Dispersal largely explains the Gondwanan distribution of the ancient tropical clusioid plant clade

Brad R. Ruhfel1,5,6, Claudia P. Bove1, C. Thomas Philbrick4, and Charles C. Davis5

PREMISE OF THE STUDY: The clusioid clade (Malpighiales) has an ancient fossil record (~90 Ma) and extant representatives exhibit a pantropical distribution represented on all former Gondwanan landmasses (Africa, Australia, India, Madagascar, and South America) except Antarctica. Several biogeographers have hypothesized that the clusioid distribution is an example of Gondwanan vicariance. Our aim is to test the hypothesis that the modern distribution of the clusioid clade is largely explained by Gondwanan fragmentation.

METHODS: Using a four gene, 207-taxon data set we simultaneously estimated the phylogeny and divergence times of the clusioid clade using a Bayesian Markov chain Monte Carlo approach. Ancestral Area Reconstructions (AARs) were then conducted on a distribution of 1000 trees and summarized on a reduced phylogeny.

KEY RESULTS: Divergence time estimates and AARs revealed only two or four cladogenic events that are potentially consistent with Gondwanan vicariance, depending on the placement of the ancient fossil *Paleclusia*. In contrast, dispersal occurred on > 25% of the branches, indicating the current distribution of the clade likely reflects extensive recent dispersal during the Cenozoic (~65 Ma), most of which occurred after the beginning of the Eocene (~56 Ma).

CONCLUSIONS: These results support growing evidence that suggests many traditionally recognized angiosperm clades (families and genera) are too young for their distributions to have been influenced strictly by Gondwanan fragmentation. Instead, it appears that corridors of dispersal may be the best explanation for numerous angiosperm clades with Gondwanan distributions.

KEY WORDS Bonnetiaceae; Calophyllaceae; Clusiaceae; dispersal; Gondwana; Hypericaceae; Malpighiales; Podostemaceae; vicariance

Tropical forests contain tremendous angiosperm diversity (Morley, 2000) and are included in 15 of the 25 biodiversity hotspots of the world (Myers et al., 2000). When this diversity arose, and how it has been maintained, are a central focus of ecology and evolutionary biology. Principally important in these ideas is determining the place and time of origin of large, ecologically important, tropical clades. However, detailed studies of such clades are relatively few. To better understand the assembly of the tropical forest biome, and determine the roles of ancient vicariance vs. more recent dispersal, we need to elucidate the timing and origin of major plant clades that inhabit these regions (Pennington and Dick, 2004).

The extant tropics are especially well-represented on the former fragments of Gondwana, (Africa, Australia, India, Madagascar, and South America). Numerous tropical plant clades share a Gondwanan distribution pattern, and the separation of these landmasses has been used to explain the distribution of many such groups (e.g., Raven and Axelrod, 1974). However, fragmentation of Gondwana began in the Jurassic (McLoughlin, 2001; Upchurch, 2008), well before many angiosperm lineages existed (Bell et al., 2010; Magallón et al., 2015). Some portions of Gondwana, however, retained connections until more recently, thus allowing for the possibility that more recent separations influenced angiosperm diversification since the Upper Cretaceous. For example, South America and Africa separated between 110-80 Ma and the final connections between South America, Antarctica, and Australia were severed between 55-41 Ma (McLoughlin, 2001; Sanmartín, 2002; Scher and Martin, 2006; Upchurch, 2008; Beaulieu et al., 2013).
An emerging paradigm regarding many pantropical distributions is that relatively few angiosperm clades exhibit ages consistent with Gondwanan time frames and that current distributions are more likely the result of long-distance, transoceanic dispersal or dispersal through land connections in the Northern hemisphere (e.g., Renner et al., 2001; Davis et al., 2002; Davis et al., 2004; Tripp and McDade, 2014). The largely southern temperate clades Proteaceae and Nothofagaceae are classic examples of Gondwanan vicariance (Manos, 1993; Weston and Crisp, 1994). However, more recent divergence time estimates of even these groups suggest that strict vicariance scenarios are overly simplistic. Instead, initial Gondwanan vicariance, combined with recent dispersal, appears to have played a key role in facilitating their distributions (Knapp et al., 2005; Barker et al., 2007). A similar combination of vicariance and recent dispersal appears to be true in some tropical clades with Gondwanan distributions, including Centropetalaceae (Malpighiaceae, Elatinaceae); Hernandiaceae, Monimiaceae, and Myrtaceae (Sytma et al., 2004; Michalak et al., 2010; Renner et al., 2010; Cai et al., in press). The paucity of examples in which the breakup of Gondwana explains the distributions of tropical angiosperm clades may be due in part to the concentration on formal taxonomic groups (Beaulieu et al., 2013), which in many instances are far too young to be consistent with Gondwanan vicariance (Bell et al., 2010). Furthermore, extinction combined with prevalent dispersal and large geographic areas may hinder our ability to clearly detect ancient biogeographical patterns in angiosperm distributions (Pennington and Dick, 2004; Upchurch, 2008; Clayton et al., 2009; Beaulieu et al., 2013). Thus, the need to expand biogeographical studies that address Gondwanan vicariance to include both older and larger angiosperm clades, especially those without formal taxonomic designations, is warranted (Beaulieu et al., 2013; Cai et al., in press).

