Global history of the ancient monocot family Araceae inferred with models accounting for past continental positions and previous ranges based on fossils

Lars Nauheimer1,3, Dirk Metzler2 and Susanne S. Renner1

1Systematic Botany and Biology, Department of Biology, University of Munich (LMU), Menzinger-Str. 67, 80638 Munich, Germany; 2Evolutionary Biology, Department of Biology, University of Munich (LMU), Grosshaderner-Str. 2, 82152 Planegg-Martinsried, Germany; 3Present address: Molecular Evolution and Systematics of Plants, Institute of Biology, University Leipzig, Johannisallee 21-23, 04103 Leipzig, Germany

Summary

- The family Araceae (3790 species, 117 genera) has one of the oldest fossil records among angiosperms. Ecologically, members of this family range from free-floating aquatics (Pistia and Lemna) to tropical epiphytes. Here, we infer some of the macroevolutionary processes that have led to the worldwide range of this family and test how the inclusion of fossil (formerly occupied) geographical ranges affects biogeographical reconstructions.
- Using a complete genus-level phylogeny from plastid sequences and outgroups representing the 13 other Alismatales families, we estimate divergence times by applying different clock models and reconstruct range shifts under different models of past continental connectivity, with or without the incorporation of fossil locations.
- Araceae began to diversify in the Early Cretaceous (when the breakup of Pangea was in its final stages), and all eight subfamilies existed before the K/T boundary. Early lineages persist in Laurasia, with several relatively recent entries into Africa, South America, South-East Asia and Australia.
- Water-associated habitats appear to be ancestral in the family, and DNA substitution rates are especially high in free-floating Araceae. Past distributions inferred when fossils are included differ in nontrivial ways from those without fossils. Our complete genus-level time-scale for the Araceae may prove to be useful for ecological and physiological studies.

Introduction

Molecular clocks calibrated with critically assessed fossils over the past 15 yr have provided major insights into plant evolution (e.g. Schneider et al., 2004; Smith & Donoghue, 2008; Smith et al., 2010; Clarke et al., 2011; Crisp & Cook, 2011; Nagalingum et al., 2011; for a review of earlier plant clock studies, see Renner, 2005). Because rate heterogeneity is a common occurrence in large datasets (Smith & Donoghue, 2008; Smith et al., 2010), the introduction of relaxed molecular clock approaches presented a major advance for studies of macroevolution and biogeography (Sanderson, 1997, 2002; Thorne et al., 1998; Drummond et al., 2006; Himmelmann & Metzler, 2009). Nevertheless, challenges remain. The most important among them concern the handling of age constraints on calibration nodes (Yang & Rannala, 2006; Ho & Phillips, 2009; Clarke et al., 2011; Wilkinson et al., 2011) and how best to incorporate fossils (Manos et al., 2007; Doyle & Endress, 2010). The latter topic is multi-faceted because fossils provide two kinds of information: the minimum age of a group and its presence in some geographical area at a specific time. It is this latter information that has proven difficult to incorporate into ancestral area reconstructions (AARs) (but see Clayton et al., 2009).

Inferring past geographical ranges of plant clades is important if we are to understand the speed with which floras adapted to climate change in situ as opposed to the arrival of climatically pre-adapted lineages from other regions and the extinction of competitively inferior local lineages. Only large clades that occur in both temperate and tropical climates and that have a fossil record from different regions and geological eras are suitable for testing and improving the approaches available for the reconstruction of changing past ranges. Benchmark studies have focused on the tree family Simaroubaceae, the first plant group for which fossils were incorporated into a maximum likelihood (ML) biogeographical model (Clayton et al., 2009), and the gymnosperm family Cupressaceae (Mao et al., 2012). These are ancient (Jurassic to Cretaceous) families, and migration pathways available to them would have changed greatly after the breakup of Pangea and Gondwana, and more recent events, such as the closure of the Isthmus of Panama or the deterioration of the Beringian land bridge. Changes in connectivity between areas can be explicitly modelled using ML (Ree & Smith, 2008; Ree &
Sanmartín, 2009), and several studies have explored this option (Table 1).

Here, we use a complete genus-level phylogeny of the ancient monocot family Araceae to develop a synthetic approach that consists of, first, the evaluation of the sensitivity of time inference to different implementations of prior constraints on node ages and, second, the modelling of the changing continental connections through time and the incorporation of the best-supported fossils directly into AARs. We compare the results obtained with and without information on past (no longer occupied) ranges to arrive at a plausible scenario for the earliest history of Araceae. The family Araceae comprises 3790 species in 117 genera (Boyce & Croat, 2011). They are among the horticulturally most important families of monocotyledons and have received much attention from pollination ecologists (Chouteau et al., 2008). Although the family Araceae is most diverse in the tropics, it includes a few genera in the subtropics and temperate regions of North America, Eurasia, the Mediterranean region and Australia that occur in habitats ranging from bogs to deserts (Fig. 1). Molecular and morphological phylogenetic analyses have included representatives of most of the genera (Cabrera et al., 2008; Cusimano et al., 2011). A formal biogeographical analysis, however, has never been attempted, and a review of the family’s biogeography was pre-molecular and, in hindsight, interpreted many unnatural groups (Mayo, 1993).

