American Journal of Botany

American Journal of Botany 98(4): 1–15. 2011.

Relationships within the Araceae: comparison of morphological patterns with molecular phylogenies[1]

Natalie Cusimano[2,11], Josef Bogner[3], Simon J. Mayo[4], Peter C. Boyce[5], Sin Y. Wong[6], Michael Hesse[7], Wilbert L. A. Hetterscheid[8], Richard C. Keating[9], and Jim C. French[10]

[2]Department of Biology, LMU Munich, D-80638 Munich, Germany; [3]Augsburger Str. 43a, D-86368, Gersthofen, Germany; [4]Royal Botanic Gardens Kew, Richmond, Surrey, TW9 3AE, U.K.; [5]BRT Research Associate, Forest Herbarium (BKF), The Office of Forest and Plant Conservation Research, National Park, Wildlife and Plant Conservation Department, 61 Phahonyothin Road, Chatuchak, Bangkok 10900, Thailand; [6]Department of Plant Science and Environmental Ecology, Faculty of Resource Science and Technology, Universiti Malaysia Sarawak 94300 Samarahan, Sarawak, Malaysia; [7]Department of Structural and Functional Botany, University of Vienna, 1030 Vienna, Austria; [8]Von Gimborn Arboretum, Velperengh 13, 3941 BZ Doorn, Netherlands; [9]Missouri Botanical Garden, P.O. Box 299, St. Louis, Missouri 63166-0299, USA; and [10]188 San Jose Court, San Luis Obispo, California 93405, USA

- *Premise of the study:* The first family-wide molecular phylogeny of the Araceae, a family of about 3800 published species in 120 genera, became available in 1995, followed by a cladistic analysis of morpho-anatomical data in 1997. The most recent and comprehensive family-wide molecular phylogeny was published in 2008 and included species from 102 genera. We reanalyzed the molecular data with a more complete genus sampling and compared the resulting phylogeny with morphological and anatomical data, with a view to contributing to a new formal classification of the Araceae.
- *Methods:* We analyzed 113 aroid genera and 4494 aligned nucleotides that resulted from adding 11 genera to the 2008 molecular matrix. We also analyzed 81 morphological characters in the context of the molecular phylogeny, using an extended version of the 1997 morpho-anatomical data set.
- *Key results:* The resulting maximum-likelihood phylogeny is well resolved and supported, and most of the 44 larger clades also have morphological or anatomical synapomorphies as well as ecological or geographic cohesion. Of the 44 clades, 16 are here newly circumscribed and informally named. However, some relationships remain poorly supported within the Aroideae subfamily. The most problematic placement is *Calla* within Aroideae, which conflicts with the distribution of morphological, anatomical, and palynological character states.
- *Conclusions:* The comparison of the molecular analysis with morphological and anatomical data presented here represents an important basis for a new formal classification for the Araceae and for the understanding of the evolution of this ancient family, a monocot group known in the fossil record from the early Cretaceous.

Key words: Araceae; *Calla*; character evolution; classification; Lemnoideae; molecular phylogeny; phenotypic characterization.

In recent years, the phylogeny and evolution of the Monocots have come under intense scrutiny with the rapid development of molecular phylogenetic systematics (e.g., Barfod et al., 2010) and these studies have highlighted the position of the Araceae as an early-diverging Monocot clade, within which the duckweeds (Araceae, Lemnoideae) have evolved (Cabrera et al., 2008). During the same period, exciting new fossil discoveries have been made in the Araceae (e.g., Friis et al., 2004). These have pushed back the history of the family to the early Cretaceous and justify an increased focus on the study of phylogeny and character evolution in this family.

Since the landmark study of French et al. (1995), phylogenies based on molecular data have been the primary basis for interpreting patterns of relationships in the Araceae at the suprageneric level (Barabé et al., 2002; Renner and Zhang, 2004; Renner et al., 2004; Rothwell et al., 2004; Tam et al., 2004; Gonçalves et al., 2007; Cabrera et al., 2008, Gauthier et al., 2008; Wong et al., 2010). The most comprehensive molecular analysis to date has been provided by Cabrera et al. (2008), a study that effectively settled the long-standing question of the relationships of the duckweeds (the former Lemnaceae, now Araceae subfamily Lemnoideae), using a matrix of 102 aroid genera (including the duckweeds) and 5188 aligned base pairs of chloroplast DNA. In attempting to transform such phylogenies into a formal classification, it is desirable to compare them with phenotypic data sets so as to highlight clades that are supported by distinctive morphological or anatomical synapomorphies and those that are not, but are nevertheless well supported by molecular synapomorphies. Keating (2002), for example, was able to interpret his vegetative anatomical data using the phylogeny of French et al. (1995), leading him to propose a new formal classification of the family. Bogner and Petersen (2007) presented an updated version of the classification of Mayo et al. (1997), which itself emerged as a result of comparison of morpho-anatomical data with French et al.'s (1995) molecular tree.