The clusioid clade provides a test case of the potential impact of Gondwanan fragmentation on tropical angiosperms. Features of the clusioids that make them especially amenable for this purpose include their: (1) well-sampled and strongly supported phylogeny (Ruhfel et al., 2011); (2) pantropical distribution; (3) ancient fossil record (~90 Ma; Crepet and Nixon, 1998; Ruhfel et al., 2013); and (4) ancient molecular divergence time estimates for the clade (Xi et al., 2012). This clade includes five monophyletic families (Bonnetiaceae, Calophyllaceae, Clusiaceae, Hypericaceae, and Podostemaceae (Ruhfel et al., 2011; Xi et al., 2012). Moreover, phylogenetic analyses of the clusioids have further clarified the placement of Paleoclusia (Ruhfel et al., 2013), one of the oldest rosid macrofossils (Crepet et al., 2004; Schönenger and von Balthazar, 2006). These analyses provide a more informed use of Paleoclusia as an age constraint for molecular divergence-time estimation. In addition to Paleoclusia, fossil pollen from the clusioid clade is known from the Eocene (Pachydermites, ~45 Ma; Germeraad et al., 1968; Salard-Cheboldaef, 1979). These fossils directly indicate that the clade is ancient, a finding that is further supported by molecular divergence-time estimates of Malpighiales, i.e., stem group clusioids were estimated to be of Albian-Cenomanian age (~96-107 Ma) and crown group clusioids to be of Turonian-Coniacian age (~88-91 Ma; Xi et al., 2012). Although no previous biogeographical study has been conducted on the entire clusioid clade, studies of genera in Clusiaceae, Hypericaceae, and Podostemaceae indicate that at least some intercontinental disjunctions are more consistent with recent dispersal than Gondwanan vicariance (Dick et al., 2003; Kita and Kato, 2004; Meseguer et al., 2013; Nürk et al., 2015).

Each of the clusioid families is distributed primarily in the tropics and is represented on two or more former Gondwanan landmasses (see Appendix S1, see Supplemental Data with the online version of this article). These transcontinental disjunctions provide numerous opportunities to test the influence of Gondwanan fragmentation vs. more recent dispersal. These disjunctions either arose repeatedly though continental breakup of these southern landmasses, were achieved more recently via dispersal when these land areas were more separate, or though a combination of dispersal and vicariance. Several researchers have commented on the possibility that the clusioid clade and its constituent subclades are ancient. Raven and Axelrod (1974) hypothesized that various clusioid clades date to times when Africa and South America were in close proximity to one another. Others (Robson, 1977; Kato, 2006) have proposed similar biogeographical hypotheses that invoke Gondwanan vicariance for members of this clade. The goal of our study is to test the hypothesis that the modern distribution of the clusioid clade is due to ancient Gondwanan fragmentation. To accomplish this we will estimate divergence times using Bayesian methods that allow for lineage-specific rate heterogeneity. Next, we will reconstruct ancestral ranges of all major clusioid clades using methods that account for uncertainty in phylogenetic relationships and divergence-time estimates. And finally, we will identify patterns of dispersal under conditions when vicariance seems unlikely.

**MATERIALS AND METHODS**

**Taxon sampling and methods**—We used the expanded taxon sampling scheme from Ruhfel et al. (2011). This included 194 clusioid species sampled for three plastid genes (matK, ndhF, and rbcL) plus the mitochondrial gene matR. To this data set we added new data for seven species of Podostemaceae (Castelnovia multipartita Tul. & Wedd., Castelnovia princeps Tul. & Wedd., Cipoia ramosa C.P.Bove, C.T.Philbrick & Novelo, Hydrodiscus koyamae (M.Kato & Fukushima) KoI & M.Kato, Lophogynne lucunosa (Gardner) C.P.Bove & C.T.Philbrick, Macarea clavigera P.Royen, and Saxicolella amicorum J.B.Hall) and one species of Clusiaceae (Tovomitopsis paniculata (Spreng.) Planch. & Triana). These taxa include five Podostemaceae genera that were not included in Ruhfel et al. (2011): Cipoia, Hydrodiscus, Lophogynne, Macarea, and Saxicolella. Cipoia, Macarea, and Tovomitopsis paniculata have never been included in a molecular phylogenetic analysis. This sampling includes all major morphological and biogeographical representatives of the five clusioid families.

We also added new data for previously unsampled gene regions and in some instances, replaced previously sampled, but lesser quality gene sequences. Molecular methods, sequence assembly, and alignment strategy followed Ruhfel et al. (2011). Data matrices and trees are available in the Dryad Data Repository (Ruhfel et al., 2016, http://dx.doi.org/10.5061/dryad.q4h2r). We added a total of six, six, seven, and 13 sequences for matK, ndhF, rbcL, and matR, respectively. The final alignment contained 207 taxa (5 outgroups, 198 clusioids). Voucher information is provided in Appendix S2 (see Supplemental Data with the online version of this article). We specified the following as outgroups following Xi et al. (2012): Bruguierua (Rhizophoraceae), Ctenolephon (Ctenolophonaceae), Irvingia (Irvingiaceae), Pandan (Pandaceae), and Ochna (Ochnaceae).
Phylogenetic analyses and divergence time estimation—We implemented a Bayesian Markov chain Monte Carlo (MCMC) approach to simultaneously estimate phylogenetic history and divergence times using BEAST version 1.6.1 (Drummond and Rambaut, 2007). Data were partitioned by gene region following Ruhfel et al. (2011) and a GTR + Γ model with four rate categories was applied to each partition with base frequencies estimated from the data. We implemented a relaxed molecular clock (uncorrelated lognormal; Drummond et al., 2006) and a Yule tree prior. A starting tree was created using RAxML version 7.2.6 (Stamatakis, 2006) with branch lengths adjusted for time using PATHd8 version 1.0 (Britton et al., 2007); branch lengths and topology satisfied all topological and fossil prior constraints (see paragraph below). We ran five independent MCMC chains for 50 million generations each sampling every 1000 steps to obtain an estimated sample size (ESS) greater than 200 for all parameters. We assessed convergence and stationarity of estimated parameter values using Tracer version 1.5 (Rambaut and Drummond, 2009). When the independent chains converged, the samples of each run were combined after discarding the burnin (~25% of each run), using LogCombiner version 1.6.1. TreeAnnotator version 1.6.1 was then used to generate a maximum clade credibility (MCC) tree and estimate the mean node age, 95% highest posterior density (HPD) of divergence time estimates, and posterior probability for all nodes in the topology.

