **Charr J-C**, **Tran HTM**, ***et al***. **2019**. Evaluation of chloroplast genome annotation tools and application to analysis of the evolution of coffee species. *PLoS One* **14**: e0216347. doi:10.1371/journal.pone.0216347.

Heath TA, **Huelsenbeck JP**, **Stadler T. 2014**. The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences* **111**: E2957–E2966. doi:10.1073/pnas.1319091111.

Herngreen GFW, **Kedves M**, **Rovnina LV**, **Smirnova SB. 1996**. Cretaceous palynofloral provinces: a review. In: **Jansonius J**, **McGregor DC**, eds. *Palynology: principles and applications*, **vol. 3**. Dallas: American Association of Stratigraphic Palynologists, 1157–1188.

Ho SYW. 2014. The changing face of the molecular evolutionary clock. *Trends in Ecology & Evolution* **29**: 496–503. doi:10.1016/j.tree.2014.07.004.

Hörandl E, **Appelhans MS. 2015**. *Next generation sequencing in plant systematics. (Regnum Vegetabile, 158)*. Königstein: Koeltz Scientific Books.

Huang Y, **Shi S. 2002**. Phylogenetics of Lythraceae *sensu lato*: a preliminary analysis based on chloroplast *rbc*L Gene, *psa*A-*ycf*3 spacer, and nuclear rDNA internal transcribed spacer (ITS) sequences. *International Journal of Plant Sciences* **163**: 215–225. doi:10.1086/338392.

Huelsenbeck JP, **Larget B**, **Alfaro ME. 2004**. Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Molecular Biology and Evolution* **21**: 1123–1133. doi:10.1093/molbev/msh123.

Inglis PW, **Cavalcanti TB. 2018**. A molecular phylogeny of the genus *Diplusodon* (Lythraceae), endemic to the campos rupestres and cerrados of South America. *Taxon* **67**: 66–82. doi:10.12705/671.5.

Inglis PW, **Pappas M de CR**, **Resende LV**, **Grattapaglia D. 2018**. Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. *PLoS One* **13**: e0206085. doi:10.1371/journal.pone.0206085.

Iturralde-Vinent MA, **Macphee RDE. 1999**. Paleogeography of the Caribbean region: implications for Cenozoic biogeography. *Bulletin of the American Museum of Natural History* **238**: 1–95. http://hdl.handle.net/2246/1642.

Jin J-J, **Yu W-B**, **Yang J-B**, ***et al***. **2020**. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology* **21**: 241. doi:10.1186/s13059-020-02154-5.

Johnson LAS, **Briggs BG. 1984**. Myrtales and Myrtaceae: a phylogenetic analysis. *Annals of the Missouri Botanical Garden* **71**: 700–756. doi:10.2307/2399159.

Kalyaanamoorthy S, **Minh BQ**, **Wong TKF**, **von Haeseler A**, **Jermiin LS. 2017**. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**: 587–589. doi:10.1038/nmeth.4285.

Katoh K, **Standley DM. 2013**. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772–780. doi:10.1093/molbev/mst010.

Koehne E. 1881. Lythraceae monographice describuntur. *Botanische Jahrbücher fur Systematik, Pflanzengeschichte und Pflanzengeographie* **1**: 436–458.

Koehne E. 1886. Dir Geographische Verbreitung der Lythraceen. *Botanische Jahrbücher fur Systematik, Pflanzengeschichte und Pflanzengeographie* **7**: 1–61.

Koehne, E. 1903. Lythraceae. In: **Engler A**, ed. *Das Pflanzenreich IV. 216*. Leipzig: Wilhelm Engelmann, 1–326.

Kurtz S, **Choudhuri J V**, **Ohlebusch E**, **Schleiermacher C**, **Stoye J**, **Giegerich R. 2001**. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research* **29**: 4633–4642. doi:10.1093/nar/29.22.4633.

Lakhanpal RN, **Guleria JS. 1981**. Leaf-impressions from the Eocene of Kachchh, Western India. *Journal of Palaeosciences* **28**: 353–373. doi:10.54991/jop.1981.1423.

Lanfear R, **Calcott B**, **Ho SYW**, **Guindon S. 2012**. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* **29**: 1695–1701. doi:10.1093/molbev/mss020.

Lehwark P, **Greiner S. 2019**. GB2sequin - A file converter preparing custom GenBank files for database submission. *Genomics* **111**: 759–761. doi:10.1016/j.ygeno.2018.05.003.

Linder CR, **Rieseberg LH. 2004**. Reconstructing patterns of reticulate evolution in plants. *American Journal of Botany* **91**: 1700–1708. doi:10.3732/ajb.91.10.1700.