The availability of the morpho-anatomical data set of Mayo et al. (1997), molecular sequence data of Cabrera et al. (2008), and restriction-site data of French et al. (1995) prompted us to

[1] Manuscript received 1 May 2010; revision accepted 3 January 2011.

For discussions and comments on the manuscript we thank Susanne S. Renner, Judy Jernstedt, Mark P. Simmons and two anonymous reviewers. We thank Lidia Cabrera, Gerardo Salazar and Mark Chase for letting us use their molecular data matrix, and Thomas B. Croat for help with the original morpho-anatomical data matrix.

[11] Author for correspondence (e-mail: cusimano@bio.lmu.de)

doi:10.3732/ajb.1000158

carry out a new study with a view to contributing to a new formal classification of the Araceae. Here, we report the results of analyses of these three data sets (augmented versions in the case of the morpho-anatomical and sequence data). Both separate and combined analyses (total evidence methods) were undertaken during this study, but in light of the results, we concluded that the most useful approach for the purposes of a future formal classification would be to trace the morpho-anatomical characters onto the molecular sequence phylogeny. Our aim is to highlight well-supported clades with distinct phenotypic characterization and show those areas of conflict that require further investigation.

The main part of the discussion thus concentrates on 44 robust molecular clades, most of which are also supported by morphological synapomorphies, reinforcing proposals previously made by Cabrera et al. (2008) for modifying the family classification. Sixteen of these clades are newly presented here, but they are currently considered informal taxonomic groups. All early-diverging subfamilies and the relationships between them are well supported. The major subclades within Aroideae are each well supported, but the relationships between them are still not resolved. The only significant case in which molecular and morphological data seem to contradict one another is the location of *Calla* within the Aroideae.

BACKGROUND

The most detailed modern taxonomic study of the genera of the Araceae is the monograph by Mayo et al. (1997; *The Genera of Araceae*, hereafter GoA), which, although it treats the morphology of most of the currently recognized genera, did not include the duckweeds, then still regarded as a separate family. To provide the framework for the GoA classification, Mayo and colleagues carried out a maximum-parsimony (MP) cladistic analysis, using a matrix of 63 morphological and anatomical characters for 107 genera (including three outgroups), assembled from the literature and by examination of living, spirit, and herbarium materials. This cladistic study (the first MP analysis of morpho-anatomical data within Araceae) was motivated by, and based on, the pioneering work of Grayum (1984, 1990, 1992), who made a very wide-ranging revision of the taxonomic literature of the family and conducted a comprehensive SEM pollen study that provided the basis for his classification, itself derived by informal cladistic methodology.

It was French et al. (1995), however, who published the first computer-generated cladistic analysis of the Araceae, and they were also the first authors to present a within-family cladogram based on molecular patterns (restriction-site data from chloroplast DNA), using a matrix of 88 aroid genera and 488 characters. Their cladistic results, along with those of the independent morpho-anatomical GoA study, were first made public at the International Symposium *Monocotyledons: Systematics and Evolution* held at the Royal Botanic Gardens, Kew, in 1993 (Rudall et al., 1995) and led to discussions about combining the two matrices for further studies. This did not come to fruition, but the GoA classification eventually published in 1997 was strongly influenced by French et al.'s (1995) results, as can be seen from the discussion in Mayo et al. (1997, chapter 21). The GoA cladistic analysis was presented orally at the Tokyo International Botanical Congress in 1993 but never published in full, and the morpho-anatomical matrix on which it was based is presented for the first time here, albeit in a new and extended version.