Topological constraints were enforced in our BEAST analyses to accommodate fossil constraints and incorporate recent phylogenetic discoveries. Relationships among our outgroups were constrained to well-supported relationships (≥ 70 bootstrap percentage [BP] or 0.95 posterior probability [PP]) in Xi et al. (2012): (1) *Irvingia malayana* Oliv. ex A.W.Benn. (Irvingiaceae) + *Panda oleosa* Pierre (Pandaceae), (2) *Bruguiera gymnorrhiza* (L.) Savigny (Rhipidophoraceae) + *Clelophoton engleriavers* Mildbr. (Clelophotonaceae), and (3) *Ochna multiflora* DC. (Ochnaceae s.l.) + the clusioid clade. We also enforced constraints within the Garchinieae clade (Clusiaceae). These constraints were based on the well-supported findings of Sweeney (2008) who used two nuclear genes to resolve relationships within the clade. To ensure that our data were not in conflict with these constraints, we performed an alternative topology test using the approximately unbiased biased test (AU; Shimodaira, 2002) as implemented in the R software package, scaleboot version 0.3-2 (Shimodaira, 2008). The constrained maximum likelihood topology could not be rejected by our data (p = 0.12).

Four fossil age-constraint priors (two clusioid, two nonclusioid) were enforced as probability distributions (Ho and Phillips, 2009). The root node was set to a uniform distribution between 89.3 Ma and 125 Ma. The former age corresponds to the minimum age of *Paleocclusia*, the oldest known fossil within Malphigiales (Crepet and Nixon, 1998). The later date corresponds to the earliest evidence of tricolpate pollen (Magallon et al., 1999; Sanderson and Doyle, 2001), a synapomorphy of the eudicot clade to which the Malphigiales belong (APG III, 2009). In addition to the root node constraint, we included three additional fossils (Table 1), two of which are clusioids, to estimate divergence times. These three fossil age constraints were modeled as lognormal distributions with separate means and standard deviations (Table 1). While there are many factors to consider when assigning the mean and standard deviation values to these lognormal prior distributions (Ho and Phillips, 2009) we have taken a conservative approach to determining these values. The minimum age of each fossil constraint was assigned based on the youngest boundary of the geological stage in which the fossil was found [geological time scale following Gradstein et al. (2004)]. For example, the *Paleocclusia* fossil is Turonian (89.3-93.5 Ma) and therefore the minimum age for this prior was set as the youngest age of that stage, 89.3 Ma. The means of the lognormal prior distributions were chosen so that the median value was approximately at the midpoint of the stage. The standard deviation was chosen so that the 97.5% quantile fell roughly at the lower boundary of the layer. Thus, the tail of each prior distribution extends well beyond the age of the fossil, allowing for the possibility that the origin of the group is much older (Table 1). Due to the uncertain phylogenetic position of *Paleocclusia* we conducted two independent analyses using alternative placements of this age constraint as suggested by Ruhfel et al. (2013). In the first analysis, *Paleocclusia* was placed at the most recent common ancestor (MRCA) of Bonnetiaceae and Clusiaceae (BC placement). For the second analysis this constraint was placed in a more nested position at the crown node of the clusioid clade (CC placement; MRCA of Clusiaceae + Symphoniaceae). We did not use Eocene fossils of *Hyperi cum* (Meseguer and Sanmartin, 2012) as an age constraint as these fossil seeds share similarities with *Elatine* (Elatinaceae), *Nymphaea* (Nymphaeaceae), and Passifloraceae s.l. (Stevens, 2001 onward).

Ancestral Area Reconstructions—Ancestral area reconstructions (AARs) were conducted in a likelihood framework using the dispersal-extinction-cladogenesis model (Ree and Smith, 2008) as implemented in a modified version of LAGRANGE (Beaulieu et al., 2013). Rather than reconstructing “splits” as in traditional LAGRANGE analyses, we reconstructed “states”, an option available in the C++ version of the program. States were chosen because summarizing split results over a set of topologies with differing phylogenetic relationships is not easily interpreted (S. Smith, University of Michigan, Ann Arbor, MI, personal communication). To account for phylogenetic and divergence time uncertainty, we conducted analyses on 1000 trees randomly selected from the posterior distribution of dated trees. We used a composite Akaike weight (interpreted as the probability of the given scenario over a set of alternative scenarios) to summarize AARs (Beaulieu et al., 2013). Confidence for the optimal reconstruction at each node is evaluated with the evidence ratio for the reconstruction with the highest Akaike weight vs. all other models (Beaulieu et al., 2002). The evidence ratio represents the relative evidence of one scenario being the most supported. Evidence ratios < 2.7 can be considered as weak support for a particular scenario, while an evidence ratio > 7.4 would indicate reasonable evidence for that model (Burnham et al., 2002; Anderson, 2007). However, explicit cutoff values for interpreting this statistic are discouraged (Burnham et al., 2002; Anderson, 2007). Prior to implementing LAGRANGE, we pruned our dated topologies to leave a single placeholder (87 total; Appendix S3, see Supplemental Data with the online version of this article) for each genus that could be meaningfully scored for our biogeographic areas. Where additional phylogenetic studies have been conducted, single terminals representing subclades within genera (e.g., *Garinia*) were also retained. Tree pruning preserved branch lengths and was conducted using the R package APE version 2.6-2 (Paradis et al., 2004). To ensure that our target trees were topologically identical, and thus our biogeographic analyses directly comparable, we resolved one uncertain node (< 50 PP) in the pruned BC topology based on phylogenetic evidence from Ruhfel et al. (2013). The clade containing *Kayea*, *Mammea*, and *Poeciloneuron*, is supported as sister to the Calophyllum + Mesua
clade in a combined morphological and molecular phylogenetic analysis (Ruhfel et al., 2013). The sister group relationship of these clades was not recovered in the BC MCC topology (Appendix S4, see Supplemental Data with the online version of this article) but was recovered in the CC MCC topology (Appendix S5, see Supplemental Data with the online version of this article).

