Dissertation
der Fakultät für Biologie der
Ludwig-Maximilians-Universität München

The Araceae as a Study System: Intron Evolution, Diversification Analyses, and Evolutionary Classification

Natalie Cusimano

Munich 2011

Erklärung

Diese Dissertation wurde im Sinne von §12 der Promotionsordnung von Prof. Dr. Susanne S. Renner betreut. Ich erkläre hiermit, dass die Dissertation nicht einer anderen Prüfungskommission vorgelegt worden ist und dass ich mich nicht anderweitig einer Doktorprüfung ohne Erfolg unterzogen habe.

Ehrenwörtliche Versicherung

Ich versichere hiermit ehrenwörtlich, dass die vorgelegte Dissertation von mir selbständig, ohne unerlaubte Hilfe angefertigt wurde.

Natalie Cusimano

1. Gutachter: Prof. Dr. Susanne S. Renner
2. Gutachter: Prof. Dr. Dirk Metzler

Dissertation eingereicht am: 29.09.2009
Mündliche Prüfung am: 24.01.2011

Note

In this thesis, I present the results from my doctoral research, carried out in Munich from May 2005 to October 2009 under the guidance of Prof. Susanne Renner. The results from my thesis have contributed to five manuscripts presented in Chapters 2 to 6. I also gave several presentations listed below. Except for Chapters 3 and 6, I generated all data and conducted all analyses myself. Writing and discussion involved collaboration with Susanne Renner. For Chapter 3, some of the sequences were generated by M. Barrett. My contributions to the manuscript in Chapter 6 were data analysis, writing and discussion with the co-authors.

Papers

Cusimano, N., Zhang, L.-B. and S.S. Renner. 2008. Reevaluation of the *cox1* group I intron in Araceae and angiosperms indicates a history dominated by loss rather than horizontal transfer. *Molecular Biology and Evolution* 25: 265-276.

Cusimano, N., Barrett, M., Hetterscheid, W.L.A. and S.S. Renner. 2010. A phylogeny of the Areae (Araceae) implies that *Typhonium, Sauromatum*, and the Australian species of *Typhonium* are distinct clades. *Taxon* 59(2): 439-447.

Cusimano, N. and S.S. Renner. The handling of missing species in diversification rate analyses – with empirical examples illustrating a new approach. In review at *Systematic Biology.* (2012: published by Cusimano, N., Stadler, T. and S.S. Renner in *Systematic Biology*: Advance Access published February 14. DOI: 10.1093/sysbio/sys031.)

Cusimano N. and S.S. Renner. 2010. Slowdowns in diversification rates from real phylogenies may not be real. In review at *The American Naturalist.* (2010: published in *Systematic Biology* 59(4): 458-464.)

Cusimano, N., Bogner, J., Mayo, S.J., Keating, R.C., Boyce, P.C., Wong, S.Y., Hetterscheid, W.L.A., Hesse, M. and J.C. French.

Relationships within the Araceae: Comparison of morphological patterns with molecular phylogenies. In preparation. (2011: published in *American Journal of Botany* 98(4): 654-668.)

Seminars

Cusimano, N. December 6, 2006. Evolution des *Pistia* clades: Vertikaler oder horizontaler Transfer des mitochondrialen *cox1* Introns? *Systematic Botany Seminar Series – LMU*. Munich, Germany.

Cusimano, N. January 16, 2008. Phylogeny and biogeography of the Areae (Araceae). *Systematic Botany Seminar Series – LMU*. Munich, Germany.

Cusimano, N. and S.S. Renner. April 10, 2008. Reevaluation of the *cox1* intron in Araceae and angiosperms indicates a history dominated by loss rather than horizontal transfer. *Systematics 2008*. Göttingen, Germany.

Cusimano, N. and S.S. Renner. August 11, 2008. Under the weather at different times: Evolution of the *Arum-Typhonium* clade in SE Asia and the Mediterranean area. Invited seminar, *Monocots IV Conference*. Copenhagen, Denmark.

Posters

Cusimano, N. and S.S. Renner. July 2007. Polyploidy, aneuploidy, and chromosome number evolution of *Arum* and its allies (Areae, Araceae). *The Origin and Evolution of Biota in Mediterranean Climate Zones - an integrative Vision*. Zurich, Switzerland.

Cusimano, N., Mayo, S.J. and J. Bogner. July 2009. Relationships within the Araceae: Comparison of MorphologicalPatterns with Molecular Phylogenies.*10th International Aroid Conference*, Nancy, France.

Part of my research was funded as follows:

- The three-month research period (Oct – Dec 2005) in the lab of William Friedman at the University of Colorado at Boulder by a grant from the NSF

Research Coordination Network

- Travels to the conferences in Zürich and Göttingen by the EES^{LMU} global learning fund

- The research visits to the Araceae collection of Wilbert Hetterscheid in Wageningen (Aug 2006) and to Vienna for practial work in the lab of Privatdozent Dr. Hanna Weiss-Schneeweiss at the University of Vienna (Sep 2007) by the "Gesellschaft der Freunde des Botanischen Gartens München"

- Travel to the Monocots IV Conference by the Münchner Universitätsgesellschaft

Für Omina.

Contents

List of Figures

List of Tables

Chapter 1

General Introduction and Discussion

1.1 Summary

This thesis addresses five issues concerning the evolution and diversification of plants, focusing on the family Araceae as a study system: (1) The evolution of an intron in the *cox1* gene, (2) testing the monophyly of the genus *Typhonium*, (3) testing the handling of missing species in diversification analyses, presenting a new approach, (4) the influence of species sampling on the inference of diversification patterns, and (5) the evolution of morphological and anatomical traits, chromosome numbers, and embryology based on reconstructed phylogenetic relationships. These issues are investigated in the Araceae as a whole and in several clades within the family: the *Pistia* clade, the Areae, and two subclades of Areae, namely *Typhonium* and *Arum* and their relatives. I also studied the evolution of the *cox1* intron throughout the angiosperms and used a published phylogeny of *Momordica* to test methods for diversification analyses.

In the first part (Chapter 2), I investigated the history of a widespread, but not universal, intron in the *cox1* gene of Araceae and other angiosperms. Newly generated sequence data and analyses support vertical inheritance within the *Pistia* clade, however, I also discovered a clear *cox1* intron loss, the first such case documented in the angiosperms. Based on a relaxed molecular clock, the intron has been present in the Araceae for at least 70 million years. Analyses of the *cox1* intron sequences and their flanking coconversion tracts in the context of an angiosperm phylogeny point to a few early gains, followed by numerous losses, contradicting the prevailing interpretation of thousands of horizontal transfer events.

The reconstructed Areae phylogeny clearly reveals that the genus *Typhonium* is not monophyletic, but splits instead in three clades. The largest clade includes the type species, *T. trilobatum*. The second clade comprises all analyzed Australian endemics, and the third all former *Sauromatum* species. In Chapter 3, I focus on the latter clade, discussing morphology and chromosome numbers, both supporting the molecular results, and I address the necessary taxonomic changes, namely the resurrection of the genus *Sauromatum* and five new combinations. Additionally, I present an updated key to the Areae genera and a new key for the *Sauromatum* species.

Large time-calibrated phylogenies are now readily obtained and are increasingly being used to infer diversification patters. However, inferring rates of di-

versification is statistically challenging, and the sensitivity of methods when their underlying assumptions are not met is poorly understood. A major problem in diversification analysis is incomplete species sampling. Chapter 4 presents a review of plant diversification studies, which revealed that the majority of published studies is based on species sampling of <60% and that the handling of missing species is inconsistent. I applied different approaches for inferring diversification rate changes over time and for the handling of missing species (using two non-nested subclades of the Areae), and then present a new approach for handling the missing species. My new approach yielded clearly different results than the "classical" methods. It involves model-based data augmentation and imputation, and is therefore statistically preferable. The classical methods only test the incomplete phylogeny, and additionally are based on the assumption that species sampling is random. The observation that this might not be the case, led me to explore the effects of different sampling strategies (random, cutting off whole clades, sampling deep nodes) in a complete empirical and two simulated phylogenies with constant diversification rates (Chapter 5). Results revealed that downturns in diversification rate are inferred with high confidence when deep nodes are oversampled.

In collaboration with other Araceae taxonomists, palynologist and morphologists, I conducted analyses of a morphological and anatomical data matrix that comprises 81 characters coded for 109 genera, and of a completed molecular Araceae data matrix including 113 species representing all genera (Chapter 6). Most of the relationships found in the resulting phylogeny are well supported by morphological-anatomical characters. However, relationships of major clades within the Aroideae subfamily remain unresolved, and the inclusion of *Calla* in the Aroideae is contradicted by several morphological characters. On this basis, 47 clades could be described of which 19 are newly discovered, forming the basis of a new formal classification. In still ongoing projects, I have gathered original data on the chromosome numbers of some Areae and reviewed the embryology and early endosperm development of the *Pistia* clade (Sections 1.4 and 1.5).

1.2 Introduction to the Study Groups

Araceae

The Araceae, a family in the early-diverging monocot order Alismatales, comprises about 4000 species in 113 genera (Bogner and Petersen, 2007; Boyce and Wong, 2008, 2009; Wong et al., 2009; Cusimano et al., 2009). Most of the species are tropical herbs. The oldest fossils attributed to Araceae are 120 my old (Friis et al., 2004), and it is clear that Araceae have undergone multiple radiations and waves of extinction (especially in the northern hemisphere) over their long history. Araceae make up an important part of nearly all perhumid tropical biomes of the Old as well as of the New World, where the family shows a high diversity in terms of species numbers and life forms (geophytes, helophytes, (hemi-) epiphytes, free-floating aquatics). This diversity is also reflected in the wide variety of habitats occupied, and the ecology and morphology of vegetative and generative structures. Especially characteristic are the leaves of Araceae with their venation patterns ("You can identify nearly every genus by a simple fragment of the leaf from the midrib to the edge", Josef Bogner, pers. com., Copenhagen Aug 2008). Another striking feature are the inflorescences, which consist of a fleshy axis, called the spadix that bears small, mostly highly reduced flowers, arranged in spirals and subtended by the leaf- or petal-like bract. Flowers can be bisexuell and inflorescences then looking as in Fig. 1.1A, or unisexuell. In monoecious species the female flowers are positioned at the base of the spadix and the male flowers above them, usually separated by a sterile zone (Figs. 1.1B, 1.4, 3.3). The flowers are mostly reduced to either carpels or stamens. Below the female flowers, between female and male, or above the male flowers, sterile flowers (pistillodes or staminodes) can be present in varying combinations, numbers, shapes, colors and sizes (Fig. 1.1C). Above the flower-bearing, fertile, zone the spadix can be extended in a sterile appendix (Mayo et al., 1995, 1997).