There is no doubt that members of the family Araceae were geographically widespread in the Cretaceous (Stockey et al., 1997, 2007; Friis et al., 2004, 2006, 2010; Bogner et al., 2005, 2007; Wilde et al., 2005; Herrera et al., 2008). The quality of several of their fossils not only allows the robust calibration of DNA substitution rates, but, importantly, also provides information about past ranges of certain clades. We used this fossil record in Bayesian analyses with relaxed molecular clock models and biogeographical likelihood analyses. Our primary questions are as follows: when and where did the family Araceae undergo its early diversification; what is the geographical and temporal context of the family’s two aquatic lineages (Pistia, Lemnoideae); and what has been the impact on Araceae of the climate change from the Palaeocene/Eocene high temperatures to the Oligocene and Miocene cooling?

Materials and Methods

Sampling and alignment of DNA

We augmented the datasets of Cabrera et al. (2008) and Cusimano et al. (2011) by sequencing from disjunctly distributed or newly described genera, namely Alcatisa, Amorphophallus, Aponalis, Arisema, Colocasia, Hestia, Homalomena, Nephthytis, Ooia, Pichinia and Rhaphidophora. The final alignment comprised 132 Araceae plus Acorus (Acoraceae), sister to all other monocots, and Tofieldia (Tofieldiaceae), to represent the sister clade of the Araceae (Supporting Information Table S1 provides species, author names, herbarium voucher information and GenBank accession numbers).

Table 1 Biogeographical studies that have used maximum likelihood models with constrained dispersal probabilities and the ways in which they have accommodated uncertainty

<table>
<thead>
<tr>
<th>Author</th>
<th>Root age (Ma)</th>
<th>Number of time slices</th>
<th>Number of areas</th>
<th>Comparison with equal dispersal probability models</th>
<th>Comparison of dispersal null model</th>
<th>Comparison with DNA</th>
<th>Fossil ranges included</th>
<th>AAR, ancestral area reconstruction; DIVA, dispersal vicariance analysis (Ronquist, 1997).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clayton et al. (2009)</td>
<td>70</td>
<td>4</td>
<td>7</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Smith (2009)</td>
<td>60</td>
<td>4</td>
<td>4</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Bouchenak-Khelladi et al. (2010)</td>
<td>49</td>
<td>2</td>
<td>4</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Saed et al. (2010)</td>
<td>60</td>
<td>4</td>
<td>4</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Salvo et al. (2011)</td>
<td>50</td>
<td>4</td>
<td>4</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Courte à-Mazuecos &amp; Vargas (2011)</td>
<td>30</td>
<td>5</td>
<td>3</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Valenzuela et al. (2012)</td>
<td>75</td>
<td>4</td>
<td>4</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Ibarra-Murillo et al. (2012)</td>
<td>25</td>
<td>4</td>
<td>4</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Mao et al. (2012)</td>
<td>150</td>
<td>3</td>
<td>3</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>This Study</td>
<td>150</td>
<td>3</td>
<td>3</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

AAR, ancestral area reconstruction; DIVA, dispersal vicariance analysis (Ronquist, 1997).
Like previous family-wide analyses, we relied on chloroplast loci, namely the \textit{trnL} intron and spacer, the \textit{matK} gene and partial \textit{trnK} intron, and the \textit{rbcL} gene. We used standard primers, except for \textit{matK} for which we used the Araceae-adapted primers of Cusimano \textit{et al.} (2010). Sequences were edited and aligned manually using Sequencher 4.7, and regions of uncertain alignment were excluded, leading to the removal of 881 nucleotides, mostly as a result of several autapomorphic indels and microsatellite regions in the \textit{trnL} intron. The final matrix included 4343 aligned positions, 2987 of which belonged to coding regions; it was deposited in TreeBASE (submission 12268).

**Phylogenetics and relaxed molecular clock dating**

Phylogenetic analyses of the concatenated sequence data were performed under ML optimization, using RAxML 7.2.6 (Stamatakis \textit{et al.}, 2008) with separate GTR + \(\Gamma\) substitution models for coding and noncoding regions, and the fast bootstrap option, using 1000 replicates. Throughout this article, > 85\% bootstrap support (BS) is considered as medium support and 95–100\% as strong support. Chronograms (phylogenies with branch lengths scaled to geological time) were estimated using two Bayesian methods: TreeTime (Himmelmann & Metzler, 2009) and BEAST (Drummond \textit{et al.}, 2006, 2012; Drummond & Rambaut, 2007).

In BEAST (versions 1.6.2 and 1.7.0), rate variation across branches is modelled as uncorrelated and log-normally distributed. We used a pure birth (Yule) tree prior and the substitution models recommended by jModeltest: TPM1uf + \(\Gamma\) for the non-coding region and GTR + \(\Gamma\) for the coding region taking into account codon positions. An additional analysis applied the simpler JC + \(\Gamma\) substitution model to assess possible over-parameterization. We ran 12 \(\times\) 20 million generations of Monte Carlo Markov chains (MCMC) for the more complex substitution model and 6 \(\times\) 15 million generations for the simple model. Convergence was analysed in Tracer (1.5; Rambaut & Drummond, 2007), and runs were continued until effective sample sizes (ESSs) were > 200. Separate runs reached similar posterior probabilities (PPs) and, after the exclusion of appropriate burn-in fractions, they were concatenated in LogCombiner (1.6.2, part of the BEAST package) and resampled at a lower density to obtain a final sample of \(c.\) 10 000 trees. Maximum clade credibility trees with mean node heights were constructed using TreeAnnotator (1.6.2, part of the BEAST package). We also ran analyses with empty alignments (‘prior-only’ option) and compared the resulting posterior divergence times to assess the influence of prior settings.