Eight biogeographic areas were circumscribed for our analyses, mostly following Buerki et al. (2011) (Appendix S6, see Supplemental Data with the online version of this article): (1) North America, Central America, and the Caribbean; (2) South America; (3) Eurasia; (4) Africa; (5) Madagascar, plus the Comoros, Seychelles, and Mascarenes; (6) India and Sri Lanka; (7) Southeast Asia (those regions west of Wallace’s Line, but not part of continental Eurasia); and (8) Australia (those regions east of Wallace’s line including New Caledonia and the Pacific Islands). Here, we include India as separate from Southeast Asia to account for terminals solely distributed there. Throughout the text we use the term “New World” to describe the area including North America, Central America, the Caribbean, and South America, while the term “Old World” refers to the area including Eurasia, Africa, Madagascar, India and Sri Lanka, Southeast Asia, and Australia as defined above. Terminal ranges were obtained from the literature and explanations for our area coding can be found in Appendix S3.

We applied an unconstrained model and a model that incorporates information on biologically feasible ancestral ranges and dispersal probabilities scaled according to area connections through four stratified windows of time, reflecting changing land configurations during the period of interest. In the unconstrained model, all combinations of areas were allowed and no constraints were imposed on dispersal from one area to another. This approach has two disadvantages. First, the inclusion of all possible ranges allows the possibility that unrealistic ancestral ranges may be inferred (e.g., North America + Australia). Second, all dispersal events are considered equally probable. In our second model, we imposed assumptions that reflect our knowledge of the current distributions of clusioid taxa and of land configurations through time. This is relevant because the clusioid clade is ≥ 90 Ma, and thus spans a period of great change in climate and continental configurations. We limited the maximum range size in our transition matrix to three areas in most cases. Most (~90%) of our terminal taxa have range sizes of three areas or fewer and larger ranges are unlikely to be maintained over long periods of time without being further reduced via cladogenesis. However, including some terminal taxa with larger ranges (> 3) was unavoidable. This is because LAGRANGE requires any ranges present in terminal taxa to be included in the transition matrix, i.e., nine taxa in our data set have distributions of four areas or greater (e.g., Mammea, present in six areas; Appendix S3). We further reduced possible ancestral ranges by excluding those that seemed unlikely (e.g., North America + Australia). A full list of the ranges included in our second model is available in Appendix S7. We also included information on dispersal probabilities across four discrete windows of time. Dispersal probabilities and time slices used in our analyses were derived from the LAGRANGE model parameters proposed by Buerki et al. (2011). Their model considered four time slices (120-80 Ma, 80-60 Ma, 60-30 Ma, and 30-0 Ma) and three dispersal rate probabilities: (1) 1.0 for dispersal between areas that were physically connected; (2) 0.5 for areas connected through dispersal by abiotic factors such as equatorial ocean currents; and (3) 0.01 for areas that were not connected during a given period of time. Slight modifications to the Buerki et al. model were necessary to also include India, which was not considered as a separate landmass in their analyses. Thus, we added dispersal probabilities between India and the other landmasses for each of the four windows of time (Appendix S8). Because of evidence suggesting that India maintained biological connections to Africa during the late Cretaceous and Paleogene as it drifted northwards toward Eurasia (Briggs, 2003; Ali and Aitchison, 2008), we incorporated this information into our dispersal probability matrices.

RESULTS

Phylogenetic analyses and divergence time estimation—Our BEAST analyses resulted in a robust phylogeny of the clusioid clade that was very similar to Ruhfel et al. (2011). As such, we will comment mainly on the placement of the eight newly added taxa. No strongly supported conflicts were present between the full MCMC topologies from the BC and CC analyses (Appendices S4 and S5, respectively). The neotropical Podostemaceae taxa Castelnovia multipartita, C. princeps, and Lophogyne lacunosa were strongly placed (96 PP) within the strictly neotropical Podostemaceae clade in positions that agree with Tippery et al. (2011). Neotropical Cipoia racemosa was strongly supported as a member of the primarily Old World Podostemaceae clade (100 PP). Within this clade it was well-supported (98 PP) in a clade containing mostly African and Malagasy taxa. Macarena, a monotypic genus from Colombia, was placed sister to Rhynchosclis with strong support (100 PP). The African genus Saxicolletla was strongly placed (100 PP) in the primarily Old World clade of Podostemaceae. However, its position within that clade was not well-supported. Hydrodiscus koyamae was strongly placed (>95 PP) as sister to a clade containing Hanse niella and Thawatchaia. Tomovitopsis paniculata (Clusiaceae) was strongly placed (100 PP) as sister to its congener T. saldanhae Engl. Finally, the addition of matR, ndhF, and rbcL data for Diamantina (Podostemaceae) did not resolve the placement of this taxon; Ruhfel et al. (2011) sampled only matK for this taxon. Diamantina is

---

### TABLE 1. Fossils used as age constraints in the divergence time estimations. Constraints were applied to the most recent common ancestor (MRCA) listed. Mean and standard deviation (SD) for the lognormal prior on each fossil are also given.