Macphee RDE, **Iturralde-Vinent M. 2005**. The interpretation of Caribbean paleogeography: reply to Hedges. In: Alcover JA, Bover P, eds. Proceedings of the International Symposium 'Insular Vertebrate Evolution: The Palaeontological Approach'. *Monografies de la Societat d'Història Natural de les Balears* **12**: 175–184. ISBN 84-609-6472-8

Mai DH. 1996. Zwei neue Arten von Samen aus dem deutschen Jungtertiär. *Feddes Repertorium* **107**: 299–303. doi:10.1002/fedr.4921070505.

Matzke NJ. 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in Island Clades. *Systematic Biology* **63**: 951–970. doi:10.1093/sysbio/syu056.

Maurin O, **Anest A**, **Bellot S**, ***et al***. **2021**. A nuclear phylogenomic study of the angiosperm order Myrtales, exploring the potential and limitations of the universal Angiosperms353 probe set. *American Journal of Botany* **108**: 1087–1111. doi:10.1002/ajb2.1699.

McCall RA. 1997. Implications of recent geological investigations of the Mozambique Channel for the mammalian colonization of Madagascar. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **264**: 663–665. doi:10.1098/rspb.1997.0094.

Millen RS, **Olmstead RG**, **Adams KL**, ***et al***. **2001**. Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell* **13**: 645–658. doi:10.1105/tpc.13.3.645.

Mirarab S, **Warnow T. 2015**. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* **31**: i44–i52. doi:10.1093/bioinformatics/btv234.

Morley RJ. 2003. Interplate dispersal paths for megathermal angiosperms. *Perspectives in Plant Ecology, Evolution and Systematics* **6**: 5–20. doi:10.1078/1433-8319-00039.

Morley RJ, **Dick CW. 2013**. Missing fossils, molecular clocks, and the origin of the Melastomataceae. *American Journal of Botany* **90**: 1638–1644. doi:10.3732/ajb.90.11.1638.

Morris J. 2007. *A molecular phylogeny of the Lythraceae and inference of the evolution of heterostyly*. Ph.D. Dissertation, Kent State University, OH.

Muellner-Riehl AN, **Rojas-Andrés BM. 2022**. Biogeography of Neotropical Meliaceae: geological connections, fossil and molecular evidence revisited. *Brazilian Journal of Botany* **45**: 527–543. doi:10.1007/s40415-021-00770-4.

Muller J. 1981. Fossil pollen records of extant angiosperms. *The Botanical Review* **47**: 1–142. doi:10.1007/bf02860537.

Natural Earth. 2023. *Free vector and raster map data at 1:10m, 1:50m, and 1:110m scales*. Available online: https://www.naturalearthdata.com/ (accessed on 10 January 2023).

Nguyen L-T, **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. doi:10.1093/molbev/msu300.

Palacios CR, **Rzedowski J. 1993**. Estudio palinológico de las floras fósiles del Mioceno Inferior y principios del Mioceno Medio de la región de Pichucalco, Chiapas, México. *Acta Botanica Mexicana* **24**: 1–96. doi:10.21829/abm24.1993.677.

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. doi:10.1098/rstb.2004.1532.

Pitman WC, **Cande S**, **LaBrecque J**, **Pindell J. 1993**. Fragmentation of Gondwana: the separation of Africa from South America. In: **Goldblatt P**, ed. *Biological relationships between Africa and South America*. New Haven, CT: Yale University Press, 15–34. doi:10.2307/j.ctt22726mc.6

Pletsch T, **Erbacher J**, **Holbourn AEL**, ***et al***. **2001**. Cretaceous separation of Africa and South America: the view from the West African margin (ODP Leg 159). *Journal of South American Earth Sciences* **14**: 147–174. doi:10.1016/s0895-9811(01)00020-7.

POWO. **2023**. *Plants of the World Online*. Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet; http://www.plantsoftheworldonline.org/. Retrieved 10 January 2023.

Prakash U, **Awasthi N. 1970**. Fossil woods from the Tertiary of eastern India. 1. *Palaeobotanist* **18**: 32–44. doi:10.54991/jop.1969.816.

Prakash U, **Bande MB. 1980**. Some more fossil woods from the Tertiary of Burma. *Palaeobotanist* **26**: 261–278. doi:10.54991/jop.1977.1071.

Prasad M. 1993. Siwalik (Middle Miocene) woods from the Kalagarh area in the Himalayan foot hills and their bearing on palaeoclimate and phytogeography. *Review of Palaeobotany and Palynology* **76**: 49–82. doi:10.1016/0034-6667(93)90080-e.