Nodes with PP \(\geq 0.95\) were considered to be moderately well supported and nodes with PP = 1 as strongly supported.

TreeTime (version 1.0.1; http://evol.bio.lmu.de/_statgen/software/treetime/) differs from BEAST in not using a tree prior and, instead, assuming a uniform prior for all combinations of branch lengths, conditioned on the exponentially distributed age of the root and the additional priors for node ages, as specified by the user for time calibration. The user can choose among four
models of rate change along the tree, a compound Poisson distribution, a Dirichlet distribution, an uncorrelated exponential distribution (UCED) or an uncorrelated log-normal distribution, but there is no statistical way to select the best-fitting model for one’s data. We therefore ran all four models, always using the GTR + Γ substitution model and 2 × 5 MCMC of one million generations, with a burn-in period of 10 000 generations and parameters sampled every 1000th tree. Results from two runs per rate change model were combined manually to create an input file for TreeAnnotator in which the maximum clade credibility tree with mean node heights was then obtained.

**Fossil calibration**

Table S2 lists all fossils, with the morphological arguments for their attribution to Araceae or particular nodes within that family. Of the 21 fossils, we used seven for calibration purposes and nine in the geographical analyses of past ranges (see the next section). For absolute ages, we relied on the time-scales of Walker & Geissman (2009) and Ogg (2010).

(1) The subfamily Orontioideae is represented by macrofossils from the Late Cretaceous to the Eocene of North America and Europe (Bogner et al., 2005, 2007). A fossilized infructescence (*Albertarum puerii*) from Alberta, Canada, provides a minimum constraint of 72 million years ago (Ma) for the stem node of Orontioideae (node 3 in all figures and table). Slightly younger fossils of Orontioideae are *Lysichiton austricus*, *Orontium mackii*, *O. wolfei* and *Symphlocarpus hoffmaniae* (Table S2). (2) The subfamily Lasioideae is first known from the pollen taxon *Lasioideaecladites* from the Late Cretaceous of Siberia (Hofmann & Zetter, 2010; Table S2), and we used the age of *L. bessii* (70 Ma) to constrain the stem lineage (node 28). (3) The oldest free-floating member of Araceae is represented by *Limmohiophyllum scutatum* from the Late Cretaceous of North America (a taxon occurring into the Oligocene in East Asia). We used the oldest occurrence of *L. scutatum* (65.5 Ma; Kvacck, 1995; Table S2) to constrain the stem lineage of Lemnoideae (node 6). (4) The subfamily Aroidae is first known from the Palaeocene Colombian leaf taxon *Monrichardia aquatrica* (Herrera et al., 2008; Table S2), and we used the age of this fossil (55.8 Ma) to constrain the stem node of *Monrichardia* (node 44). (5) The same formation also contains fossil leaves resembling those of the living genus *Anthurium* (Herrera et al., 2008), and we used the age of *Petrocardium cerrejonense* (55.8 Ma) to constrain the stem node of *Anthurium* (node 13). (6) The *Typhonodorum* clade (see the Results section, Fig. 2) is first represented by the leaf morphogenus *Nipophyllites*, with *N. zaisanicus* from the Palaeocene (55.8 Ma) of Kazakhstan (Wilde et al., 2005), *N. limnesiis* from the middle Eocene of America (Dilcher & Daghlìan, 1977; Wilde et al., 2005) and *N. bohemicus* from the lower Eocene of the Czech Republic (Wilde et al., 2005). We used an age of 55.8 Ma to constrain the stem node of the *Typhonodorum* clade (node 62). (7) The subfamily Monsteroideae is first known from the leaf fossil *Araciphyllites tertiaris* from the middle Eocene (47 Ma) of Germany (Wilde et al., 2005). Monsteroide leaves, however, evolved at least twice in the Araceae, once in the New World *Heteropsis* clade and once in the Old World *Rhaphidophora* clade (Cusimano et al., 2011; see the Results section, Fig. 2). The German leaves closely resemble those of living Asian species in the genera *Epipremnum*, *Rhaphidophora* and *Scindapsus*, arguing for an assignment to the stem lineage of these Old World genera. We therefore used 47 Ma to constrain the stem lineage of the *Rhaphidophora* clade (node 20). (8) Our constraint for the root (the monocot crown node) is based on the oldest monocot pollen (*Liliacidites*, 125 Ma; Doyle et al., 2008; Table S2) as the minimum boundary combined with the youngest (139 Ma) and oldest (138 Ma) median ages inferred for the monocot crown group in molecular clock studies (Bell et al., 2010; Smith et al., 2010). The 125- or 112-Ma pollen taxon *Mayoa portugallica* (Friis et al., 2004, 2010; Table S2) may also represent Araceae, but the exine structure of *Mayoa* is ‘rarely columellae-like’ (Friis et al., 2004: 16566), raising the possibility that the grain might be from a gymnosperm (Hofmann & Zetter, 2010). Inflorescences including *in situ* pollen from the Albian/Aptian of Portugal (112 Ma; Table S2) clearly represent Araceae (Friis et al., 2010), but cannot yet be confidently assigned to particular nodes.