<table>
<thead>
<tr>
<th>Fossil</th>
<th>Minimum age (Ma)</th>
<th>Mean (SD)</th>
<th>Fossil locality</th>
<th>MRCA</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ctenolophonoides costatus van Hoeken- Klinkenberg</td>
<td>65.5</td>
<td>2.5 (0.4)</td>
<td>Nigeria and Somalia</td>
<td>Ctenolophon and Bruguiera</td>
<td>(Edet and Nyong, 1994; Schrank, 1994)</td>
</tr>
<tr>
<td>Paleoclasia chevalieri Crepet &amp; Nixon</td>
<td>89.3</td>
<td>2.0 (0.5)</td>
<td>New Jersey, United States</td>
<td>BC: Bonnetiaceae and Clusiaceae s.s. CC: Clusiaceae and Symphoniaceae</td>
<td>(Crepet and Nixon, 1998; Ruhfel et al, 2013)</td>
</tr>
<tr>
<td>Pachydemites diedeni Germeraad et al.</td>
<td>40.4</td>
<td>4.5 (0.3)</td>
<td>Nigeria</td>
<td>Pentadesma and Symphonia</td>
<td>(Germeraad et al., 1968; Salard-Cheboldaeff, 1979)</td>
</tr>
</tbody>
</table>
weakenly placed (<66 PP) as sister to the strictly neotropical clade of Podostemoideae. In contrast, Ruhfel et al. (2011) weakly placed (56 ML BP) *Diamantina* as sister to the remaining Podostemoidae. The placement of *Diamantina* as sister to all other Podostemoidae agrees with other hypotheses (Philbrick et al., 2004; Rutishauser et al., 2005; Koi et al., 2006; Koi et al., 2012), but has very little support here (<5 PP). Pruned topologies used in our AARs are shown in Fig. 1 (CC) and Fig. 2 (BC).

Divergence times for major subclades using each placement of *Paleoclusia* are given in Appendix S9 (see Supplemental Data with the online version of this article). Alternative placements of *Paleoclusia* have only a minor effect on node ages, i.e., nodes are somewhat younger with the BC placement. Moreover, the 95% HPDs of most nodes overlap considerably between the BC and CC analyses except for the nodes where *Paleoclusia* was used as a constraint (Appendix S10). The effect of various branching process priors (BPP) on age estimates has received little attention (Condamine et al., 2015). The birth-death BPP generally have found either no major effect on estimated clade ages (Couvreur et al., 2014, 2015) or major differences (Condamine et al., 2015). The birth-death BPP generally estimates more nodes toward the present compared to the past (Gernhard, 2008). Thus, if estimated clusoid ages were to differ with the use of the birth-death prior it would likely not alter the main conclusions of our analyses. At the time these analyses were conducted a birth-death process prior option was not available in BEAST (Drummond et al., 2012).

**Ancestral Area Reconstructions**—Log-likelihood scores of the data under the constrained model were better when using the unconstrained model for both the BC and CC analyses (BC: unconstrained $-\ln (L) = 285.241$, constrained $-\ln (L) = 267.259$; CC: unconstrained $-\ln (L) = 287.935$, constrained $-\ln (L) = 271.699$). AARs from the unconstrained model are similar to the AARs of the more complex model. However, several nodes in the unconstrained analyses were reconstructed as having widespread ancestors, in some cases up to seven areas. We consider such widespread ancestors unlikely. Given that the constrained model is more biologically realistic (i.e., it includes information on feasible ranges as well as land connections and dispersal probabilities through time) and because the data fit the constrained analysis better as determined by likelihood scores, we focus on the results of this model (Figs. 1 and 2). Importantly, both models suggest Gondwana fragmentation as having influenced the early history of the clusioid clade. In the unconstrained model the crown node and the first two divergence events within the clusioid clade were reconstructed as having ranges identical to those using our constrained approach.

The inferred biogeographical history of the clusioid clade is very similar when using alternative placements of *Paleoclusia* (compare Figs., 1 and 2 for the CC and BC analyses, respectively) and AARs differ at only the crown nodes of Podostemaceae and Clusiaceae. Support for AARs is relatively low in the early divergence events of the clusioid clade (i.e., from the crown node of the clusioids to the crown nodes of the five families). For these nodes, Akaike weights ($w_i$) are $<0.5$ and evidence ratios are mostly $<2.7$. The crown node of the clusioids is inferred to have a range of South America + Africa ($w_i = 0.388$, evidence ratio $= 1.70$; CC $w_i = 0.370$, evidence ratio $= 1.58$). This node appears to have experienced a vicariance event with the MRCA of Bonnetiaceae + Clusiaceae inheriting a South American range ($w_i = 0.280$, evidence ratio $= 1.26$; CC $w_i = 0.369$, evidence ratio $= 1.95$), and the MRCA of Calophylleae + Hypericaceae + Podostemaceae clade inheriting an African range ($w_i = 0.401$, evidence ratio $= 2.53$; CC $w_i = 0.341$, evidence ratio $= 2.00$). The crown node of Hypericaceae + Podostemaceae is inferred as African ($w_i = 0.483$, evidence ratio $= 2.82$; CC $w_i = 0.435$, evidence ratio $= 2.48$). Crown group ancestral ranges for each family are as follows: Bonnetiaceae, South America ($w_i = 0.465$, evidence ratio $= 1.18$; CC $w_i = 0.451$, evidence ratio $= 1.11$); Clusiaceae, South America + Africa ($w_i = 0.306$, evidence ratio $= 1.54$) in the BC placement, South America ($w_i = 0.266$, evidence ratio $= 1.10$) in the CC placement; Calophylleae, South America + Africa + Australia ($w_i = 0.26$, evidence ratio $= 1.81$; CC $w_i = 0.318$, evidence ratio $= 1.62$); Hypericaceae, Africa ($w_i = 0.252$, evidence ratio $= 1.46$; CC $w_i = 0.257$, evidence ratio $= 1.55$); and for Podostemaceae, Africa ($w_i = 0.287$, evidence ratio $= 1.11$) with the BC placement, South America + Africa ($w_i = 0.258$, evidence ratio $= 1.08$) with the CC placement.