Privé-Gill C. 1981. Quelques bois de Dicotylédones éocènes (Lutétien supérieur) du Bassin de Paris. *Palaeontographica Abteilung B* **177**: 119–135.

QGIS Development Team. **2023**. *QGIS Geographic Information System*. Open Source Geospatial Foundation. Available online: http://www.qgis.org (accessed on 10 January 2023).

R Core Team. **2023**. *R: A Language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Rambaut A, **Suchard MA**, **Xie D**, **Drummond AJ. 2014**. *Tracer v1.6*. http://beast.bio.ed.ac.uk/Tracer.

Raven PH, **Axelrod DI. 1974**. Angiosperm biogeography and past continental movements. *Annals of the Missouri Botanical Garden* **61**: 539. doi:10.2307/2395021.

Reginato M, **Almeda F**, **Michelangeli FA**, ***et al.*** **2022**. Historical biogeography of the Melastomataceae. In: **Goldenberg R**, **Michelangeli FA**, **Almeda F**. eds. *Systematics, evolution, and ecology of melastomataceae*. Cham: Springer. doi:10.1007/978-3-030-99742-7_4

Renner S. 2004. Plant dispersal across the tropical Atlantic by wind and sea currents. *International Journal of Plant Sciences* **165**: S23–S33. doi:10.1086/383334.

Ronquist F, **Teslenko M**, **van der Mark P**, ***et al***. **2012**. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model selection across a large model space. *Systematic Biology* **61**: 539–542. doi:10.1093/sysbio/sys029.

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. doi:10.1006/bijl.2001.0542.

Shaw J, **Lickey EB**, **Beck JT**, ***et al***. **2005**. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* **92**: 142–166. doi:10.3732/ajb.92.1.142.

Shi S, **Huang Y**, **Tan F**, **He X**, **Boufford DE. 2000**. Phylogenetic analysis of the Sonneratiaceae and its relationship to Lythraceae based on ITS sequences of nrDNA. *Journal of Plant Research* **113**: 253–258. doi:10.1007/pl00013926.

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* **106**: 20359–20364. doi:10.1073/pnas.0903410106.

Simon MF, **Pennington T. 2012**. Evidence for adaptation to fire regimes in the tropical savanna of the Brazilian Cerrado. *International Journal of Plant Sciences* **173**: 711–723. doi:10.1086/665973.

Smith SA, **Moore MJ**, **Brown JW**, **Yang Y. 2015**. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evolutionary Biology* **15**: 150. doi:10.1186/s12862-015-0423-0.

Smith SA, **Walker-Hale N**, **Walker JF**, **Brown JW. 2020**. Phylogenetic conflicts, combinability, and deep phylogenomics in plants. *Systematic Biology* **69**: 579–592. doi:10.1093/sysbio/syz078.

Srivastava R, **Guleria JS. 2006**. *A catalogue of Cenozoic (Tertiary) plant megafossils from India (1989-2005)*. Lucknow: Diamond Jubilee Special Publication, Birbal Sahni Institute of Palaeobotany. ISBN 81-86382-09-7

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313. doi:10.1093/bioinformatics/btu033.

Stettler JM, **Stevens MR**, **Meservey LM**, ***et al***. **2021**. Improving phylogenetic resolution of the Lamiales using the complete plastome sequences of six *Penstemon* species. *PLoS One* **16**: e0261143. doi:10.1371/journal.pone.0261143.

Stevens PF. 2000. Botanical systematics 1950-2000: change, progress, or both? *Taxon* **49**: 635–659. doi:10.2307/1223969.

Straub SC, **Parks M**, **Weitemier K**, **Fishbein M**, **Cronn RC**, **Liston A. 2012**. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. *American Journal of Botany* **99**: 349–364. doi:10.3732/ajb.1100335.

Stull GW, **Soltis PS**, **Soltis DE**, **Gitzendanner MA**, **Smith SA. 2020**. Nuclear phylogenomic analyses of asterids conflict with plastome trees and support novel relationships among major lineages. *American Journal of Botany* **107**: 790–805. doi:10.1002/ajb2.1468.

Sytsma KJ, **Litt A**, **Zjhra ML**, ***et al***. **2004**. Clades, clocks, and continents: historical and biogeographical analysis of Myrtaceae, Vochysiaceae, and relatives in the Southern Hemisphere. *International Journal of Plant Sciences* **165**: S85–S105. doi:10.1086/421066.

Teixeira A, **Iannetta P**, **Binnie K**, **Valentine TA**, **Toorop P. 2020**. Myxospermous seed-mucilage quantity correlates with environmental gradients indicative of water-deficit stress: *Plantago* species as a model. *Plant and Soil* **446**: 343–356. doi:10.1007/s11104-019-04335-z.