Using the above fossils, we devised two prior constraint schemes. The first consisted of uniform priors with hard minimum bounds for the seven fossils and a normal distribution with a soft minimum and maximum bound for the root (mean, 132 Ma; standard deviation, 4.25 Ma). The maximum bound on the uniform priors was set to 500 Ma, an age sufficiently high to effectively give all possible ages up to the soft maximum constraint at the root an equal probability. The second scheme consisted of gamma-distributed priors for all eight constraints, resulting in a higher probability (compared with the uniform priors) of ages falling close to the minimum constraint. For the root, our gamma distribution had an offset of 123.9, a shape parameter of 2 and a scale parameter of 3.07, which permitted 5% of the inferred ages to be < 125 Ma, 5% to be > 138.5 Ma, with 90% falling in between these two dates (based on the same rationale as above, constraint 8). For the remaining seven fossils, we chose gamma distributions with shape parameters 2 and offsets and scales set so that 5% of the ages could be younger than the respective fossil and 5% could be older than the earliest monocot pollen, *Liliacidites*.

Four earlier studies have dated groups of Araceae using fossil calibrations. Nie et al. (2006) used *Albertarum puerii* (72 Ma; our constraint 1) as a minimum constraint for the six species of Orontioideae living today, whereas we place this fossil at the relevant stem node. Renner & Zhang (2004) and Mansion et al. (2008) assigned *Nitophyllites zaisanicus* (our constraint 6) to the split between *Typhonodorum* and *Peltandra*, whereas we assign this fossil to the stem of this clade. They also assigned a 45-Ma leaf fossil, *Caladiosoma messelense* (Wilde et al., 2005), to an apparent *Alocasia*/*Colocasia* node, since shown to have been an artefact (Nauheimer et al., 2012). Lastly, Renner et al. (2004a) used a controversial fossil from the Miocene Latah Formation near Spokane (16–18 Ma) to constrain the split between *Arisaema triphyllum* from North America and *A. amurense* from Korea, China and Russia (Renner et al., 2004b), resulting in relatively old ages inferred for that genus.
Fig. 2 Maximum likelihood tree for the Araceae based on 4343 aligned nucleotides of plastid DNA. Support values based on 1000 bootstrap replicates are indicated by the circles at the nodes: white ≥ 70, grey ≥ 85 and black ≥ 95%. Clade names follow Cusimano et al. (2011).
Biogeographical analyses

For AARs, we relied on the dispersal–extinction–cladogenesis (DEC) model, implemented in Lagrange (Ree et al., 2005; Ree & Smith, 2008). It uses the information contained in genetic branch lengths and allows the incorporation of changing dispersal probabilities across areas and time. We devised two time-slice models, one with bins of 0–30, 30–90 and 90–150 Ma, and the other with bins of 0–30, 30–60, 60–90 and 90–150 Ma (Table S3). The oldest bin captures the plate tectonic situation before the breakup of West Gondwana, and the youngest the period during which the Central American land bridge and South-East Asia formed. The middle bins tried to capture connectivity via the North Atlantic land bridge and Antarctica, and differed in the way in which India connects to Eurasia (Table S3). Our nine operational geographical units were Eurasia (A), Africa (B), Madagascar (C), South-East Asia and India (D), Australia (E), North and Central America (F), South America (G) and Antarctica (H). To accommodate the mostly globally distributed water-associated and free-floating taxa, we created a ninth category, ‘water-associated’ (I), assigned to the marine Alismatales families Cymodoceaceae, Posidoniaceae, Ruppiaceae and Zosteraceae, the freshwater aquatics Alismataceae, Aponogetonaceae and Potamogetonaceae, plants of marshy coastal habitats (Juncaginaceae), and free-floating Araceae (Lemnoideae, Pistia). Remusatia was coded as present only in area D because most of its species occur in the Himalayan foothills and the Western Ghats of India; the widespread R. vivipara, occurring in Africa, Asia and Australia, is especially adapted to bird dispersal. Similarly, Pothos, occurring on Madagascar with one widespread species (P. scandens), was only coded for areas A, D and E, where it has several species, and Sauromatum was only coded for areas A and D, although it also has one species (S. venosum) ranging from Africa to tropical China. Spathiphyllum was coded for South America and Central North America, where 44 species occur, but not for South-East Asia, where S. commutatum and S. solomonense occur on the Philippines and New Guinea; its monospecific sister group, Holochlamys, is endemic on New Guinea. In the absence of molecular phylogenetic evidence on the relationships of the two Asian Spathiphyllum species, we left it unwise to code this genus as present in Asia because its two Asian species may turn out instead to belong to Holochlamys.

To create a tree that included all families of Alismatales (the order to which the Araceae family belongs), we manually added one representative per family to a newick tree file obtained from BEAST. The topology and divergence times were constrained to match the results of the large monocot chronogram of Janssen & Bremer (2004). The enlarged tree had 145 tips (132 Araceae and 13 outgroups) and became the input tree for Lagrange. To create Python script input files (with the tree of choice and the area connectivity probability matrices; Table S3), we used the Lagrange online configurator tool. Ancestral areas were limited to maximally two, and a relative probability of > 66.6% was considered to be strong support for an ancestral range scenario.