When considering the 95% HPD of estimated node ages, twenty-two area reconstructions potentially consistent with Gondwana vicariant events occur in the clusioid clade (Appendix S11, see Supplemental Data with the online version of this article; CC, Fig. 1; BC, Fig. 2). However, only two of these events in the BC analysis and four in the CC analysis occur within a time window when the areas involved would have been contiguous or in close proximity. These putative vicariant events involve the separation of South America and Africa (80-110 Ma) or South America, Antarctica, and Australia (41-55 Ma). Events involving the fragmentation of South America and Africa occur in the CC analysis (Fig. 1) at the crown nodes of the clusioids, Calophylleae, and Podostemaceae. A vicariant event at crown Podostemaceae does not occur in the BC analysis as the AARs differ (Fig. 2). The separation of South America, Antarctica, and Australia causing a vicariant event at the crown Calophylleae node is present in the CC analysis. The biogeographic scenario for this node and its descendants in the BC analysis is the same, but the age is slightly too young (95% HPD = 19.4-40.1 Ma).

Forty-nine dispersal events into new areas are inferred using either placement of *Paleoclusia*, indicating that a dispersal event occurs on ~28% of the branches. Thirty-two events occur among regions of the Old World, nine within the New World, and eight between the New and Old Worlds. The majority of these range expansion events occur well after the fragmentation of Gondwana, even when considering the 95% HPDs. In the BC analyses 37 of 49 (~76%) inferred dispersal events likely occur after the beginning of the Eocene (55.8 Ma) and of those 28 (~57%) were since the beginning of the Oligocene (33.9 Ma). In the CC analyses 35 out of 49 (71%) occurred since the beginning of the Eocene and 24 of those (49%) since the beginning of the Oligocene.

**DISCUSSION**

Our divergence time estimates for the clusioids are similar to those published elsewhere with some exceptions. On average, our dates are older when compared to Xi et al. (2012) who placed *Paleoclusia* deeper in the phylogeny at the crown node of the clusioid clade. Similarly, our dates for crown Hypericaceae are 10-20 Ma older than in recent studies (Meseguer et al., 2015; Nürk et al., 2015). Regardless of this variation, the biogeographic history of the clusioids is characterized by more recent dispersal rather than ancient vicariance. We first discuss putative examples of Gondwanan
FIGURE 1 Chronogram and biogeographic reconstructions of the clusioid clade with the *Paleoclusia* fossil placed at crown Clusiaceae (CC placement). The constrained maximum-likelihood biogeographic model included information on biologically feasible ranges as well as land connections and dispersal probabilities through four stratified windows of time. Eight areas were included in the analysis: North America (NA), South America (SA), Eurasia (EA), Africa (Af), Madagascar (Md), India (In), Southeast Asia (Se), and Australia (Au); see text for full circumscriptions of areas and Appendix S3 (see Supplemental Data with the online version of this article) for details regarding areas scored for each terminal. Filled boxes to the left of taxon
vicariance, then move to patterns of dispersal. We divide our discussion of dispersal patterns into three categories: (1) between the Old and New Worlds; (2) within the New World; and (3) within the Old World. We focus on the mean ages inferred in our divergence time estimates for dispersal events.

**Ancient Gondwanan vicariance in the clusioid clade**—We identified two or four examples of putative Gondwanan vicariance depending on the placement of *Paleoclusia*. Of these four events, three involved the separation of South America and Africa. One of these events was present only in the CC analysis. The fourth event, recovered only in the CC analysis, involved the separation of South America, Antarctica, and Australia. All other cladogenic events involving Gondwanan landmasses were far too young to be attributable to vicariance, i.e., in most cases age estimates would need to be tens of millions of years older. This suggests that as biogeographic studies are expanded to examine more inclusive and older clades (e.g., Malpighiales), it is possible that vicariance events associated with the break-up of Gondwan will be more common (Beaulieu et al., 2013; Cai et al., in press).

The placement of *Paleoclusia* within the clusioid clade is somewhat uncertain, yet is important for estimating divergence times and ancestral areas. This is underscored by the different number of Gondwanan vicariance events recovered depending on its placement as an age constraint; divergence times in the CC analysis were older and thus more nodes were of an age potentially attributable to the separation of Gondwana. At the time of deposition, the locality of *Paleoclusia* (New Jersey, USA) was in Southern Laurasia in a subtropical to tropical environment (Crepet and Nixon, 1998). However, most extant members of the clusioid clade are now found only in similar environments in more southern regions that are mostly on former Gondwanan fragments (Appendix S1). No early diverging lineages in our AARs include North America, at least in part because *Paleoclusia* was not included in our AARs. If the more nested CC placement is correct, the occurrence of *Paleoclusia* in Southern Laurasia may be indicative of a broader historical Clusiaceae distribution including South America and regions of Laurasia (e.g., North America). If the BC placement is correct, a broader historical distribution of the Bonnetiaceae + Clusiaceae clade is implied.

**Gondwanan distributions in the clusioid clade are largely the result of post-Eocene dispersal**—Dispersal rates are exceptionally high in the clusioid clade with a dispersal event along > 25% of the branches in our analyses. Moreover, > 70% of dispersal events have occurred since the beginning of the Eocene (55.8 Ma). Changing land configurations and accompanying climatic changes since the origin of the clusioids likely presented opportunities for movement between areas during some windows of time but not others (McLoughlin, 2001; Sanmartin, 2002). Moreover, we detect disjunction patterns involving similar areas in distantly related clades suggesting common dispersal corridors. Relevant to this discussion is the dispersal biology of these plants. Dispersal occurs by various means within the clusioid clade. Wind and water dispersal are common in Calophyllaceae, Hypericaceae, and Podostemaceae, and dispersal by birds and mammals occurs in Calophyllaceae, Clusiaceae, Hypericaceae, and possibly Podostemaceae (Cook and Rutishauser, 2007; Stevens, 2007a, b). Oceanic fruit dispersal is likely in some widespread taxa (e.g., *Calophyllum* and *Mammeea*; Stevens, 2007a), but in *Symphonia* vegetative propagules are more likely for marine dispersal (Dick et al., 2003). Vegetative propagules may also be water-dispersed in Podostemaceae (Philbrick et al., 2015), but it is unlikely that these freshwater plants could survive extended saltwater exposure.