Araceae morphology is relatively well known, partly because the family received the attention of some of the best plant morphologists of the 19th and 20th century, including H. W. Schott, A. Engler, P. van Tieghem, and W. Troll as well as that of many, aroid "aficionados," partly because many species are in cultivation and horticulturally important. Most of today's Araceae researchers are professionals,

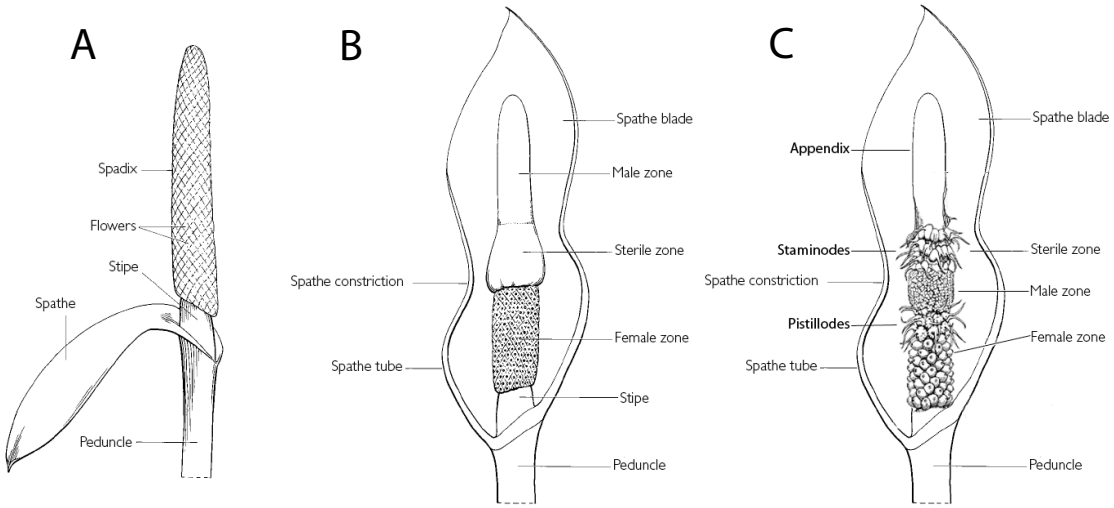


Figure 1.1: Inflorescences types: **A** bisexual flowered spadix with a simple undifferentiated spathe; **B** unisexual flowered undifferentiated spadix with female flowers in the lower part of the spadix and male flowers above, and with a spathe divided into a limb (blade) and a convolute lower tube; **C** bisexual flowered spadix with pistillodes, staminodes, and a spadix appendix (after Mayo et al., 1997).

such as Josef Bogner, Peter Boyce, Tom Croat, Simon Mayo, Jin Murata, Marija Bedalov and Wilbert Hetterscheid. These workers have conducted much field work, have cultivated specimens and have specialized on subgroups without loosing the overview over the whole family. The state of Araceae systematics is still best represented in the book "The genera of Araceae" by Mayo et al. (1997), updated by Bogner and Petersen (2007). In the last two years six new/resurrected genera have been published *Bakoa* and *Schottariella* belonging to the Schismatoglottidae (Boyce and Wong, 2008, 2009), and *Philonotion* (Wong et al., 2009), or will be published in near future, namely *Sauromatum* (Cusimano et al., 2009), *Lazarum* and *Schottariopsis*.

French et al. (1995) were the first to test these morphology-based hypotheses about Araceae evolution and phylogenetics with molecular data from chloroplast restriction sites involving species from 86 of the genera. The main question they addressed was the position and relationships of the two clades of free-floating Araceae, namely the monotypic genus *Pistia*, the water lettuce, one of the World's worst weeds, and the Lemnaceae (duckweeds), at that time treated as their own

family (*Landoltia, Lemna, Spirodela, Wolfia* and *Wolfiella*). Because of their reduced habits, these aquatic clades lack taxonomically important characters, and were therefore difficult to compare and consequently also difficult to place. Linnaeus (1754) nevertheless placed the genus *Pistia* in the Araceae, correctly as it turned out. Grayum (1990) was among the first to suggest that the Lemnaceae might be nested within the Araceae family and that these two clades are not closely related. Their free-floating life form would thus have evolved independently. The analysis of French et al. (1995, Fig. 1.2) confirmed these hypotheses: *Lemna* and *Pistia* are both nested in the Aroideae, but in distant positions: *Pistia* was in a well-supported with 13 highly derived geophytic Aroideae, whereas *Lemna* is more early-diverging in the Aroideae. Furthermore, with *Acorus* as outgroup (Grayum, 1987; Duvall et al., 1993), French et al. (1995) found the Orontioideae as the first-diverging clade, followed by four other major clades, the Lasieae, the Pothoideae, Philodendroideae and Aroideae (as sister to *Calla*) sensu Grayum (1990, with minor rearrangements). This first molecular analysis was followed by several investigations focussing on subclades of the Araceae, i.e., *Symplocarpus*, Thomsonieae, *Schismatoglottis, Amorphophallus, Arisaema,* Monsteroideae, Orontioideae, Spathicarpeae, *Philodendron, Pistia* clade (Wen et al., 1996; Grob et al., 2002, 2004; Barabé et al., 2004; Jung et al., 2004; Tam et al., 2004; Renner and Zhang, 2004; Renner et al., 2004; Nie et al., 2006; Gonçalves et al., 2007; Gauthier et al., 2008; Mansion et al., 2008).

In September 2008, Cabrera et al. published another molecular study involving a nearly complete genus sampling with 102 of the at that time 108 species (only species of *Anaphyllum, Croatiella, Furtadoa, Asterostigma (= Incarum pavonii), Theriophonum* and *Zomicarpa* were missing; Fig. 1.3). Their focus was the exact position of the duckweeds within the Araceae, which was still dubious. Duckweeds came out as the second-diverging clade in the Araceae, after the also aquatic Orontioideae-Gymnostachydoideae clade ("proto-aroids"). In addition, the monophyly of the subfamilies Pothoideae, Monsteroideae and Lasioideae was supported. Monophyly of the subfamily Aroideae requires including Calloideae (*Calla*) in Aroideae. Monophyly of some tribes in their current determination (e.g., Colocasieae, Monstereae) needs further testing.

Over the past 10 years, Simon Mayo and Josef Bogner have extended and completed the morphological-anatomical data matrix on which they based their cladis-

tic analysis in "The Genera of Araceae" (Mayo et al., 1995, 1997). Now it includes 81 characters of all genera except *Bakoa*, *Schottariella*, *Philonotion*, *Sauromatum* and *Lazarum*. I explored the phylogenetic signal in these data. I also reanalyzed the coded RFLP data of French et al. (1995) and added 11 *matK* sequences to the data set of Cabrera et al. (2008) for a now complete genus sampling of 113. This manuscript is presented as Chapter 6. Most of the five (non-monogeneric) subfamilies and the 20 (non-monogeneric) tribes are revealed by molecular data in addition to 19 new clades of different taxonomic levels fitting mostly well with morphology. We refrain from formal (Latin) names for these new clades that we discuss. Nevertheless, our informal system forms a basis for a new formal classification of the Araceae.

Pistia Clade

Pistia stratiotes and 13 other genera formed a well-supported clade in the RFLP tree of French et al. (1995, Fig. 1.2). For lack of a formal name, this clade is currently being referred to as the *Pistia* clade (Renner and Zhang, 2004). To investigate the biogeography of the *Pistia* clade, Renner and Zhang (2004) improved the sampling by including representatives of 16 genera including 36 of the 320 species. From three molecular markers they constructed a phylogeny and estimated the divergence times constraining it with fossils. Besides the two monotypic genera *Pistia* and *Protarum* they sampled Colocasieae (*Alocasia, Ariopsis, Colocasia, Remusatia, Steudnera; Gonatanthus* has been reduced to the synonymy of *Remusatia*), Arisemateae (*Arisaema, Pinellia*) and the Areae (*Arum, Biarum, Dracunculus, Eminium, Helicodiceros, Theriophonum, Typhonium*) sensu Mayo et al. (1997). The species of the mainly Southeast Asian *Pistia* clade occur in a wide range of habitats, including many species in the temperate zone, which is striking in a family that is otherwise almost restricted to warm and humid climates: About half of the species of the Areae occur in the Mediterranean region (see below), and the genera *Arisaema, Arum*, and *Pinellia* include dozens of cold-resistant species that occur in high latitudes or altitudes, for example, *Arum* in northern Europe and the Himalayan region. Many *Arisaema* species in northern China and in the Himalayas grow at altitudes well above 4000 m. Few of the *Pistia* clade species occur in Africa and only four in North America, namely *Pistia*

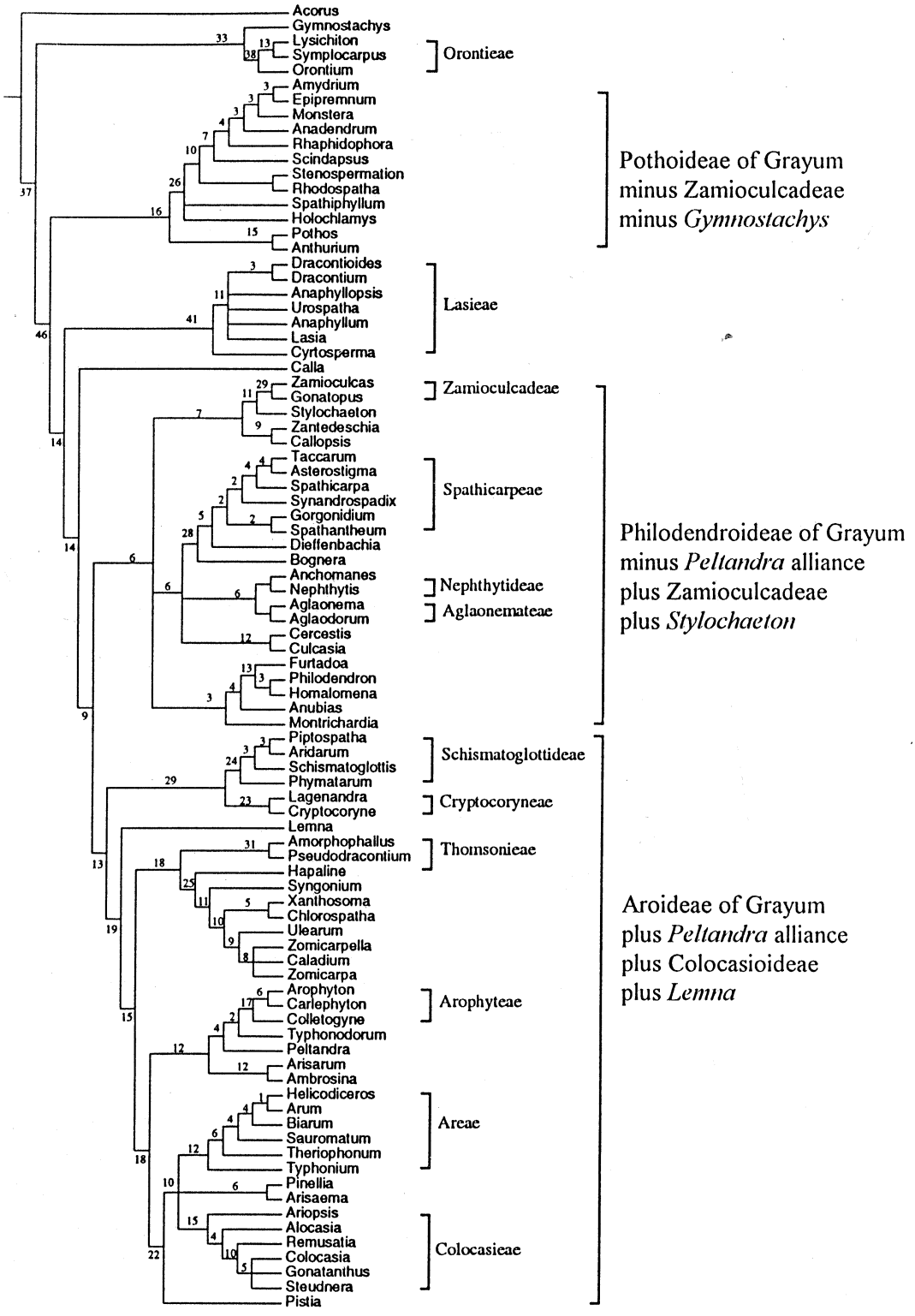


Figure 1.2: Strict parsimony consensus tree generated with PAUP, with number of restriction site changes at each node, calculated by MacClade. Tribal limits of Grayum (1990) are given only for the tribes with two or more genera that are monophyletic in the cpDNA tree. Subfamilial limits are shown for three polyphyletic Grayum (1990) subfamilies. (French et al., 1995; their Fig. 1)