To integrate fossil ranges into the reconstructions, we added them manually in the newick chronogram (as performed for the Alismatales outgroups, above). Each fossil was inserted along the stem lineage of the group to which it had been assigned, with its age determining where it was placed. In addition, each fossil was given either a short branch length (1 Ma), simulating an extinct range, or a long branch length (the fossil’s age), simulating a range occupied for a long time (until today). When fossils are given short branch lengths, the DEC model, which considers branch lengths as proportional to time, will treat any range shifts indicated by their geographical occurrences as evidence for rapid geographical change (and, conversely, for long branch lengths). One of the seven fossils used for clock calibration, Limnobium scutatum, was not used in the AAR because it was too young to be assigned to the relevant stem lineage, whereas three others that had not been used as constraints were added because they contributed geographical information (Lysichiton australis, Orontium mackeii and Keratocarpa allenbyense; Table S2).

Results

Ages and substitution rates inferred using different clock models and calibration priors

Results from BEAST runs with an empty alignment revealed no contradictions among the prior constraints and showed that the PPs with the complete alignment differed from those without, indicating that the signal in the data overwrote the priors. The topologies of the ML phylogeny (Fig. 2) and the Bayesian relaxed clock chronograms (Figs 3, S2) differ only in the placement of a few statistically unsupported nodes. Although the testing of generic boundaries is not the topic of this study, four of the seven genera for which we included more than one species turned out to be polyphyletic. The Bornean species of Nephthytis (N. bintulensis) groups with other South-East Asian genera, and the African type species of the genus (N. afzelii) groups with other African genera (nodes 118 and 119). The Asian representative of the large genus Homalomena groups with another Asian genus (Furtadoa), and both are sister to the American genus Philodendron, whereas the sampled American species of Homalomena, which has few species on that continent, falls elsewhere (nodes 113 and 115). The Asian genera Alocasia and Colocasia are also polyphyletic (nodes 70, 72, 93).

Combining independent Bayesian MCMCs yielded ESSs > 200, indicating that the posterior estimates were not unduly influenced by autocorrelation. The parameter-intensive complex substitution model needed > 200 million MCMC generations to reach the > 200 ESS threshold, whereas the simple model needed 50 million. Ages for statistically supported nodes in the resulting chronograms under complex or simple substitution models differed by only 4.1% on average (Table S4).

The ages obtained from the TreeTime analyses using the compound Poisson process, the Dirichlet model, UCED or an uncorrelated log-normal distribution are shown in Table S4. The root node (monocot crown group) was between 204 and 146 Ma, but these drastic differences did not consistently carry through to the tips. TreeTime differs from BEAST in using no tree prior and
instead assuming a uniform prior for all combinations of branch lengths, conditioned on the exponentially distributed age of the root and the additional priors specified by the user for time calibration. This may be the reason for the generally older root ages in TreeTime, whereas the ages higher up in the tree were not that different from those obtained with BEAST (Table S4). We know of no statistic for choosing among the four models in TreeTime, but note that the UCED model yielded ages closest to those from BEAST. Below, we focus on the BEAST chronogram from the constraint scheme using the uniform priors for the seven fossils and the simple substitution model, rather than the chronogram obtained with the gamma priors and complex substitution models (all results are shown in Table S4).

With BEAST, the Araceae stem lineage (Fig. 3, node 1) is dated to 135 Ma and the Araceae crown group (node 2) to 121.7 Ma (95% confidence intervals (CIs) on all ages are shown in Table S4). Six of the eight subfamilies (marked with capital letters in Fig. 3) existed by the Late Cretaceous (<100–80 Ma), the Lemnoideae even a bit earlier (103.6 Ma); the Zamioculca- doidea evolved just around the K/T boundary (67 Ma). The crown groups of most subfamilies are much younger than their stems, most extremely so in the Lasiioideae (stem, 90 Ma; crown, 26 Ma). By contrast, the most species-rich subfamily, the Aroidae (1573 species in 75 genera), diversified into 10 major lineages between 87 Ma (node 38) and 62.3 Ma (node 61).

The posterior age distributions (Fig. S3) obtained for three of the constrained nodes are substantially shaped by their priors Nitrophyllites zaisanicus (estimated mean 57.5 Ma/constrained 55.8 Ma), Araciphyllites tertarius (51.38 Ma/47 Ma) and Petrocardium cerrejonense (64.51 Ma/55.8 Ma), whereas the fossils Lasioideaecidites hessi, Limnobodium scutatum and Montrichardia aquatic hardly affected the posterior distribution of the nodes to which they were assigned.

Plastid DNA substitution rates across the Araceae vary from values of 1.23 × 10^{-4} to 2.19 × 10^{-3} substitutions per site per million years (Table S5, Fig. S1). The average rate is 4.12 × 10^{-4} and the median 3.47 × 10^{-4}. The highest rates (> 10^{-3}) occur on branches leading to the free-floating Lemnoideae and Pistia, the stem of the aquatic Cryptocoryne and Lagenandra, and the branches between nodes 2 and 6.