**Dispersal between the Old and New Worlds**—Eight dispersal events between the Old and New Worlds were inferred. In all but one instance, these appear to be limited to dispersal between Africa and the New World (five events) or between Australia (or Australia + Southeast Asia) and South America (two events). Dispersal from South America to Africa is inferred once, but dispersal from Africa to the New World occurred at least four times, suggesting a difference in dispersal rates between these two hemispheres. The South America to Africa dispersal event occurred between the Upper Cretaceous and Eocene (~90-50 Ma) along the stem branch of Clusiaceae (CC, Fig. 1) or along the branch leading to crown Clusiaceae (BC, Fig. 2). If this event occurred early enough it might have been via direct overland dispersal or by island-hopping as South America and Africa were still in close proximity at this time (Morley, 2003). However, if dispersal occurred during the Eocene, this would likely be attributed to either long-distance dispersal over the Atlantic Ocean or possibly an overland boreotropical route. Two dispersal events in the opposite direction (Africa to South America) occur during a similar window of time along the stem branch of Calophyllaceae, and in the early history of Podostemaceae. Given the length of the branch subtending Calophyllaceae, this event could be attributed to overland dispersal or long-distance dispersal. The branch in Podostemaceae where this event occurs depends on the placement of *Paleoclusia*. In the BC analysis, either long-distance dispersal or an overland boreotropical route through the Northern Hemisphere is more likely, while in the CC analysis direct overland dispersal between South America and Africa would have been possible as they were in close proximity at this time. Dispersal between Africa and the New World is present in three other areas of the topology, i.e., the stem branch of *Hypericum*, within Vismieae, and within Garpinciae. In *Hypericum* and Vismieae, migration through the northern hemisphere is possible, but in Garinceae a more recent dispersal sometime after the latest Miocene was likely during a time when migration via Laurasia may not have been available for tropical taxa.

Dispersal from South America into Australia (or Australia + Southeast Asia) possibly occurred twice, once each in Bonnetiaceae and Clusiaceae. The branch along which dispersal occurred in Bonnetiaceae is long (> 35 Ma), thus dispersal could have occurred...
FIGURE 2 Chronogram and biogeographic reconstructions of the clusioid clade with the *Paleclusia* fossil placed at the crown node of the Bonnetiaceae + Clusiaceae clade (BC placement). The constrained maximum-likelihood biogeographic model included information on biologically feasible ranges as well as land connections and dispersal probabilities through four stratified windows of time. Eight areas were included in the analysis: North America (NA), South America (SA), Eurasia (EA), Africa (Af), Madagascar (Md), India (In), Southeast Asia (Se), and Australia (Au); see text for full circumscriptions of areas and Appendix S3 (see Supplemental Data with the online version of this article) for details regarding areas scored for each terminal.
anytime between the late Eocene and the mid Miocene. If dispersal occurred early it could have been via Antarctica or the North Atlantic Land Bridge as these routes may have been available to tropical elements during the Paleocene-Eocene Thermal Maximum (Morley, 2003; Penington and Dick, 2004). If more recent, it might either have occurred through the northern hemisphere or via long distance dispersal across the southern oceans, perhaps facilitated by the Antarctic Circumpolar Current (Sammartin et al., 2007). Dispersal from South America to New Caledonia in Symphoniae is too recent to have occurred via direct land connections between Antarctica and Australia; paleoclimates were not likely conducive for migration during this time. This leaves open the possibilities of long distance dispersal across the Pacific (Heads, 2010) or a migration through Laurasia.

**Dispersal within the New World**—Dispersal events are unidirectional from South America into Northern America for tropical clusioid taxa, a pattern broadly consistent with numerous plant groups (Burnham and Graham, 1999). The Isthmus of Panama was traditionally thought to be completely formed by \(~\text{3.5 Ma}\) but recent studies suggest connections as early as \(~\text{20 Ma}\) (Willis et al., 2014; Bacon et al., 2015). The Isthmus has long been considered an important route of biotic interchange between North and South America. Dispersal prior to the closing of the Isthmus has been suggested for many taxa, and appears more prevalent for plants than animals (Cody et al., 2010; Bacon et al., 2015). We infer nine independent events reflecting this south-to-north dispersal pattern, in Calophyllaceae, Clusiaceae, and Podostemaceae. The timeline for most of these events is not likely older than the Oligocene-Miocene transition \((~\text{23 Ma})\), but could be younger. Species-level phylogenies are needed to determine a more precise timing for these events.

**Dispersal within the Old World**—Biogeography of the clusioids in the Old World is complicated and involves at least 32 dispersal events among adjacent areas since the Upper Cretaceous. This is likely the result of a complex geological history, particularly in tropical Southeast Asia (Hall, 1998) (including areas scored as India, Eurasia, India, and Southeast Asia). Our inferred clusioid dispersal events between these regions occur between the Upper Cretaceous and Eocene \((~\text{100-34 Ma})\), but these are fewer and mostly involve movements out of Africa into Eurasia, India, and Madagascar. Most dispersal events occur after the Eocene-Oligocene boundary \((~\text{33.9 Ma})\) and primarily involve movements out of India, Southeast Asia, and Eurasia into areas immediately adjacent to these regions. During this period continental Southeast Asia and Sundaland experienced wide fluctuations in sea level and rainforest cover (Hall, 1998; Morley, 2007) that likely affected dispersal corridors. Rainforest cover, in particular, was severely reduced between 35 and 20 Ma, yet before and after this event, extended periods of suitable rainforest climate existed (Morley, 2007).