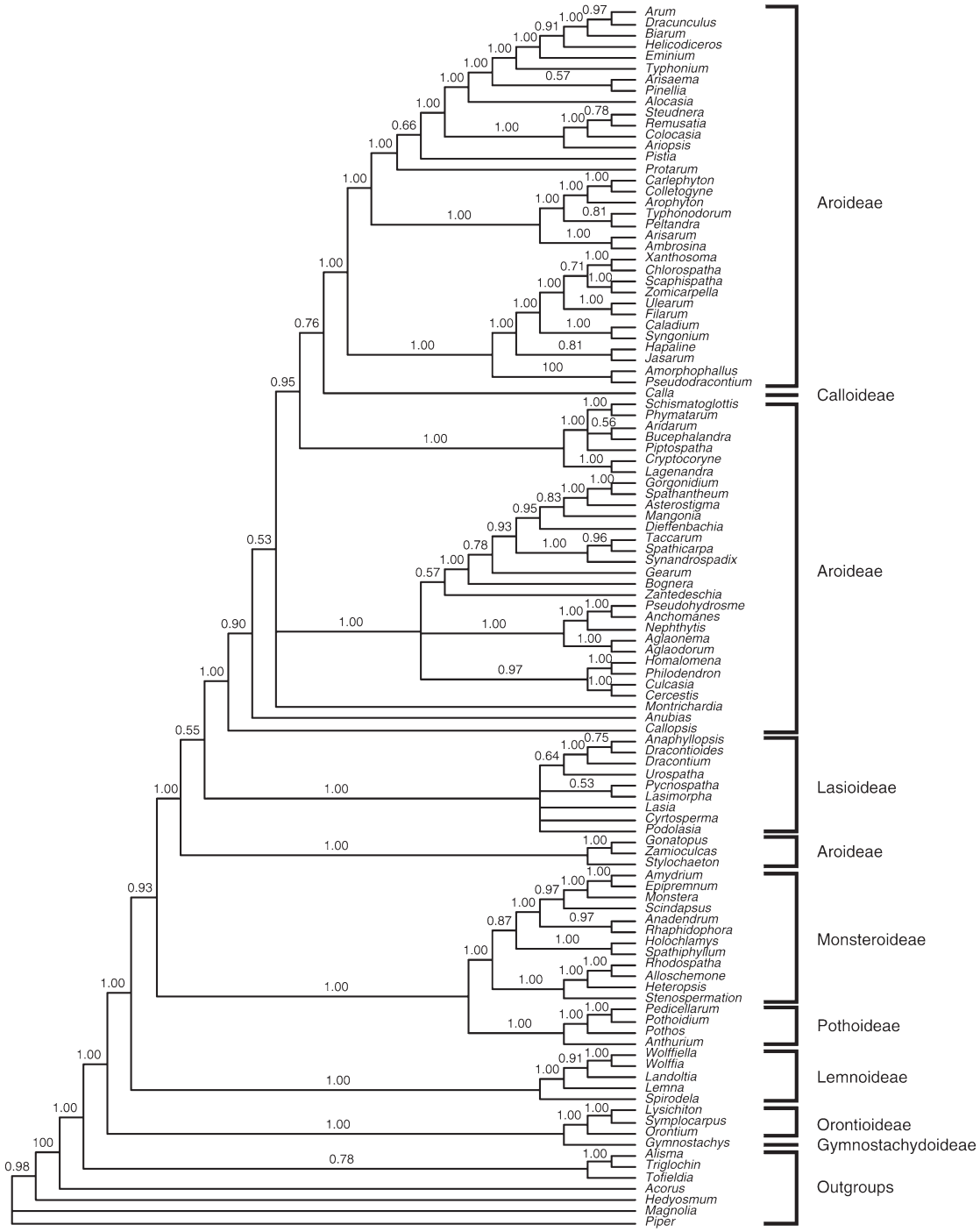


Figure 1.3: Bayesian summary tree from analysis of all regions combined. Numbers above branches are posterior probabilities. Bars indicate subfamilies recognized by Mayo et al. (1997). (Cabrera et al., 2008, their Fig. 2).

and three species of *Arisaema*. Whereas *Pistia* itself is a worldwide aquatic weed, *Protarum sechellarum* is endemic to the Seychelles Islands.

My paper published in Molecular Biology and Evolution on the mitochondrial *cox1* intron in plants used the *Pistia* clade as study system to test the hypothesis of vertical vs. horizontal transfer of the intron (Cusimano et al., 2008, Section 1.3 and Chapter 2). In this study, as well as in the study of diversification patterns within the Areae (Chapter 4), I used the *Pistia* clade to infer divergence time estimates, because Areae have no fossils, and so I used outgroup fossils for calibration.

Areae

The Areae sensu Mayo et al. (1997), a subclade of the *Pistia* clade, includes the name-giving genus of the family, *Arum*, and and six other genera. Whereas about one half of the species occur in Southeast Asia (*Typhonium* s.l. occurs in seasonal habitats of tropic Australia, Southeast Asia, and India, with one species also in Africa, and *Theriophonum* is endemic to India), the other half is centered in the Mediterranean region, a habitat that is too dry and too cold for most Araceae: *Arum* itself ranges from the Himalayan region to the West-Mediterranean, and in cold temperate Europe it extends to England and Southern Sweden. *Dracunculus vulgaris* occurs in scrub, woodland and macchia vegetation, under trees and shrubs, but also on stony open ground in the middle and eastern Mediterranean; *D. canariensis* occurs in forest and scrub vegetation on the Canary Islands; and *Helicodiceros muscivorus* occurs on limestone and granit rock crevices near the sea on Corsica, Sardinia, and the Balearic Islands. Finally, *Biarum* and *Eminium* include the most drought-adapted species in all Araceae, with ranges extending deep into the dry areas of the Near East (including the Negev desert), North Africa, and Southern Spain. There are only four other species of Araceae occurring in the Mediterranean region: the three species of the genus *Arisarum* and *Ambrosina bassii*.

I reconstructed a phylogeny of the Areae from sequences of three different markers and estimated the divergence times. I sampled species throughout all genera with a focus on *Typhonium* species (see below). Two subclades of this phylogeny, the Mediterranean *Arum* clade and *Typhonium*, were used to test different methods of correcting for missing species in diversification analyses. A review of such

analyses conducted in plant clades revealed a species sampling of <60% in the majority of the studies and inconsistencies in the application of the methods (Chapter 4). We tested three widely used methods for assessing diversification patterns over time together with the standard method for correcting for missing species. Additionally, I developed a new approach that is based on the assumption that species sampling is generally not random and that correction for missing species should be done by data augmentation and imputation. The classic method only tests the incomplete phylogeny and has the problem that it assumes that species sampling is random. The observation that this might not be the case led me to explore the effects of different sampling strategies (random, cutting off whole clades, sampling deep nodes) in a complete empirical phylogeny of the Cucurbitaceae genus *Momordica* and two simulated phylogenies, all with constant diversification rates (Chapter 5). Results revealed that downturns in diversification rate are always inferred with high confidence when deep nodes are oversampled. This introduces an important bias in diversification estimates inferred from undersampled phylogenies.

Finally, I was interested in investigating relationships and genus circumscriptions in Areae. Recently, two genera have been reduced to the synonymy of *Typhonium* because of lacking morphological differentiation and intermediate species: *Lazarum*, including only one Australian endemic, namely *L. mirabile*, was transferred into *Typhonium* in 1997 (Hay, 1997) and *Sauromatum* in 2000 (Hetterscheid and Boyce, 2000). Some *Typhonium* species had always be seen as close to *Sauromatum* because of morphological similarity (*T. brevipilosum, T. giganteum, T. hirsutum, T. horsefieldii*; Hetterscheid and Boyce, 2000). However, two studies suggested that *Typhonium* is might not be monophyletic: In the analysis of French et al. (1995), *Sauromatum* and *Typhonium* do not form a clade and Renner and Zhang (2004) found that all Areae might be embedded in *Typhonium*, therefore making it polyphyletic. Additionally, a restriction fragment analysis of chloroplast DNA (Sriboonma et al., 1993) and the analysis of a morphological data matrix (Sriboonma et al., 1994) could not reveal *Typhonium* as monophyletic and only failed to recover its polyphyly because outgroup choice was not apropriate. In Chapter 3, I show that *Typhonium* is indeed not monophyletic (Fig. 3.1). Instead, the 52 included species (of a total of 72 *Typhonium* species) fall into three distinct clades: The largest clade includes the type species of the genus, *Typhonium trilo-*

Figure 1.4: Longitudinally-opened spathes of 16 *Typhonium* species. Photos by Wilbert Hetterscheid.

batum (Figs. 1.4 and 3.1). A second clade is composed of species belonging to the former genus *Sauromatum* plus 5 five additional species (Figs. 3.1 and 3.3). A third clade is composed of all included Australian endemics (Fig. 3.1). I make the necessary transfers, present an updated key to the Areae genera and a new key for the *Sauromatum* species.

References

Barabé, D., C. Lacroix, A. Bruneau, A. Archambault, and M. Gibernau. 2004. Floral development and phylogenetic position of *Schismatoglottis* (Araceae). Int. J. Plant Sci. 165:173–189.

Bogner, J. and G. Petersen. 2007. The chromsome numbers of the aroid genera. Aroideana 30:82–90.

Boyce, P. C. and S. Y. Wong. 2008. Studies on Schismatoglottidae (Araceae) of Borneo VII: *Schottarum* and *Bakoa*, two new genera from Sarawak, Malaysian Bornea. Botanical Studies 49:393–404.

Boyce, P. C. and S. Y. Wong. 2009. *Schottariella mirifica* P.C. Boyce & S.Y. Wong: a new name for *Schottarum sarikeense* (Araceae: Schismatoglottideae). Botanical Studies 50:269–271.

Cabrera, L. I., G. A. Salazar, M. W. Chase, S. J. Mayo, J. Bogner, and P. Davila. 2008. Phylogenetic relationships of aroids and duckweeds (Araceae) inferred from coding and noncoding plastid DNA. Am. J. Bot. 95:1153–1165.

Cusimano, N., M. Barrett, W. L. A. Hetterscheid, and S. S. Renner. 2009. A phylogeny of the Areae (Araceae) implies that *Typhonium, Sauromatum*, and *Lazarum* are distinct clades. Taxon accepted, provided minor revisions.

Cusimano, N., L.-B. Zhang, and S. S. Renner. 2008. Reevaluation of the *cox1* group I intron in Araceae and angiosperms indicates a history dominated by loss rather than horizontal transfer. Mol. Biol. Evol. 25:265–276.

Duvall, G. H., M.ad Learn Jr., L. E. Eguiarte, and M. Y. Clegg. 1993. Phylogenetic analysis of *rbcL* sequences identifies *Acorus calamus* as the primal extant monocotyledon. Proc. Natl. Acad. Sci. USA 90:4641–4644.

French, J. C., M. G. Chung, and Y. K. Hur. 1995. Chloroplast DNA phylogeny of the Ariflorae. Pages 255–275 *in* Monocotyledons: Systematics and Evolution (P. J. Rudall, P. Cribb, D. F. Cutler, and C. J. Humphries, eds.). Royal Botanic Gardens, Kew, England.

Friis, E. M., K. R. Pedersen, and P. R. Crane. 2004. Araceae from the Early Cretaceous of Portugal: Evidence on the emergence of monocotyledons. Proc. Natl. Acad. Sci. USA 101:16565–16570.

Gauthier, M.-P. L., D. Barabe, and A. Bruneau. 2008. Molecular phylogeny of the genus *Philodendron* (Araceae): delimitation and infrageneric classification. Bot. J. Linnean Soc. 156:13–27.