Ancestral areas inferred with and without information from fossil ranges

We ran AARs with each of the two time-slice models and compared the resulting global likelihoods; the three-time-slice model (Table S3) resulted in a higher likelihood than the four-time-slice model and was therefore preferred. The ancestral areas inferred with ancient ranges (from fossils) included and assigned short branches are shown in Fig. 3; those inferred from only the geographical ranges of living genera are shown in Fig. S2. Nodes that changed when fossils were given long branches are shown in the inset in Fig. 3. The probabilities for ancestral areas obtained with the three fossil insertion models are given in Table S5. The inferred ancestral areas for most of the 131 ingroup nodes had probabilities of > 66.6%.

The inclusion of fossil ranges in the AARs had the greatest effects in early-diverging lineages and in nodes close to fossils. Without fossil ranges, the origin of the Araceae (node 1), its first divergence (node 2) and the Gymnostachydoidea lineage (node 3) are reconstructed as water associated. With fossil ranges included, these lineages are reconstructed as originating in West Laurasia (North America). Whether fossils were simulated as extinct lineages by giving them short branches or as still living lineages by giving them long branches affected seven aroid nodes (Fig. 3, inset; Table S5). For example, node 44 is inferred as having Asian descendants in the short-branched model, but South American ones in the long-branched model. The short-branch model may be more realistic as these fossil lineages obviously no longer occupy their former ranges.

A striking geographical disjunction in the Aroidae involves Hapaline (node 52; seven species), the only Asian member of an otherwise South American clade (its sister group is the monotypic South American genus Jasarrus). This divergence apparently dates to the Eocene/Oligocene boundary at 34 Ma (19–49 Ma). A similarly unusual disjunction is that between Peltandra, with two species in Florida and the eastern USA, and its closest relatives (Typhonodorum and three other genera) in East Africa, Madagascar and adjacent islands (Mayo, 1993; our Fig. 3, node 64). AARs with or without fossil ranges (nodes 62 and 9) infer a Eurasian origin of the stem lineage of this clade (Madagascar then probably reached by over-water dispersal from Africa or Asia). Another case of apparent trans-oceanic dispersal involves node 27, the Monsteroideae genus Spathiphyllum, with 44 species in Central and South America, two on the Philippines and New Guinea (not sequenced), and Holochlamys, with a single species in New Guinea. This split is dated to 21 Ma (7–36 Ma) and, according to our AARs, is a result of trans-Laurasian range expansion (Fig. S2). This inference should probably be viewed sceptically because of the incomplete sampling of the relevant species.

Fig. 3 Results from divergence dating with BEAST and ancestral area reconstruction (AAR) with fossils included and assigned short branch lengths. Grey bars at nodes show 95% confidence intervals. Exact ages are shown in Supporting Information Table S5. Black dots with a white centre indicate posterior probability (PP) of ≥ 0.95, entire black dots PP = 1. Araceae taxon names are in black, outgroup names in grey and added fossil taxa in red. Branches are coloured by area: Eurasia in red (A), Africa in yellow (B), Madagascar in ochre (C), South-East Asia and India in orange (D), Australia in purple (E), North America in dark blue (F), South America in green (G), Antarctica in grey (H) and world-wide-distributed water-associated taxa in turquoise (I). Letters to the left of taxon names show the area coded for that taxon. Only the AAR with the highest probability is shown; those with bold numbers have a probability ≥ 66.6%. Underlined node numbers mark nodes sensitive to whether fossils were given long or short branches. Letters in boxes next to taxon names indicate subfamilies: Gymnostachydoideae (G), Orontioideae (O), Lemnoideae (Le), Pothoideae (P), Monsteroideae (M), Lasiioideae (La), Zamioculca doideae (Z), Aroidae (A).
Discussion

This study provides the first complete genus-level chronogram for the Araceae and a biogeographical analysis that not only incorporates formerly occupied ranges but also treats the incorporated fossils in different ways and tests the fit of different time-slice models. Table 1 summarizes the methodological progress in AARs over the past few years and shows how our study differs from previous approaches. The point of including fossils on short or long ‘genetic’ branches was to simulate fossil lineages that either went extinct shortly after the age of the respective fossil or that persisted for a long time. The primary questions we wished to answer in this study were when and where the Araceae family underwent its early diversification, the time of evolution of the family’s two aquatic lineages (Pistia, Lemnoideae) and the impact of the climate change over the past 60 million years. The answers to these questions are provided in the three Discussion sections below. We also briefly discuss the unexpectedly high substitution rates in aquatic Araceae.

All AAR hinges on the correct inference of time. We therefore inferred divergence times using different approaches. The TreeTime program (Himmelmann & Metzler, 2009) yielded surprisingly older root ages than obtained with BEAST, whereas ages near the tips inferred from the two approaches were in better agreement, especially under the UCED model (Table S4). The two programs differ in using a tree prior (BEAST) or not (TreeTime), and this may affect root ages. BEAST runs carried out with simple or complex substitution models, and using gamma distributions or uniform distributions on the fossil prior constraints, yielded similar node ages (Table S4), the main difference being that results were obtained much more quickly with the simpler models.