Movements from India into adjacent areas of Eurasia, Southeast Asia, and Australia occur in *Craxotyllum* (Hypericaceae), *Dalzelia* (Podostemaceae), *Kayea* and *Mammeea* (Calophyllaceae), and Gariinclaeae (Clusiaceae). Some geological estimates date the collision of India with Eurasia between 50 and 55 Ma (Scotese et al., 1988; Najman et al., 2010), but other evidence suggests that this occurred more recently, between 35 and 40 Ma (Ali and Aitchison, 2008; Bouilhol et al., 2013). Most of our inferred dispersal events out of India are consistent with overland dispersal routes becoming available during more recent periods. Changing land configurations and climate also likely influenced movement out of Southeast Asia and Eurasia into nearby areas during this period (Hall, 1998; Morley, 2007). Clades with an ancestral range in Southeast Asia or Southeast Asia + Australia that disperse into nearby areas include *Ploiarium*, a subclade of Gariinclaeae, and a subclade of Calophyllaceae. Origination in Eurasia, most likely what is now continental tropical Southeast Asia, with subsequent movement into nearby areas (e.g., India) occurs in several Podostemoidae genera. Similar “into-India” dispersal during this period has been invoked for amphibians (Van Bocxlaer et al., 2009).

The final repeated pattern of dispersal detected in our results is movement into, but not out of, Madagascar. This pattern occurs at least eleven times: Clusiaceae \((\times 4)\); Calophyllaceae \((\times 2)\); Hypericaceae \((\times 3)\); and Podostemaceae \((\times 2)\). These events appear to originate mostly in Africa, but dispersal into Madagascar from Southeast Asia, India, Eurasia, and Australia are also inferred. These events do not seem to be clustered in time and occur throughout the Upper Cretaceous and Cenozoic. These results are consistent with evidence that most Malagasy clades dispersed from Africa to Madagascar during the Cenozoic (Yoder and Nowak, 2006).

**CONCLUSIONS AND FUTURE DIRECTIONS**

Our results provide an improved phylogeny for the clusioid clade and have placed three taxa not previously included in molecular phylogenetic analyses. Furthermore, our results suggest that most clusioid clades are too young for their distributions to have been influenced strictly by Gondwanan fragmentation. Separation of Africa and South America appears to have played a role in the earliest history of the clade, though this was only detectable by examining an ancient clade with no formal taxonomic designation. Current clusioid distribution most likely reflects extensive dispersal, most of which occurred since the beginning of the Eocene \((~\text{56 Ma})\). These results further reinforce the idea that dispersal has likely played a major role in the assembly, maintenance, and distribution of tropical diversity since the Upper Cretaceous (Renner et al., 2001; Davis et al., 2002; Davis et al., 2004; Pennington and Dick, 2004; Clayton et al., 2009; Willis et al., 2014).

Future work on the biogeography of the clusioid clade should concentrate on two main topics (i.e., fossils and phylogenies). The placement of *Paleoclusia* has an effect on both divergence time estimates for the clusioid clade, particularly dates for stem and crown Clusiaceae, as well as some AARs. Given the material available for
Paleclusia it is unlikely that a more precise placement for this fossil will be obtained. However, if vegetative material of Paleclusia is discovered, this may greatly improve our knowledge of its phylogenetic placement (i.e., vegetative characters appear to be important for placing clusioid taxa scored only for morphology (Ruhfel et al., 2013)). Furthermore, reliable fossils are presently lacking for several clusioid subclades, particularly Calophyllaceae and Podostemaceae. This is problematic as nodal points farther from fossil calibration points are more difficult to estimate (Linder et al., 2005). Particularly useful would be fossils of Podostemaceae. This clade seems to have an exceptionally elevated rate of molecular evolution (Davis et al., 2007; Ruhfel et al., 2011) that may confound age estimates (Smith and Donoghue, 2008). Paleobotanists should also be on the lookout for fossil tropical clusioids in the northern hemisphere, which may help determine if any clades dispersed from the New World to the Old World via land connections. In addition, species-level phyllogenies are needed for many clusioid clades (genera, tribes, and subfamilies). Increased phylogenetic resolution within these clades would be particularly helpful in elucidating the biogeographic history of the group. These include especially: Calophyllum, Clusiaceae, Garcineriae, Mammea, Podostemoideae, and Vismieae. Well-sampled and resolved phyllogenies for these clades will avoid the complications of scoring widespread terminal taxa, and allow a more informed view of the timing and location of biogeographic events in the clusioid clade.

ACKNOWLEDGEMENTS
The authors thank the following individuals for providing tissues, help with data analysis, comments on the manuscript, and discussions on the clusioid fossil record: M. C. E Amaral, A. Baresch, J. M. Beaulieu, R. T. Pennington, C. A. Jaramillo, S. Mathews, S. A. Smith, P. S. Soltis, D. E. Soltis, and Z. Xi. We would also like to thank the reviewers of this manuscript for their time, and for providing helpful comments and suggestions that improved the quality of our work. Funding was provided by the National Science Foundation (NSF) DDE grant OISE-0936076 to B.R.R. and C.C.D. and the NSF Assembling the Tree of Life grant DEB-0622764 to C.C.D. B.R.R. was also supported by the American Society of Plant Taxonomists, the Arnold Arboretum, David Rockefeller Center for Latin American Studies, the Department of Organismic and Evolutionary Biology at Harvard University, and the Garden Club of America. C.P.B. was supported by the Brazilian National Council for Scientific and Technological Development (CNPq) grants PROTAX 562251/2010-3, REFLORA 563534/2010-9, and Productivity Grant 307870/2014-6.

LITERATURE CITED


Ruhfel, B. R., P. F. Stevens, and C. C. Davis. 2013. Combined morphological and molecular phylogeny of the clusioid clade (Malpighiales) and