Gonçalves, E. G., S. J. Mayo, M.-A. Van Sluys, and A. Salatino. 2007. Combined genotypic-phenotypic phylogeny of the tribe Spathicarpeae (Araceae) with reference to independent events of invasion to Andean regions. Mol. Phylogenet. Evol. 43:1023–1039.

Grayum, M. H. 1987. A summary of evidence and arguments supporting removal of *Acorus* from the Araceae. Taxon 36.

Grayum, M. H. 1990. Evolution and Phylogeny of the Araceae. Ann. Mo. Bot. Gard. 77:628–697.

Grob, G., B. Gravendeel, and M. Eurlings. 2004. Potential phylogenetic utility of the nuclear FLORICAULA/LEAFY second intron: comparison with three chloroplast DNA regions in *Amorphophallus* (Araceae). Mol. Phylogenet. Evol. 30:13–23.

Grob, G., B. Gravendeel, M. Eurlings, and W. Hetterscheid. 2002. Phylogeny of the tribe Thomsonieae (Araceae) based on chloroplast *matK* and *trnL* intron sequences. Syst. Bot. 27:453–467.

Hay, A. 1997. Two new species and a new combination in Australian *Typhonium* (Araceae tribe Areae). Edinb. J. Bot. 54:329–336.

Hetterscheid, W. L. A. and P. C. Boyce. 2000. A reclassification of *Sauromatum* Schott and new species of *Typhonium* Schott (Araceae). Aroideana 23:48–55.

Jung, Y., E. Song, S. Chun, K. Jang, M. Kim, S. Kang, and S. Kim. 2004. Phylogenetic analysis of plastid *trnL-trnF* sequences from *Arisaema* species (Araceae) in Korea. Euphytica 138:81–88.

Linnaeus, C. 1754. Araceae. Pages 411–415 *in* Genera Plantarum. Facsimile. J. Cramer, Weinheim.

Mansion, G., G. Rosenbaum, N. Schoenenberger, G. Bacchetta, J. A. Rossello, and E. Conti. 2008. Phylogenetic analysis informed by geological history supports multiple, sequential invasions of the Mediterranean basin by the angiosperm family Araceae. Syst. Biol. 57:269–285.

Mayo, S. J., J. Bogner, and P. C. Boyce. 1995. The Arales. Pages 277–286 *in* Monocotyledons: Systematics and Evolution (P. Rudall, C. P, C. DF, and H. CJ, eds.). Royal Botanic Gardens, Kew, England.

Mayo, S. J., J. Bogner, and P. C. Boyce. 1997. The genera of Araceae. The Trustees, Royal Botanical Gardens, Kew, UK.

Nie, Z.-L., H. Sun, H. Li, and J. Wen. 2006. Intercontinental biogeography of subfamily Orontioideae (*Symplocarpus, Lysichiton*, and *Orontium*) of Araceae in eastern Asia and North America. Mol. Phylogenet. Evol. 40:155–165.

Renner, S. S. and L.-B. Zhang. 2004. Biogeography of the *Pistia* clade (Araceae): Based on chloroplast and mitochondrial DNA sequences and Bayesian divergence time inference. Syst. Biol. 53:422–432.

Renner, S. S., L.-B. Zhang, and J. Murata. 2004. A chloroplast phylogeny of *Arisaema* (Araceae) illustrates tertiary floristic links between Asia, North America, and East Africa. Am. J. Bot. 91:881–888.

Sriboonma, D., M. Hasebe, N. Murakami, J. Murata, and K. Iwatsuki. 1993. Phylogeny of *Typhonium* (araceae) inferred from restriction fragment analysis of chloroplast dna. J. Plant Res. 106:11–14.

Sriboonma, D., J. Murata, and K. Iwatsuki. 1994. A revision of *Typhonium* (Araceae). J. Fac. Sci. Univ. Tokyo, III 14:255–313.

Tam, S., P. Boyce, T. Upson, D. Barabe, A. Bruneau, F. Forest, and J. Parker. 2004. Intergeneric and infrafamilial phylogeny of subfamily Monsteroideae (Araceae) revealed by chloroplast *trnL-F* sequences. Am. J. Bot. 91:490–498.

Wen, J., R. Jansen, and K. Kilgore. 1996. Evolution of the eastern Asian and eastern North American disjunct genus *Symplocarpus* (Araceae): Insights from chloroplast DNA restriction site data. Biochem. Syst. Ecol. 24:735–747.

Wong, S. Y., P. C. Boyce, Sofiman, A. Othman, and C. P. Leaw. 2009. Molecular phylogeny of tribe Schismatoglottideae based on two plastid markers and recognition of a new tribe, Philonotieae, from the neotropic. Taxon in press.

1.3 The *cox1* Intron in Flowering Plants

Background and Previous Analyses

Group I introns are spread through the whole organismal diversity. They are found in bacteria and bacteriophages as well as in eukaryotes, and here in a wide variety of protists, fungi and plants. They have not yet been found in Archaea and only in two species of Metazoa, the sea anemone *Metridium senile* (Actinaria; Beagley et al., 1996) and the sponge *Tetilla* (Spirophorida; Rot et al., 2006). Group I introns are not restricted to a specific gene or cell compartment, but instead found in many different genes of bacterial, mitochondrial and plastidal genomes and the ribosomal RNA of nuclear genomes. They are lacking in nuclear genes. The same type of intron can be present in different positions within a gene (Dujon, 1989; Haugen et al., 2005). Group I introns, as well as group II introns, are ribozymes, i.e., self-splicing elements that auto-catalyze their excision from the RNA-transcript so that they do not influence the functionality of the genes they are inserted themselves into. The two types of introns are distinguished by their different secondary structure and the resulting different splicing mechanisms: Group I introns require an external guanosin as cofactor, in group II introns an adenine residue within the intron acts as nucleophile (Bhattacharya, 1998; Kelchner, 2002).

A striking feature of most group I introns (as well as of group II introns) is their mobility. They are able to insert themselves at the DNA level into the homologous position of an intronless allele by a process called gene conversion or "homing" (Dujon, 1989). It was first described from genetic crosses of fungi (Jacquier and Dujon, 1985), green algae (*Chlamydomonas*, Lemieux and Lee, 1987), T-even phages (Bell-Pedersen et al., 1989; Quirk et al., 1989) and the slime mold *Physarum polycephalum* (Muscarella and Vogt, 1989). These mobile genetic elements contain an open reading frame (ORF) encoding a site-specific endonuclease that creates a double-staggered strand-break at the highly specific target site of the intronless (intron$^-$) gene. The break is then repaired by the double-strand-break-repair pathway, forming a Holliday junction, which is resolved by using the intron containing (intron$^+$) allele as template (Fig. 1.5). Because of the nucleolytic degradation of the cleaved recipient and branch migration, coconversion of exon sequences flanking the intron is common, whereat the exon stretch 3' to the intron is copied from

the donor (Bell-Pedersen et al., 1989; Muscarella and Vogt, 1989; Wenzlau et al., 1989; Belfort and Perlman, 1995, Fig. 1.5).

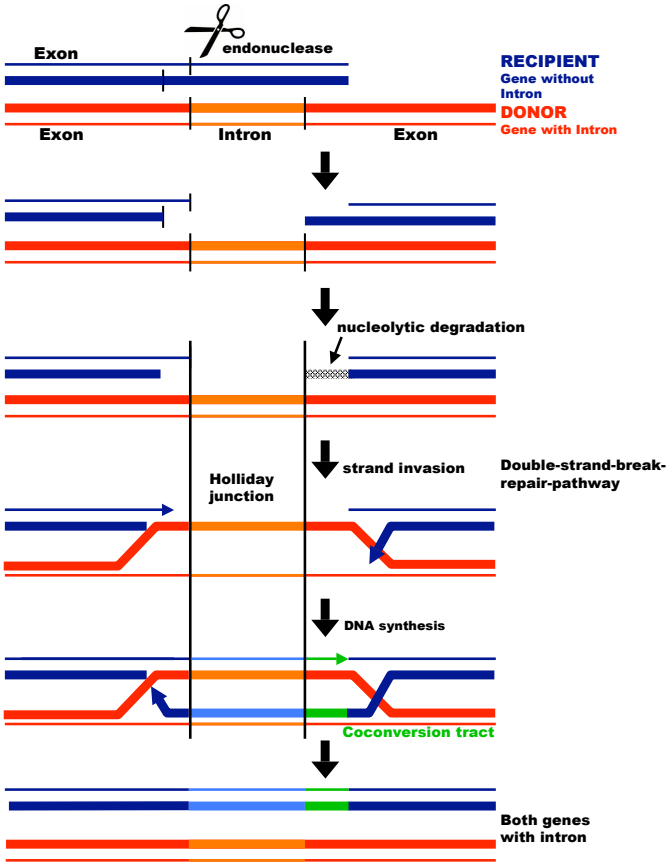


Figure 1.5: Scheme for the insertion of an intron from an intron-containing donor sequence by the double-strand-break-repair pathway (see text).

Another mechanism is the so-called intron transposition. This refers to the insertion of an intron deriving from a different site of the same gene or another gene (Dujon, 1989). Although this phenomenon has not yet been demonstrated for group I introns in its entirety, it is a plausible explanation for the similarity of introns found in different genetic locations (Belfort and Perlman, 1995).

The protein encoded by the ORF of the intron sometimes shows a maturase function in addition to the endonuclease function that is important for the correct folding and splicing of the transcribed intron (Delahodde et al., 1989; Wenzlau

et al., 1989; Haugen et al., 2005). The two functions of the proteins encoded by the ORF have complementary ways of acting: insertion (transposase/endonuclease) and splicing of the intron (maturase; Delahodde et al., 1989). It is assumed that the maturase function is newly acquired after an intron has been inserted into a genome (Belfort, 2003; Haugen et al., 2005). Despite the above-described wide distribution of introns, they are not thought to be ancient relicts in the genomes. On the contrary, studies point to let assume a more recent origin for many of them. Evidence supporting this idea is: (1) the high similarity of intron sequences in organisms belonging to distantly related taxa, whereas the respective exon sequences do not show high similarity (Lang, 1984; Quirk et al., 1989; Nishida and Sugiyama, 1995; Vaughn et al., 1995; Adams et al., 1998; Cho et al., 1998; Cho and Palmer, 1999); (2) the similarity between introns in genes in different genome compartments (Lonergan and Gray, 1994; Turmel et al., 1995; Haugen et al., 2005); and (3) the homing and transposition ability of introns (see above). One possible explanation for those observations is horizontal gene transfer, i.e., a gene transfer between non-related organisms or different cell compartments.