Thiel M, **Gutow L. 2004**. The ecology of rafting in the marine environment. I: The floating substrata. In: **Gibson RN**, **Atkinson RJA**, **Gordon JD**, eds. *Oceanography and Marine Biology: An Annual Review,* Vol. **24**.Boca Raton: Taylor & Francis, 181–264. doi:10.1201/9780203507810.

Tillich M, **Lehwark P**, **Pellizzer T**, ***et al***. **2017**. GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* **45**: W6–W11. doi:10.1093/nar/gkx391.

Tsai AY-L, **McGee R**, **Dean GH**, **Haughn GW**, **Sawa S. 2021**. Seed mucilage: biological functions and potential applications in biotechnology. *Plant and Cell Physiology* **62**: 1847–1857. doi:10.1093/pcp/pcab099.

Udvardy MDF. 1975. *A classification of the biogeographical provinces of the world. IUCN Occasional Paper 18*. Morges, Switzerland: IUCN. doi:10.1017/S0376892900019226

Vasconcelos TNC, **Proença CEB**, **Ahmad B**, ***et al***. **2017**. Myrteae phylogeny, calibration, biogeography and diversification patterns: Increased understanding in the most species rich tribe of Myrtaceae. *Molecular Phylogenetics and Evolution* **109**: 113–137. doi:10.1016/j.ympev.2017.01.002.

Vasconcelos TNC, **Alcantara S**, **Andrino CO**, ***et al***. **2020**. Fast diversification through a mosaic of evolutionary histories characterizes the endemic flora of ancient Neotropical mountains. *Proceedings of the Royal Society B: Biological Sciences* **287**: 20192933. doi:10.1098/rspb.2019.2933.

Vasconcelos T, **Boyko JD**, **Beaulieu JM. 2021**. Linking mode of seed dispersal and climatic niche evolution in flowering plants. *Journal of Biogeography* **50**: 43–56. doi:10.1111/jbi.14292.

Viana DS, **Gangoso L**, **Bouten W**, **Figuerola J. 2016**. Overseas seed dispersal by migratory birds. *Proceedings of the Royal Society B: Biological Sciences* **283**: 20152406. doi:10.1098/rspb.2015.2406.

Wang J, **He W**, **Liao X**, ***et al***. **2023**. Phylogeny, molecular evolution, and dating of divergences in *Lagerstroemia* using plastome sequences. *Horticultural Plant Journal* **9**: 345–355. doi:10.1016/j.hpj.2022.06.005.

Waters JM. 2008. Driven by the West Wind Drift? A synthesis of southern temperate marine biogeography, with new directions for dispersalism. *Journal of Biogeography* **35**: 417–427. doi:10.1111/j.1365-2699.2007.01724.x.

Wen J, **Harris A**, **Kalburgi Y**, ***et al***. **2018**. Chloroplast phylogenomics of the New World grape species (*Vitis*, Vitaceae). *Journal of Systematics and Evolution* **56**: 297–308. doi:10.1111/jse.12447.

Wicke S, **Schneeweiss GM**, **dePamphilis CW**, **Müller KF**, **Quandt D. 2011**. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology* **76**: 273–297. doi:10.1007/s11103-011-9762-4.

Wijninga VM. 1996. A Pliocene *Podocarpus* forest mire from the area of the high plain of Bogotá (Cordillera Oriental, Colombia). *Review of Palaeobotany and Palynology* **92**: 157–205. doi:10.1016/0034-6667(95)00101-8.

Wu P, **Xu C**, **Chen H**, **Yang J**, **Zhang X**, **Zhou S. 2021**. NOVOWrap: An automated solution for plastid genome assembly and structure standardization. *Molecular Ecology Resources* **21**: 2177–2186. doi:10.1111/1755-0998.13410.

Yu Y, **Blair C**, **He X. 2020**. RASP 4: Ancestral state reconstruction tool for multiple genes and characters. *Molecular Biology and Evolution* **37**: 604–606. doi:10.1093/molbev/msz257.

Zhang X-F, **Landis JB**, **Wang H-X**, **Zhu Z-X**, **Wang H-F. 2021**. Comparative analysis of chloroplast genome structure and molecular dating in Myrtales. *BMC Plant Biology* **21**: 219. doi:10.1186/s12870-021-02985-9.

Zizka A, **Azevedo J**, **Leme E**, ***et al***. **2019**. Biogeography and conservation status of the pineapple family (Bromeliaceae). *Diversity and Distributions* **26**: 183–195. doi:10.1111/ddi.13004.