MATERIALS AND METHODS

Character Matrix and Data Analyses—Most of the morpho-anatomical data were gathered during the preparation of the genus descriptions for GoA, when the morphology and anatomy of the stem, leaf, inflorescence, fruits, and seeds were reexamined using existing taxonomic literature supplemented by observations made from specimens in the herbaria, spirit and living collections of the Royal Botanic Gardens Kew, and the Munich Botanical Garden. The morphological and anatomical characters used here are mostly documented by Mayo et al. (1997), Grayum (1984, 1990, 1992), and Keating (2002), together with the literature cited in those works. We have added data sets for the lemnoid genera (*Lemna, Spirodela, Landoltia, Wolffia, Wolffiella* from Landolt, 1986, 1998 and Landolt and Kandler, 1987) and for more recently published genera not included in GoA. The morphological and anatomical characters are described in Appendix S1 (see Supplemental Data online at http://www.amjbot.org/cgi/content/full/ajb.1000158/DC2). Where no references are given, GoA is our primary information source. The resulting matrix consists of 81 characters for 109 genera of Araceae and one outgroup taxon, *Acorus*. In the original matrix, polymorphic characters were coded as ambiguities, but for the present analysis, where possible, we inferred an ancestral character state (IAS) for polymorphic characters because this has been found to yield more reliable results in analyses of higher-level taxa (Simmons, 2001, and references therein). The IAS matrix is presented in Appendix S2 (see Supplemental Data; for editable versions of the IAS matrix as well as the original morpho-anatomical matrices, see http://scratchpad.cate-araceae.org/). The chloroplast restriction site (RFLP) data matrix of French et al. (1995) included 88 aroid genera and 488 characters, with *Acorus* as outgroup (for editable version, see http://scratchpad.cate-araceae.org/).

The alignments of the six chloroplast markers of Cabrera et al. (2008; *rbcL*, *matK*, partial *trnK* intron, partial *tRNA-Leu* gene, *trnL–trnF* spacer, and partial *tRNA-Phe* gene), including 102 aroid genera and seven outgroup taxa, were obtained from TreeBase (for GenBank numbers, see Cabrera et al., 2008: appendix 1). We completed it by adding sequences from the six accepted genera not then included (*Anaphyllum, Croatiella, Furtadoa, Theriophonum, Zomicarpa*, and *Asterostigma* (the single *Asterostigma* species sampled by Cabrera et al. (2008) is now classified as *Incarum pavonii*), two recently published genera, *Bakoa* and *Schottariella* (Boyce and Wong, 2008, 2009), two other genera recently resurrected (*Philonotion,* Wong et al., 2010; *Sauromatum*, Cusimano et al., 2010), and an Australian genus composed of species previously assigned to *Typhonium* (the name *Lazarum* A. Hay is available but cannot yet be applied, pending availability of new material; Cusimano et al., 2010), giving a total of 113 genera. We also added a second accession of *Calla* as a check, and used *Tofieldia* (Tofieldiaceae), *Acorus* (Acoraceae), and *Hedyosmum* (Chloranthaceae) as outgroups. Appendix 1 shows the sources of the sequences added for the additional species. Sequences of two different species (*Typhonium horsfieldii*, *T. hirsutum*) have been combined to represent the genus *Sauromatum*. Several of the sequences of the additional species were available in GenBank, and the missing sequences were generated according to the methods described in Cabrera et al. (2008) and deposited in GenBank (accession nos. HQ687765–HQ687767). Alignment of the new sequences was first done automatically in MacClade 4.08 (Maddison and Maddison, 2005) and afterwards adjusted visually, trying to maximize similarity (Simmons, 2004). Unlike Cabrera et al. (2008), we did not use gap-coding, because it did not increase support or resolution either in the resulting phylogeny from Cabrera et al.'s (2008) MP analysis or from our partitioned MrBayes analyses of the gap-coded molecular data (data not shown). The sequence data matrix and the resulting trees were deposited in TreeBase (http://purl.org/phylo/treebase/phylows/study/TB2:S11083).

All three data sets were analyzed with a Bayesian Markov-chain Monte Carlo (MCMC) approach (Yang and Rannala, 1997), using MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). Searches for the RFLP and morphological data sets relied on an F81 and a JC model, respectively; transition rates of the binary (RFLP) model relied on the stationary frequencies of 0 and 1; rates of the morphology model were all equal; to correct for the coding bias (i.e., all absent and/or all present characters are not detected), coding was set to "variable". The search for the sequence data relied on the GTR + Γ model with the number of gamma categories set to four (Yang, 1993). We used a flat Dirichlet prior for the relative nucleotide frequencies and rate parameters, a discrete uniform prior for topologies, and an exponential distribution (mean 10) for the γ-shape parameter and all branch lengths.

Bayesian runs were started from independent random starting trees and repeated four times. Markov chain Monte Carlo runs extended for 10 million generations for the RFLP and morphological data and 8 million runs for the molecular sequence data, with trees sampled every 100th generation (resulting in 100 000 and 80 000 trees, respectively, for each run). Besides the standard