Michel and Westhof (1990) distinguished 11 subgroups of group I intron based on comparative sequence analysis. Several of them can be found at different insertion sites of the cytochrome oxidase subunit 1 gene (*cox1*) in the mitochondrial genome of fungi and green algae. In higher plants, only one type of intron (IB1 type) has been found in the *cox1* gene. After the discovery of several group I introns in the *cox1* gene in the liverwort *Marchantia polymorpha* (Ohta et al., 1993), *Peperomia polybotrya* was the first angiosperm for which a group I intron in the *cox1* gene has been reported, in a study involving 25 angiosperm species (Vaughn et al., 1995; Adams et al., 1998). Cho et al. (1998) surveyed the presence or absence of the intron in the *cox1* gene in over 300 land plants representing the diversity of angiosperms. In 48 of the species they found the intron, all at the homologous site in the *cox1* gene, with a highly similar length and sequence, but distributed very disjunctly over the organismal phylogeny. The phylogeny resulting from just the intron sequences turned out to be incongruent with the organismal one: distantly related taxa grouped together such as *Maranta* and *Hydrocotyle* (Marantaceae/Apiaceae), *Heliotropium* and *Rhamnus* (Boraginaceae/Rhamnaceae). On the other hand, closely related taxa grouped in very distant positions such *Maranta* and *Hedychium* (Zingiberales), *Ilex* and *Hydro-*

cotyle (Aquifoliaceae/Apiaceae). These observations and also the differences in coconversion tracts in closely related taxa, led Cho et al. (1998) to infer at least 32 independent gains of the intron. In addition, the lack of intron in early-diverged species in groups that otherwise include intron$^+$ taxa, supported their hypothesis. Cho and Palmer (1999) investigated the distribution of the *cox1* intron at a finer scale, focusing on the Araceae family. They screened 14 genera and found an intron in six of them (*Philodendron, Zamioculcas, Arisaema, Pistia, Amorphophallus, Xanthosoma*). From this they argued for at least three and more likely five independent horizontal gene transfers and no loss, again based on the distribution of the intron$^+$ taxa on the organismal phylogeny, the incongruent intron phylogeny, and the number of changes in the coconversion tracts. Only for one clade, *Arisaema triphyllum/Pistia stratiotes*, did they infer vertical transmission of the intron because of congruence between intron and organismal phylogenies and the identical coconversion tract. Cho and Palmer (1999) favored a multiple-gain-of-the-intron scenario over a multiple-loss scenario mainly for two reasons: Each cell contains hundreds of mitochondrial genomes. Genes loosing the intron will therefore have a low chance of fixing the intron$^-$ state because of the homing ability of all other intron copies still present. Another argument is that genes that lost their intron should retain the altered coconversion tract like a footprint (Cho et al., 1998). Palmer et al. (2000, S. 6965)'s final conclusion was, "Given that we have still sampled only a tiny fraction of the >300,000 species of angiosperms, we are confident that the intron has been horizontally acquired at least hundreds of times during angiosperm evolution and probably over 1,000 times. Equally remarkably, all of these transfers seem to have occurred very recently, in the last 10 million years or so of angiosperm evolution." As possible donor they proposed either a set of closely related fungi or one initial fungus-to-plant transfer followed by numerous plant-to-plant transfers.

Main Conclusions of the Present Work (Chapter 2)

My paper on the *cox1* intron (Cusimano et al., 2008, Chapter 2) tested the hypothesis put forward by Cho and Palmer (1999) of vertical inheritance of the *cox1* intron in the clade including *Arisaema triphyllum/Pistia stratiotes*, our *Pistia* clade, and also in general that of horizontal gene transfer of the *cox1* intron in

the angiosperms.

I first analyzed all available angioperm *cox1* sequences, resulting in a sampling of 179 taxa, of which 110 are intron$^+$, and 69 intron$^-$ (Chapter 2, Fig. 3). The resulting maximum likelihood phylogeny of angiosperm *cox1* introns, including those of Araceae, revealed many natural grouping of these taxa up to the family level (Chapter 2, Fig. 4). The few incongurences between the intron phylogeny and the angiosperm tree are best explained by low signal in the intron sequences (sequence similarity among the 110 introns ranges from 91% to identical) and long-branch attraction among a few taxa with high mitochondrial substitution rates.

Analysis of the 179 coconversion tracts reveals 20 types of tracts (11 of them only found in single species, all involving silent substitutions; Chapter 2, Figs. 2, S1). The distribution of these tracts on the angiosperm phylogeny showed that most of the early-diverging taxa lack the intron. Most of the intron$^-$ species throughout the whole phylogeny have a 0 coconversion tract type. The most common coconversion tract type among intron$^+$ angiosperms is what we call the 6+T(+A) type (see Chapter 2 for an explanation of this annotation). Another common tract type is 4+T. The intron phylogeny revealed two well-supported clades, each having one of the two most common tract types (Chapter 2, Fig. 4). There are several derivative tract types arising from gradual back mutation of the coconverted nucleotides, a pattern found in several clades, indicating that the coconversion tracts are not as stable as supposed by Palmer and colleagues.

Next we addressed the hypothesis of vertical transmission of the *cox1* intron in the *Pistia* clade. To assess the time of the intron gain (or loss) in angiosperms, we conducted molecular clock dating of the *Pistia* clade tree and screened it for intron presence/absence and coconversion tract type (Chapter 2, Fig. 1). The intron was present in all but two species (*Protarum sechellarum, Croat and Dzu 77954*) of the *Pistia* clade, all having the same coconversion tract (6+T), except for the two intron lacking species (0). According to the argumentation of Cho et al. (1998) and Cho and Palmer (1999), these data point to vertical transmission of the intron. Consequently, this case represents the first clear loss of the *cox1* intron, along with loss of the coconversion tract pattern. In addition, we showed that the intron must have persisted for at least 70 Myr in this clade contrary to Cho and Palmer (1999)'s claim that the *cox1* intron in angiosperms is young.

We also showed that the number of coconversion tract types found within an

angiosperm order correlates with the number of species analyzed in that order (Chapter 2, Fig. S3). This points to the need for of finer scale analyses with complete taxon sampling.

Lastly we addressed Cho and Palmer's hypothesis of fungi as donors for the *cox1* intron in plants. While we found high sequence similarity among the 110 angiosperm introns, all known putative homologs from fungi are highly different, so that none of them can be considered as possible donor (Chapter 2, Fig. S2). However, sampling in fungi is exceedingly sparse.

Together, these results suggest that the *cox1* intron entered angiosperms once or only a few times, has largely or entirely been transmitted vertically, and has been lost numerous times, with coconversion tract footprints providing unreliable signal of former intron presence. For the first time, the intron distribution pattern was investigated in a densely sampled clade, and time was added to the observed patterns.

Subsequent Analyses

In August 2008 a new paper addressing the *cox1* intron in plants was published by Palmer and colleagues (Sanchez-Puerta et al., 2008). Sampling was enlarged to 640 angiosperms, of which 129 possess the intron. Horizontal transfer is again put forward as the explanation for the apparent discrepancies between the intron and angiosperm phylogeny and the pattern of intron presence on the angiosperm tree. Coconversion tracts are again interpreted as reliable footprints of former intron presence. Horizontal transfers occur mainly within families. *Cox1* intron loss is reported from the genus *Plantago* based on the presence of intron$^+$ as well as intron$^-$ species having the same coconversion tract except for loss of RNA editing in the intron$^+$ species, indicating a loss by retroprocessing.

Sanchez-Puerta et al. (2008) hypothesize that 8 of 70 inferred HGT were accompanied by reduction in length of the coconversion tract, meaning that coconversion tracts do not necessarily look like the one of the donor, clearly reducing the importance of the tracts as footprints of horizontal transfer.

Surprisingly, (Sanchez-Puerta et al., 2008, p. 1773) again suggest, "Absent substantially more comprehensive taxon sampling and incorporation of divergence time estimates, we are unable to estimate with confidence the direction and abso-

lute timing of cox1 intron transfers. However, with respect to taxonomic rank, all cox1 intron transfers are relatively recent events."

Overall, the weight of the evidence clearly points to one or few early gains of the *cox1* intron followed by multiple losses being a more parsimonious explanation than hundreds or thousands of horizontal transfers. Functionality of the *cox1* intron endonuclease and therewith its homing ability has not yet been proven experimentally. The procedure of intron loss seems much more likely when one assumes a non-functionality of the endonuclease. Loss may actually be under positive selection because possessing the intron presents costs for the organism, and introns can easily be removed just by reverse transcription or genomic deletion. By contrast, the gain of the intron involves a highly complex procedure, requiring an isolated piece of the DNA fragment, a vector, a recipient, and a mechanism introducing this piece of DNA into a mitochondrion. The question how this new state then spreads to all other mitochondria is difficult to answer, also for the intron loss scenario, as no investigated angiosperm is heteromorphic for this character. A loss as well as a gain must therefore be an absolute event, changing the state in hundreds or thousands of mitochondrial genomes because there exists no stage in a plant's life cycle in which mitochondrion number is reduced to one. Furthermore, Goremykin et al. (2009) point in their study of mitochondrial DNA of *Vitis vinifera* to long-branch attraction, wrong model selection, limited sequence variation of plant mitochondrial genes and the analysis of paralogues as likely reasons for unexpected phylogentic groupings.

Last not least, the only event demonstrated with confidence is the loss of the intron (Cusimano et al., 2008; Sanchez-Puerta et al., 2008), with the two losses having happened in different ways.

References

Adams, K. L., M. J. Clements, and J. C. Vaughn. 1998. The *Peperomia* mitochondrial *coxI* group I intron: Timing of horizontal transfer and subsequent evolution of the intron. J. Mol. Evol. 46:689–696.

Beagley, C. T., N. A. Okada, and D. R. Wolstenholme. 1996. Two mitochondrial group I introns in a metazoan, the sea anemone *Metridium senile*: one intron contains genes for subunits 1 and 3 of NADH dehydrogenase. Proc. Natl. Acad. Sci. USA 93:5619–5623.

Belfort, M. 2003. Two for the price of one: a bifunctional intron-encoded DNA endonuclease-RNA maturase. Genes Dev. 17:2860–2863.

Belfort, M. and P. Perlman. 1995. Mechanisms of intron mobility. J. Biol. Chem. 270:30237–30240.

Bell-Pedersen, D., S. M. Quirk, M. Aubrey, and M. Belfort. 1989. A site-speciic endonuclease and co-conversion of flanking axons associated with the mobile *td* intron of phage T4. Gene 82:119–126.

Bhattacharya, D. 1998. The origin and evolution of Protists group I introns. Protist 149:113–122.

Cho, Y. and J. D. Palmer. 1999. Multiple acquisitions via horizontal transfer of a group I intron in the mitochondrial *cox1* gene during evolution of the Araceae family. Mol. Biol. Evol. 16:1155–1165.

Cho, Y., Y. L. Qiu, P. Kuhlman, and J. D. Palmer. 1998. Explosive invasion of plant mitochondria by a group I intron. Proc. R. Soc. Lond. B Biol. Sci. 95:14244–14249.

Cusimano, N., L.-B. Zhang, and S. S. Renner. 2008. Reevaluation of the *cox1* group I intron in Araceae and angiosperms indicates a history dominated by loss rather than horizontal transfer. Mol. Biol. Evol. 25:265–276.

Delahodde, A., V. Goguel, A. M. Becam, F. Creusot, J. Perea, J. Banroques, and C. Jacq. 1989. Site-specific DNA endonuclease and RNA maturase activities of two homologous intron-encoded proteins from yeast mitochondria. Cell 56:431–441.

Dujon, B. 1989. Group I intron as mobile genetic elements: facts and mechanistic speculations—a review. Gene 82:91–&.

Goremykin, V. V., F. Salamini, R. Velasco, and R. Viola. 2009. Mitochondrial DNA of Vitis vinifera and the Issue of Rampant Horizontal Gene Transfer. Mol. Biol. Evol. 26:99–110.

Haugen, P., D. M. Simon, and D. Bhattacharya. 2005. The natural history of group I introns. Trends Genet. 21:111–119.

Jacquier, A. and B. Dujon. 1985. An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. Cell 41:383–394.

Kelchner, S. A. 2002. Group II introns as phylogenetic tools: structure, function, and evolutionary constraints. Am. J. Bot. 89:1651–1669.

Lang, B. F. 1984. The mitochondrial genome of the fission yeast *Schizosaccharomyces pombe*: highly homologous introns are inserted at the same position of the otherwise less conserved *coxI* genes in *Schizosaccharomyces pombe* and *Aspergillus nidulans*. Embo J. 3:2129–2136.

Lemieux, C. and R. W. Lee. 1987. Nonreciprocal recombination between alleles of the chloroplast 23S rRNA gene in interspecific *Chlamydomonas crosses*. Proc. Natl. Acad. Sci. USA 84:4166–4170.

Lonergan, K. M. and M. W. Gray. 1994. The ribosomal-RNA gene region in *Acanthamoeba castellanii* mitochondrial DNA – A case of evolutionary transfer of introns between mitochondria and plastids? J. Mol. Biol. 239:476–499.