Different approaches have been used to assess the effect of chronogram uncertainty on AARs (Smith, 2009; Salvo et al., 2010; Buerki et al., 2011; Fernández-Mazuecos & Vargas, 2011; our Table 1); basically, ancestral areas were inferred on many chronograms, rather than just one. Because of the size of our tree and the complexity of manually adding the fossils, we refrained from using one of the TreeTime chronograms as input for an AAR, but instead relied on the BEAST chronogram obtained with uniform prior constraints, which we feel is the most conservative dating approach.

A further methodological issue is how one should choose among different time-slice models. We compared the global likelihood of models that assumed different area connectivity in three or four time bins (0–30, 30–90, 90–150 Ma or 0–30, 30–60, 60–90, 90–150 Ma; Table S3), and preferred the three-time-slice model because it had a higher likelihood. The only other study to compare the fit of different time slices (Mao et al., 2012; our Table 1) found that a more complex time-slice model fitted their data better than a simpler one, but it is not completely clear what metric to use to assess model fitting (R. Ree, Field Museum, Chicago, IL, USA, pers. comm., March 2012).

Araceae in time and space – early occupation of aquatic habitats

At the onset of the Early Cretaceous, when the Araceae diverged from the remaining Alismatales (138 Ma; CI, 130–146 Ma), the breakup of Pangea (160–138 Ma) into the supercontinents Laurasia and Gondwana was essentially complete. North America and South America, however, were still close (Smith et al., 2004). The inferred origin of the Araceae as ‘water associated’ (without the benefit of fossil range information) or Laurasian (with fossils included and assigned short or long branches) matches several lines of evidence.

An origin in wet habitats fits the ecology of, and fossils associated with, early-diverging clades in the family. The deepest divergence in Araceae is between a clade comprising the Australian subfamily Gymnostachyoidae (one species) and the North American and Asian subfamily Orontioideae (seven species) plus their sister clade comprising the remaining six subfamilies of Araceae. This divergence dates to c. 122 Ma (CI, 112–132 Ma; node 2 in all figures and tables). All living gymnostachyoid/orontioid species are restricted to wet habitats (Bogner et al., 2007). Next-oldest divergences involve the entirely aquatic Lemnoideae, dating to c. 104 Ma (CI, 93–113 Ma; node 6), and the split between the Australian Gymnostachys and the Northern Hemisphere orontioids, dating to c. 96 Ma (CI, 73–115 Ma; node 3). Considering the near-basal position of these wet habitat-adapted lineages and the large number of aquatic lineages in the Araceae sister group (most of the 13 other Alismatales families), an origin of aroids in water-associated swampy habitats is plausible. Second, Late Cretaceous and Palaeocene fossils of free-floating Araceae (Limmnobiphyllum scutatum and Cobbania corrugata; Stockey et al., 1997, 2007; Hoffman & Stockey, 1999) indicate that transitions from a terrestrial growth habit to an aquatic one had occurred early during the evolution of the family. Such transitions are known from other monocots (Cook, 1999) and, even within Araceae, a free-floating habit evolved a third time in the ancestor of the monotypic genus Pistia, a member of the derived subfamily Aroidae (Figs 3, S2, node 95, perhaps in the Eocene).

That Laurasia was a region of early Araceae evolution receives support from fossils from the late Aptian to early Albian Figueira da Foz Formation in Portugal (Friis et al., 2010). A newly discovered orontioid leaf fossil from the late Aptian Crato Formation in Brazil indicates that Araceae at that time were also in West Gondwana (C. Coiffard and B. Mohr, Natural History Museum, Berlin, Germany; see by us in May 2011; Table S2). Until c. 65 Ma, the southern tip of West Gondwana/South America provided the only overland connection between West and East Gondwana (Reguero et al., 2002: Fig. 3; Iglesias et al., 2011: Fig. 1d), and Antartica therefore could have been the route by which the gymnostachyoid clade reached Australia, where it today has a single surviving species. Our AAR, however, does not capture this, probably because of the lack of living or extinct Araceae from Antarctica.
Biogeographically, the presence of orontioid/gymnostachyoid Araceae in the Cretaceous of Brazil and in today’s Australia resembles the history of another ancient family, the Callalycanthaceae in the Laurales. Members of Calycanthaceae are first known from c. 115-Ma-old Crato Formation fossils resembling *Calycanthus* (today three species in North America and China), but also the monotypic Australian genus *Idiopernum* (Mohr & Eklund, 2003). The split between the Northern Hemisphere Calycanthaceae (nine living species) and the Australian *Idiopernum* has been dated to the Late Cretaceous, and a trans-Antarctic overland connection between South America and Australia has been invoked (Zhou et al., 2006). The orontioid Araceae, moreover, resemble Calycanthaceae in exhibiting Miocene Beringian disjunctions, in their case involving the North American/Chinese genera *Calycanthus* and *Chimonanthus* (Zhou et al., 2006) and, in the case of the Araceae, the North American/Asian genera *Symlocarpus* and *Lycychiton* (Nie et al., 2006; our Fig. 3, node 5).