Michel, F. and E. Westhof. 1990. Modelling of the three-dimensional architecture of group in catalytic introns based on comparative sequence analysis. J. Mol. Biol. 216:585–610.

Muscarella, D. E. and V. M. Vogt. 1989. A Mobile Group I Intron in the Nuclear rDNA of *Physarum polycephalum*. Cell 56:443–454.

Nishida, H. and J. Sugiyama. 1995. A Common Group I Intron between a Plant Parasitic Fungus and Its Host. Mol. Biol. Evol. 12:883–886.

Ohta, E., K. Oda, K. Yamoto, Y. Nakamura, M. Takemura, N. Nozato, K. Akashi, K. Ohyama, and F. Michel. 1993. Group I introns in the liverwort mitochondrial genome: the gene coding for subunit 1 of cytochrome oxidase shares five intron positions with its fungal counterparts. Nucleic Acids Res. 21:1297–1305.

Palmer, J. D., K. L. Adams, Y. R. Cho, C. L. Parkinson, Y. L. Qiu, and K. M. Song. 2000. Dynamic evolution of plant mitochondrial genomes: Mobile genes and introns and highly variable mutation rates. Proc. Natl. Acad. Sci. USA 97:6960–6966.

Quirk, S. M., D. Bell-Pedersen, and M. Belfort. 1989. Intron Mobility in T-Even Phages: High Fequency Inheritance of Group I Introns Promoted by Intron Open Reading Frames. Cell 56:455–465.

Rot, C., I. Goldfarb, M. Ilan, and D. Huchon. 2006. Putative cross-kingdom horizontal gene transfer in sponge (Porifera) mitochondria. BMC Evol. Biol. 6.

Sanchez-Puerta, M. V., Y. Cho, J. P. Mower, A. J. Alverson, and J. D. Palmer. 2008. Frequent, phylogenetically local horizontal transfer of the cox1 group I Intron in flowering plant mitochondria. Mol. Biol. Evol. 25:1762–1777.

Turmel, M., V. Cote, C. Otis, J. Mercier, M. Gray, K. Lonergan, and C. Lemieux. 1995. Evolutionary transfer of ORF-containing group I introns between different subcellular compartments (chloroplast and mitochondrion). Mol. Biol. Evol. 12:533–545.

Vaughn, J. C., M. T. Mason, G. L. Sperwhitis, P. Kuhlman, and J. D. Palmer. 1995. Fungal origin by horizontal transfer of a plant mitochondrial group I intron in the chimeric *coxI* gene of *Peperomia*. J. Mol. Evol. 41:563–572.

Wenzlau, J. M., R. J. Saldanha, R. A. Butow, and P. S. Perlman. 1989. A latent intron-encoded maturase is also an endonuclease needed for intron mobility. Cell 56:421–420.

1.4 Endosperm Development in the *Pistia* Clade

The Araceae are relatively well known in overall morphology (see Chapter 1.1). However, embryological data are sparse (Grayum 1991), and most originate from investigations made half a century ago, with the microscopic and staining approaches then available. In addition, investigations have concentrated on a few groups, with sampling not phylogenetically well spread out, because generic (and other, deeper) relationships in the family were poorly understood. Therefore, most of the existing data are rather crude pollen and ovary characters, while the ontogeny of microspores, megaspores, gametophytes, embryos, endosperm and surrounding tissues have rarely been documented. Investigation of these structures not only requires sophisticated embedding and microscopic techniques, but ideally also wild-collected material, because cultivated plants can show anomalous embryogenesis. Complete embryological data are available only for three species: *Peltandra virginica* Goldberg (1941), *Synandrospadix vermitoxicus* Cocucci (1966) and *Theriophonum minutum* Parameswaran (1959). The scarse existing embryological data nevertheless show that Araceae are closely related to the Alismatiflorae and indicate that the genus *Acorus* is only distantly related to Araceae (Grayum, 1987). Whether embryological data also contain phylogenetic signal at lower taxonomic levels in Araceae is unclear because the current data set is not adequate to answer this question.

One of the most controversial arguments in embryology is that of early endosperm development. Generally, three types of endosperm development have been distinguished: an *ab initio* cellular development, a free-nuclear development, and a helobial endosperm development. The latter is poorly defined; (Grayum, 1991) described helobial endosperm as follows: "unequal division of the primary endosperm nucleus results in a smaller chalazal chamber that becomes haustorial, and cell division in the micropylar chamber is ab initio free- nuclear." The fate of the chalazal domain is not clearly defined.

Friedman and co-workers have analyzed endosperm development and double fertilization in numerous studies, focusing especially on *Ginkgo, Ephedra* (Gnetales), basal angiosperms, basal monocots, and basal eudicots. Their work on endosperm development in basal angiosperms (Floyd and Friedman, 2000) revealed six different patterns of early endosperm development (one of free- nuclear

Table 1.1: Available data for early endosperm development in the seven investigated species of the *Pistia* clade. For references see text.

Species	**1st division**		**Chalazal domain**		**Micropylar**	
	Type	**Modus**	**Division**	**Ploidie**	**domain**	**Development**
Pistia stratiotes	cellular	?	none	hypertrophied	cellular	*ab initio* cellular
Ariopsis peltata	cellular	micropylar	none	hypertrophied	nuclear (first 3)	helobial
Arisaema triphyllum	nuclear	?	nuclear	hypertrophied?	nuclear	free nuclear
Arisaema wallichianum	cellular	?	none	hypertrophied	cellular	*ab initio* cellular
Arum maculatum	cellular	?	none	2457n	cellular	*ab initio* cellular
Theriophonum minutum	cellular	chalazal	none	hypertrophied	cellular	*ab initio* cellular
Typhonium trilobatum	cellular	micropylar	none	hypertrophied?	cellular	*ab initio* cellular

development, two of ab initio cellular development, and three of helobial development). These patterns result from different combinations of the modi of the first cell division (free-nuclear, ab initio cellular) and the following development of the micropylar domain (free-nuclear, ab initio cellular), and the chalazal domain (free-nuclear, ab initio cellular, none). In the *Pistia* clade all six patterns can be observed (Table 1.1). In this clade, embryological data are available for *Ariopsis peltata* (Govinde Gowda, 1980), *Arisaema triphyllum* (Gow, 1908; Pickett, 1915), *Arisaema wallichianum* (Maheshwari and Khanna, 1956), *Arum maculatum* (Erbrich, 1965; Jacobson-Paley, 1920), *Pistia stratiotes* (Shadowsky, 1931), *Typhonium trilobatum* (Banerji, 1947), and *Theriophonum minutum* (Parameswaran, 1959). Table 1.2 summarizes the up to date known embryological features of these species. No embryologic studies have focused on *Biarum*, *Eminium*, *Dracunculus*, and *Helicodiceros* (which together include 32 species).

Table 1.2: Embryological characters of species in the *Pistia* clade. Shaded: Areae; lightgrey: characters with advanced states in the *Pistia* clade.

	Perianth	Floral Sexuality	Pollen nuclei	Pollen starch	Pollen aperture	Ovule orientation	Endosperm	Nucellus type	Nucellar cap	Endothelium	Megaspore tetrad type	Position of functional megaspore	Embryo sac type	Embryogeny	Number of cells in chalazal chamber	Microspre tetrad type	Tapetum type	Pollen mother-cell cytokinesis	Pollen tube path
Ariopsis peltata	no	monoecious	3	-	inaperturate	orthotropous	?	ten.	+	?	?	?	*Polygonum*- Type	?	1	?	periplasmodial	?	?
Arisaema triphyllum			2	+			starch	ten.?	-	+ ?	iso-bilateral	-		onagrad, solanad	1	tetra-hedral		succs.	poro-gamous
Arisaema wallichianum			2	+			?	ten.	+	+	linear, T-shaped	chalazal		?	1 ?	?		succs.	?
Pinellia ternata			3	+			?	?	+	?	?	?		?	?	?		?	?
Pistia stratiotes			3	+			?	?	+	+	?	chalazal		asterad	?	?		?	poro-gamous
Arum maculatum			3	-			starch	ten.	+	+	linear	chalazal		onagrad	1	?		succs.	?
Theriophonum minutum			2	+			aleuron	ten.	+	+	linear	chalazal		onagrad	1	?		succs.	poro-gamous
Typhonium trilobatum			3	+			starch	ten.	+	+	linear	chalazal		onagrad	1	?	?	succs.	poro-gamous
Dracunculus			3	+			**No data available**												
Helicodiceros			?	?															
Sauromatum			3	+															
Eminium			?	+															
Biarum			3	+															

References

Banerji, I. 1947. Life history of *Typhonium trilobatum*. Proc. Nat. Inst. Sci. India, Pt. B 13:207–230.

Cocucci, A. E. 1966. Embriología de *Synandrospadix vermitoxicus*. Kurtziana 3:157–181.

Erbrich, P. 1965. Über Endopolyploidie und Kernstrukturen in Endospermhaustorien. Oesterr. Bot. Z. 112:197–262.

Floyd, S. K. and W. E. Friedman. 2000. Evolution of endosperm developmental patterns among basal flowering plants. Int. J. Plant Sci. 161:S57–S81.

Goldberg, B. 1941. Life history of *Peltandra virginica*. Bot. Gaz. (Crawfordsville 102:641–662.

Govinde Gowda, M. T. 1980. Endosperm in *Ariopsis peltata* Nimmo., Araceae. Curr. Sci. 49:601–603.

Gow, J. E. 1908. Embryology of *Arisaema triphyllum*. Bot. Gaz. 45:38–44.

Grayum, M. H. 1987. A summary of evidence and arguments supporting removal of *Acorus* from the Araceae. Taxon 36.

Grayum, M. H. 1991. Systematic embryology of the Araceae. Bot. Rev. 57:167–203.

Jacobson-Paley, R. 1920. Sur le haustorium e la formation de l'albumen dans *Arum maculatum*. Bull. Soc. Bot. Genève, sér. 2 12:87–92.

Maheshwari, P. and P. P. Khanna. 1956. The embryology of *Arisaema wallichianum* Hooker f. and the systematic position of Araceae. Phytomorphology 6:379–388.

Parameswaran, N. 1959. A contribution to the Embryology of *Theriophonum minutum*. Proc. Indian Acad. Sci. 50B:15–25.

Pickett, F. L. 1915. A contribution to our knowledge of *Arisaema triphyllum*. Mem. Torrey Bot. Club 16:1–65.

Shadowsky, A. E. 1931. Einige Angaben über die Embryogenie von *Pistia stratiotes* L. Ber. Deutsch. Bot. Ges 49:350–356.

1.5 Cytogenetics in Araceae and the Areae

Chromosome Numbers

Chromosome numbers of over 700 Araceae species representing all genera have been reported (Petersen, 1989, 1993; Bogner and Petersen, 2007). These are mainly counts of isolated species (for references see Petersen, 1989). Only a few studies have focused on the cytology of natural clades: *Anthurium* (Sheffer and Croat, 1983; Sheffer and Kamemoto, 1976) or *Arum* (Bedalov and Küpfer, 2006). In terms of chromosome morphology, it seems that larger chromosomes with more distally positioned centromeres are derived from smaller metacentric ones (Petersen 1993). But this conclusion has to be considered preliminary because of the very fragmentary sampling over the whole family. No analysis regarding chromosomes has been carried out in molecular-phylogenetic framework.

Larsen (1969) and Marchant (1973) argued for an Araceae base number of $x = 7$, while Petersen (1989) hypothesized a base number of $x = 14$, because $2n = 28$ is especially widespread in the family. If ancient polyploidization is as important in the angiosperms as it appears (Soltis and Soltis, 1999), perhaps $x = 7$ or an even lower number is the more likely hypothesis, with all other numbers derived through polyploidization and descending (chromosome number reduction), and ascending dysploid series (chromosome number increase).

Chromosome number changes are mostly seen in genera with a high species number, raising the question if this is a consequence of diversification or if changes in chromosome numbers contributed to rapid reproductive isolation, for example, of polyploid offspring from diploid parents.

When plotting the known chromosome numbers onto a phylogeny of the Areae and their outgroups, the following pattern becomes apparent (Fig. 1 on Poster in Fig. 1.6): Most Areae investigated to date have $x = 14$, and this is also true of most outgroups except for *Pinellia* and a few species of *Arisaema* ($x = 13$). Genera of Areae that do not have $x = 14$ are *Biarum*, *Sauromatum*, and *Theriophonum*. In *Biarum*, $x = 13$ seems to be the ancestral base chromosome number, but there

Figure 1.6 ***(facing page)*:** Poster presented at the meeting "The Origin and Evolution of Biota in Mediterranean Climate Zones - an integrative Vision" 2007 in Zurich.

Polyploidy, aneuploidy, and chromosome number evolution of *Arum* and its allies (Areae, Araceae)

Natalie Cusimano & Susanne S. Renner

Munich University (LMU), Germany, cusimano@lrz.uni-muenchen.de

Introduction

The Areae clade (Fig. 1) includes about 150 species in the genera *Arum, Biarum, Dracunculus, Eminium, Helicodiceros, Theriophonum* and *Typhonium*. They range from Australia over South East Asia to the Mediterranean region and have been introduced to (sub-) tropical Africa. Chromosome numbers are available for 40% of the species and range from $2n = 160$ to $2n = 10$, the lowest chromosome number found in Araceae. We collected data from the literature and analyzed them on a new chloroplast phylogeny that includes 89 Areae from all seven genera. We also estimated the ages of the clades.

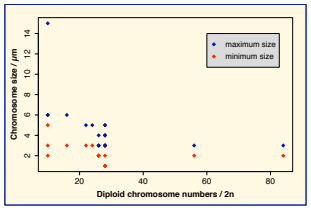


Fig. 2: Minimal and maximal chromosome sizes from chromosome counts of 19 Areae, plotted over their chromosome number. Chromosome sizes increase with decreasing chromosome number.

Discussion

The most common basic chromosome numbers in the Areae are $x = 13$ and $x = 14$. The frequency of polyploids points to numerous hybridization or autopolyploidization events. Aneuploidization and subsequent polyploidization events (or vice versa), or other complex chromosome rearrangements, seem to have played an important role only in the evolution of *Biarum* and the core *Typhonium*. In all Areae, $2n = 10$ is so far only found in the ca. 47 Ma old core *Typhonium*, which is embedded among taxa with higher chromosome numbers. This reduction of chromosome number and the negative correlation between chromosome number and size (Fig. 2) point to chromosome fusion and translocation events.

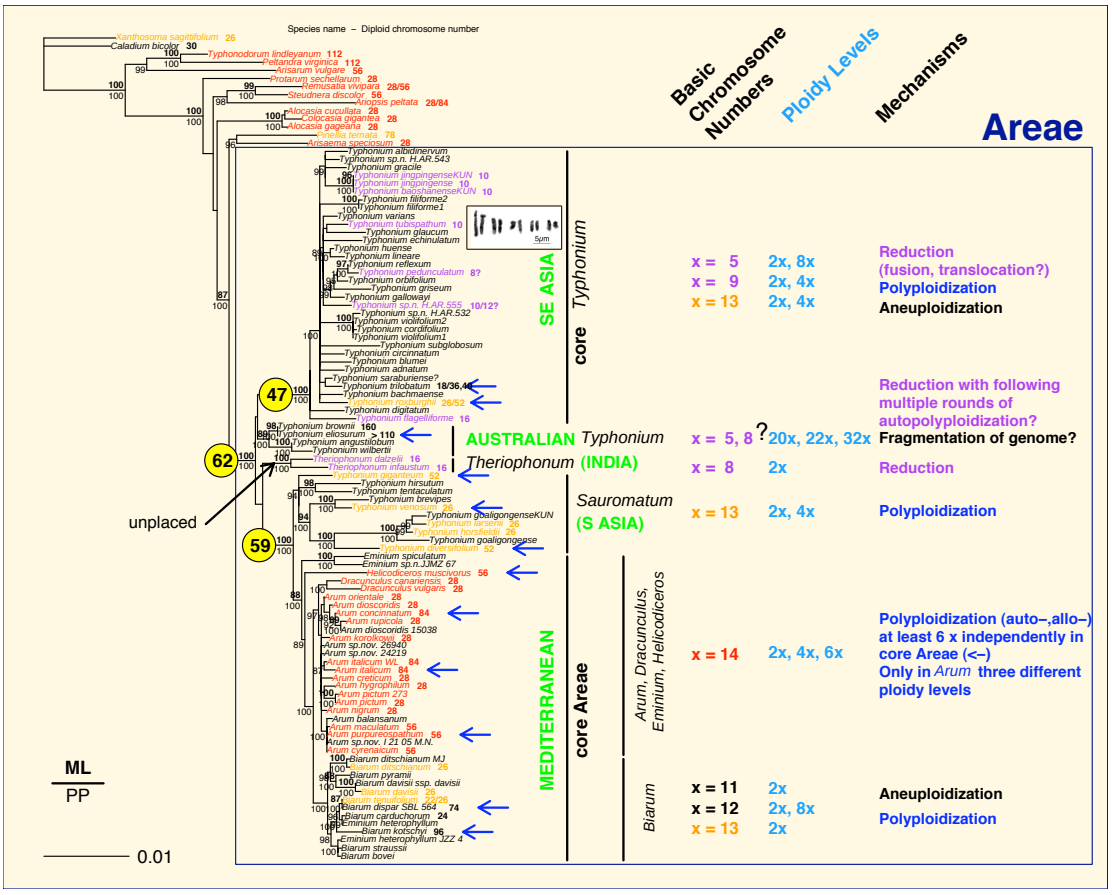


Fig. 1: Maximum likelihood phylogram of 89 Areae and 14 outgroup taxa with diploid chromosome numbers plotted next to species names. Support values at branches are maximum likelihood bootstrap percentages and posterior probabilities (ML/PP > 85 %). The karyogram is of *Typhonium tubispathum*, this is the third record of $2n = 10$ in Araceae. **Orange:** taxa with chromosome numbers based on $x = 13$; **Red:** taxa with chromosome numbers based on $x = 14$; **Purple:** hypothesized chromosome reduction events; **Blue:** hypothesized polyploidization events; Yellow circles: age in million years (SD in all cases < 8 Ma).

Phylogenetic Analyses

- Matrix (3321 bp, 335 bp parsimony informative) from 2 chloroplast genes: trnK and rpGene
- Best maximum likelihood tree search with RAxML (Stamatakis, 2006)

Maximum likelihood bootstrap analysis with RAxML (GTRCAT model, 1334 replicates)

Posterior probabilities calculation with MrBayes (GTR+I+Γ, 1,000,000 generations, sampled every 100th, temp=0.3, burnin=4750, 99 % credible set contains 10362 trees, Ronquist & Huelsenbeck, 2003)

- Dating analysis implemented in multidivtime (Bayesian dating method of Thorne & Kishino, 2002)

Literature

Ronquist, F. and J. P. Huelsenbeck. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

Thorne, J.L., Kishino, H., 2002. Divergence time estimation and rate evolution with multilocus data sets. Systematic Biology 51:689-702.

Stamatakis, A., 2006. RAxML-VI-HPC: Maximum Likelihood-based Phylogenetic Analyses with Thousands of Taxa and Mixed Models. Bioinformatics 22(21):2688-2690.

Table 1.3: Unpublished chromosome counts of *Typhonium* and *Sauromatum* species.

Species		**Chromosome No.** $2n =$
Typhonium	*saraburiense*?	18
	tubispathum	10
	pedunculatum	8, 10*
	sp. nov.(H.AR.555)	10, 12*
	filiforme	12 (+1B*)
	orbifolium/*violifolium*?	22
	varians	10
Sauromatum	*horsfieldii*	26
	hirsutum	26

are also species with $x = 12$ and $x = 11$, and one species, *B. dispar*, with $2n = 74$ ($n = x = 37$), indicating more complex changes than just multiplications of the entire genome. In *Sauromatum* ($x = 13$), as well as in *Arum* ($x = 14$), all chromosome number changes represent polyploidization events, leading up to a hexaploid level (*Arum italicum* with $2n = 84$). *Theriophonum* finally has a constant number of $2n = 16$. Apparently, speciation in these genera is associated with polyploidy and/or dysploidy (e.g., switches from $x = 14$ to $x = 13$ or perhaps vice versa), both of which may have occurred repeatedly.

While some Australian *Typhonium* species have $2n = 160$ (Briggs in Evans, 1962), one of the highest numbers reported for the family, others have $2n = 10$, the lowest chromosome number known in Araceae (Wang et al., 2002: *Typhonium jingpingense*; Zhi-Ling et al., 2007: *Typhonium baoshanense*). My own counts for additional *Typhonium* s.str. species (Table 1.3, Fig. 1.7) have revealed further cases of $2n = 10$ and other low numbers ($2n = 8$, 12, needing confirmation). Overall, *Typhonium* s.str. exhibits base chromosome numbers of $x = 4^*$, 5, 6^*, 8, 9, 11, and 13.

Genome Size

Genome size is an important cytological parameter, also referred to as C-value, where C stands for constant, as genome sizes have been found to be constant

*chromosome numbers in need of further confirmation

Table 1.4: Genome sizes of Areae species from Plant DNA C-values database; Chr.No. = Chromosome number; Est. Meth. = Estimation methods: Fe = Feulgen microdensitometry, FC: PI = flow cytometry using propidium iodide.

Species	**Chr.No.** $2n =$	**Est. Meth.**	**1C (pg)**	**Reference**
Biarum tenuifolium	22, 26	FC:PI	3,10	Zonneveld et al., 2005
Dracunculus canariensis	28	FC:PI	3,89	Zonneveld et al., 2005
Dracunculus vulgaris	28	Fe	6,83	Bennett, 1972
Arum maculatum	56	Fe	10,93	Bennett and Smith, 1976
Typhonium flagelliforme	16	Fe	5,17	Ghosh et al., 2001
Typhonium trilobatum	40	Fe	6,59	Ghosh et al., 2001

within species but highly variable between species. C-values in angiosperms span a huge range: from 0.10 pg in *Fragaria viridis* (Antonius and Ahokas, 1996), 0.16 pg in *Arabidopsis thaliana* (Bennett et al., 2003), and 63 Mbp in *Genlisea margaretae* (Greilhuber et al., 2006) to 127.4 pg in *Fritillaria assyriaca* (1pg DNA = 0,965x109 base pairs; Bennett and Smith, 1976), which represents a 1200-fold difference. Genome size does not directly correlate with organismal complexity (C-value paradox). The differences in genome size are mostly due to changes in the proportion of non-coding and repetitive DNA (both via downsizing and uploading). Physiological and ecological parameters, such as habitat, temperature, humidity, sexual system, pollination, or latitudinal range, may correlate with genome size at the intra- and interspecific level, but there is no widely accepted hypothesis explaining the observed striking differences. Large genomes may impose ecological and evolutionary constraints, for example, because cell division takes longer, and this might explain why genera with large genomes are underrepresented in extreme habitats (Knight et al., 2005; Leitch et al., 2007).

The Plant C-values Database of Royal Botanical Gardens Kew (`http://data.kew.org/cvalues/homepage.html`; Bennett and Leitch, 2005) includes 73 records for Araceae representing 30% of the family's genera and some 2% of its species. There are six reports for Areae (Table 1.4). The data do not suggest a correlation of genome size and chromosome number.