High substitution rates in aquatic Araceae

Seven of the 10 highest DNA substitution rates in the Araceae occur in the free-floating or submerged Lemnoideae, *Pistia* and Cryptocoryneae (Fig. S1, nodes 2, 6, 7 and 10). The literature on molecular substitution rate variation is vast, and there is evidence that rates in both animals and plants can vary with body size, population dynamics, lifestyle and geographical location (Lynch, 2007; Bromham, 2009). Because of this plethora of causes, it is not currently possible to decide whether it is the small body size of aquatic aroids, their mostly clonal reproduction *Calycanthus* and *Chimonanthus* (Zhou et al., 2006) and, in the case of the Araceae, the North American/Asian genera *Symlocarpus* and *Lycychiton* (Nie et al., 2006; our Fig. 3, node 5).

Extinction in the Northern Hemisphere correlated with climate cooling

Modelling work on the shapes of phylogenetic trees has shown that a long stem leading to a cluster of short branches can indicate a mass extinction (Harvey & Rambaut, 2004; Crisp & Cook, 2009). In the Araceae, the longest branch is that leading from the stem lineage of Lasioideae (Fig. 3, node 28: 90 Ma) to the crown group (node 29: 26 Ma). Lasioi fossils are known from the Late Cretaceous of Siberia and the Eocene of Canada (Smith & Stockey, 2003; Hofmann & Zetter, 2010), but, today, members of Lasioideae survive only in tropical South America, South-East Asia and Africa. The subfamily was thus once more widespread and probably experienced extinction in the Northern Hemisphere when the climate deteriorated at the end of the Oligocene.

Extinction in Eurasia and North America and survival in tropical South-East Asia, South America or Africa indeed seems to be the prevailing pattern in Araceae. Similar effects of Oligocene climate cooling and the Quaternary ice ages have been documented for many plant groups (e.g. Latham & Rickles, 1993; Tiffney & Manchester, 2001) which, today, are restricted to tropical America and/or Africa, but in the early Tertiary occurred in Europe, including Anacardiaceae (*Anacardium*), cycads (*Ceratozamia*) and Malphigiaceae (*Tetrapeteris*; Manchester et al., 2007). Which of these groups spread across the North Atlantic land bridge, linking North America and Europe by way of Greenland, and which via Beringia, requires case-by-case analyses, and, for the Araceae, remains an open question.

In summary, all eight subfamilies of Araceae formed before the K/T boundary, supporting the view that this extinction event, which was so important for large-bodied animals, had a minor impact on plants, with no major plant groups disappearing at the boundary and the damage primarily confined to the species level (Nichols & Johnson, 2008). Of the 3790 species of Araceae described so far, 18 occur in Australia, 17 in Madagascar and 129 in Africa, whereas 1525 are known from the Neotropics and the remainder from tropical Asia and the Malesian archipelago (> 1000 species in some 40 genera; Boyce & Croat, 2011). Yet, the few species in Australia, North America and Eurasia represent more ancient surviving lineages than does the entire tropical Asian region (Fig. 3).

Including past continental plate positions and extinct ranges in biogeographical models

The biogeographical approach used here combines advances made over the past 3 yr (Table 1). Specifically, we modelled changing migration pathways (using different time-slice models) and included ranges that are no longer occupied by adding fossils to the tree file. Any biogeographical model can only 'reconstruct' (infer) areas that are included in the analysis. For example, migration across Eocene Antarctica can only be inferred when Antarctic is included as an operational geographical unit and its connectivity to South America, Australia and India is part of the probability matrix. Thus, one would ideally include all known geographical ranges for a clade by incorporating the location of all of its (well-studied) fossils directly into the analysis. Such formal (quantitative, with measures of uncertainty) AAR goes beyond what can be learned from the fossil record per se because it links fossils and their locations with the distribution of living clades in a researcher-driven model.

Acknowledgements

We thank Josef Bogner for many fruitful discussions on Araceae evolution and fossils. Financial support was obtained from the German National Science Foundation (RE 603/11-1).

References


Bogner J, Hoffman GL, Aulenback KR. 2005. A fossilized aroid infructescence, *Hypocalyx orontoides* gen. nov. et sp. nov. of Late Cretaceous (Late Campanian) age

© 2012 The Authors

New Phytologist © 2012 New Phytologist Trust


**Supporting Information**

Additional supporting information may be found in the online version of this article.

**Fig. S1** Inferred DNA substitution rates in the Araceae.

**Fig. S2** Results from divergence dating and ancestral area reconstruction (AAR) without fossils.

**Fig. S3** Posterior distribution of fossil constraints from the BEAST runs with the simple substitution model (JC + F) with either uniform (left column) or gamma-shaped (right column) prior distributions.

**Table S1** List of species used in this study with author names, herbarium vouchers and GenBank accession numbers for all sequences.

**Table S2** Fossils (sorted by age) used as constraints in this study or discussed in the main text.
Table S3 Dispersal matrices employed for the maximum likelihood ancestral area reconstructions used in the four-time-slice model (90–60, 60–30), the three-time-slice model (90–30) or both (150–90, 30–0).

Table S4 Divergence age estimates under different molecular clock models in TreeTime, and prior constraint schemes (uniform or gamma) and substitution models (simple, JC + Γ; complex, TPM1uf + Γ for the noncoding region and GTR + Γ for the coding region taking into account the third codon position) in BEAST.

Table S5 Results for each node shown in Fig. 3

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the New Phytologist Central Office.