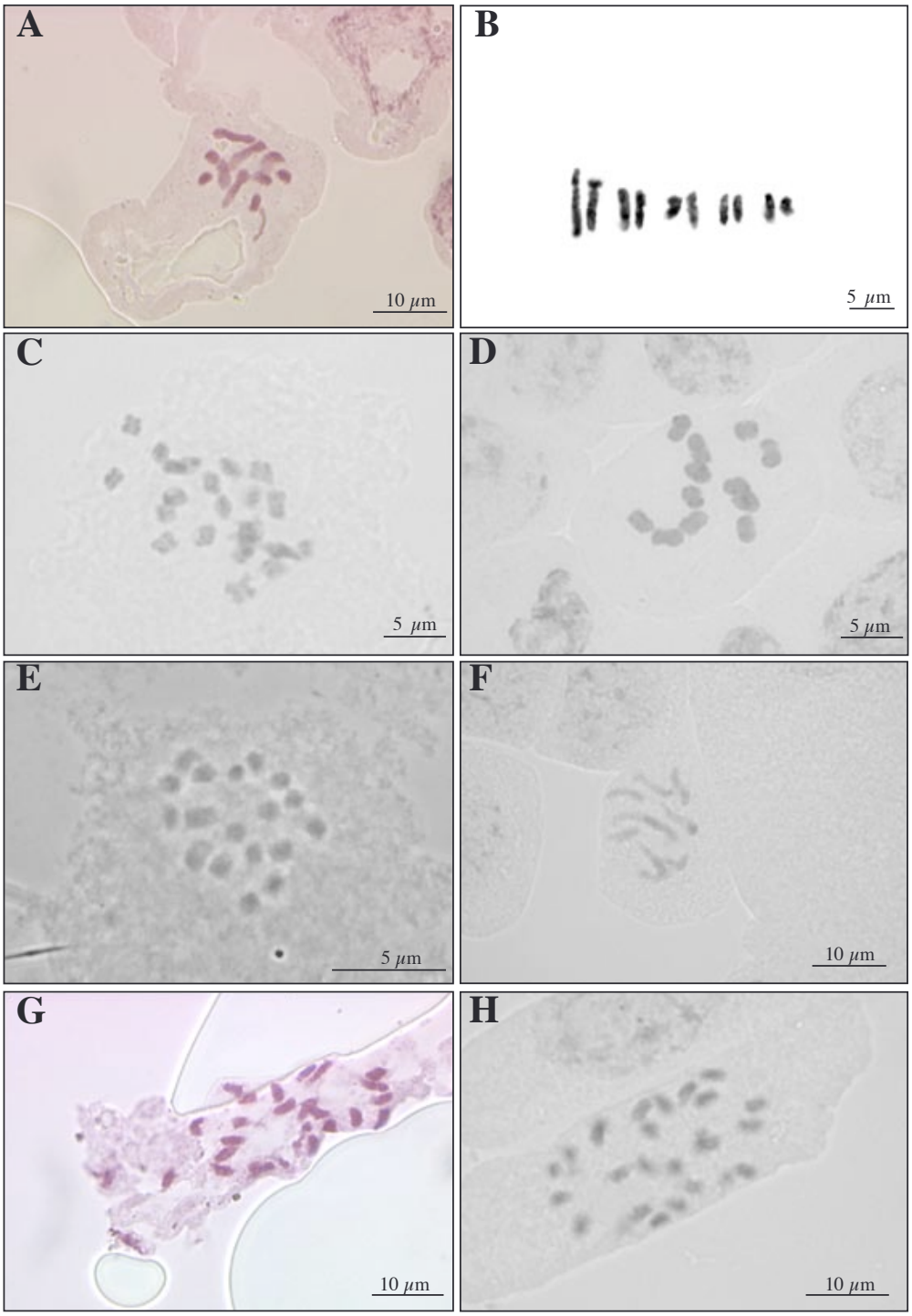


Figure 1.7: Chromosome spreads of **A, B:** *Typhonium tubispathum* $2n = 10$; **C:** *Typhonium orbifolium / violifolium* $2n = 22$; **D:** *Typhonium filiforme* $2n = 12$; **E:** *Typhonium saraburiense* $2n = 18$; **F:** *Typhonium varians* $2n = 10$; **G:** *Sauromatum horsfieldii* $2n = 26$; **H:** *Sauromatum hirsutum* $2n = 26$.

References

Antonius, K. and H. Ahokas. 1996. Flow cytometric determination of polyploidy level in spontaneous clones of strawberries. Hereditas 124:285.

Bedalov, M. and P. Küpfer. 2006. Studies on the genus *Arum* (Araceae). Aroideana 29:108–131.

Bennett, M. and I. Leitch. 2005. Nuclear DNA amounts in angiosperms: Progress, problems and prospects. Ann. Bot. 95:45–90.

Bennett, M., I. Leitch, H. Price, and J. Johnston. 2003. Comparisons with *Caenorhabditis* (similar to 100 Mb) and *Drosophila* (similar to 175 Mb) using flow cytometry show genome size in *Arabidopsis* to be similar to 157 Mb and thus similar to 25 % larger than the *Arabidopsis* genome initiative estimate of similar to 125 Mb. Ann. Bot. 91:547–557.

Bennett, M. and J. B. Smith. 1976. Nuclear DNA amounts in angiosperms. Philos. Trans. R. Soc. Lond. B Biol. Sci. 274:227–274.

Bogner, J. and G. Petersen. 2007. The chromsome numbers of the aroid genera. Aroideana 30:82–90.

Greilhuber, J., T. Borsch, K. Mueller, A. Worberg, S. Porembski, and W. Barthlott. 2006. Smallest angiosperm genomes found in Lentibulariaceae, with chromosomes of bacterial size. Plant Biol. 8:770–777.

Knight, C., N. Molinari, and D. Petrov. 2005. The large genome constraint hypothesis: Evolution, ecology and phenotype. Ann. Bot. 95:177–190.

Larsen, K. 1969. Cytology of vascular plants: III. A study of Aroids. Dan. Bot. Ark. 27:39–49.

Leitch, I. J., J. M. Beaulieu, K. Cheung, L. Hanson, M. A. Lysak, and M. F. Fay. 2007. Punctuated genome size evolution in Liliaceae. J. Evol. Biol. 20:2296–2308.

Marchant, C. J. 1973. Chromosome Variation in Araceae IV: From Acoreae to Lasieae. Kew Bull. 28:199–210.

Petersen, G. 1989. Cytology and systematics of Araceae. Nord. J. Bot. 9:119–166.

Petersen, G. 1993. Chromosome numbers of the genera of Araceae. Aroideana 6:37–46.

Sheffer, R. and T. Croat. 1983. Chromosome-numbers in the genus *Anthurium* .2. Am. J. Bot. 70:858–871.

Sheffer, R. and H. Kamemoto. 1976. Chromosome-numbers in the genus *Anthurium*. Am. J. Bot. 63:74–81.

Soltis, D. E. and P. S. Soltis. 1999. Isozyme evidence for ancient polyploidy in primitive angiosperms. Syst. Bot. 15:328–337.

Wang, Z., H. Li, and F. Bian. 2002. *Typhonium jinpingense*, a new species from Yunnan, China, with the lowest diploid chromosome number in Araceae. Novon 12:286–289.

Zhi-Ling, D., C. Shao-Tian, J. Yun-Heng, and L. Heng. 2007. *Typhonium baoshanense* Z. L. Dao & H. Li, a new species of Araceae from western Yunnan, China. Acta Phytotaxon. Sin. 45:234–238.

Chapter 2

Reevaluation of the *cox1* group I intron in Araceae and Angiosperms indicates a history dominated by loss rather than horizontal transfer

Cusimano, N., Zhang, L.-B. and S. S. Renner. 2008.
Molecular Biology and Evolution 25(2): 265-276.

2.1 Abstract

The origin and modes of transmission of introns remain matters of much debate. Previous studies of the group I intron in the angiosperm *cox1* gene inferred frequent angiosperm-to-angiosperm horizontal transmission of the intron from apparent incongruence between intron phylogenies and angiosperm phylogenies, patchy distribution of the intron among angiosperms, and differences between *cox1* exonic coconversion tracts (the first 22 nucleotides downstream of where the intron inserted). We analyzed the *cox1* gene in 179 angiosperms, 110 of them containing the intron (intron$^+$) and 69 lacking it (intron$^-$). Our taxon sampling in Araceae is especially dense to test hypotheses about vertical and horizontal intron transmission put forward by Cho and Palmer (1999, Multiple acquisitions via horizontal transfer of a group I intron in the mitochondrial *coxl* gene during evolution of the Araceae family. Mol. Biol. Evol. 16:1155-1165). Maximum likelihood trees of Araceae *cox1* introns, and also of all angiosperm *cox1* introns, are largely congruent with known phylogenetic relationships in these taxa. The exceptions can be explained by low signal in the intron and long-branch attraction among a few taxa with high mitochondrial substitution rates. Analysis of the 179 coconversion tracts reveals 20 types of tracts (11 of them only found in single species, all involving silent substitutions). The distribution of these tracts on the angiosperm phylogeny shows a common ancestral type, characterizing most intron$^+$ and some intron$^-$ angiosperms, and several derivative tract types arising from gradual back mutation of the coconverted nucleotides. Molecular clock dating of small intron$^+$ and intron$^-$ sister clades suggests that coconversion tracts have persisted for 70 million years in Araceae, whose *cox1* sequences evolve comparatively slowly. Sequence similarity among the 110 introns ranges from 91% to identical, while putative homologs from fungi are highly different, but sampling in fungi is still sparse. Together, these results suggest that the *cox1* intron entered angiosperms once, has since largely or entirely been transmitted vertically, and has been lost numerous times, with coconversion tract footprints providing unreliable signal of former intron presence.

2.2 Introduction

Most mitochondrial introns are self-splicing ribozymes that act as mobile genetic elements (Goddard and Burt, 1999; Chevalier and Stoddard, 2001; Haugen et al., 2005), with mobility depending on enzymes that the introns themselves encode. In group I introns, these enzymes belong to the LAGLI-DADG family and function as homing endonuclease, maturase, or both (Delahodde et al., 1989; Wenzlau et al., 1989; Belfort, 2003). Phylogenetic analyses suggest that group I introns in fungi, sponges, algae, and land plants have sometimes been transmitted horizontally (Lang, 1984; Lambowitz, 1989; Nishida and Sugiyama, 1995; Vaughn et al., 1995; Adams et al., 1998; Cho et al., 1998a,b; Cho and Palmer, 1999; Rot et al., 2006). A group I intron in the cytochrome c oxidase subunit I (*cox1*) gene is thought to have been transmitted horizontally as many as a 1,000 times among the 13,500 genera and 300,000 species of extant angiosperms (Cho et al., 1998b). This extrapolation was based on a survey of the *cox1* intron's distribution among 335 genera of land plants in which the authors inferred 32 separate cases of intron acquisition to account for the intron's presence in 48 of 281 species from 278 genera of flowering plants.

Horizontal transmission of introns has been inferred from three kinds of evidence (Cho et al., 1998b). First, strong incongruence between an intron phylogeny and that of angiosperms suggests independent gains rather than vertical transmission. Secondly, patchy distribution of an intron on an angiosperm phylogeny and especially the nesting of intron-containing (intron$^+$) species within large clades of intron-less (intron$^-$) species point to horizontal acquisition. Thirdly, coconversion tract analysis can provide information on whether a particular site gained or lost an intron (Bell-Pedersen et al., 1989; Adams et al., 1998; Cho and Palmer, 1999). Coconversion tracts are short stretches of flanking exon sequence that are converted to the donor DNA sequence. This is because group I introns transfer by way of a recombination/repair process initiated by a staggered double-strand break catalyzed by the intron's homing endonuclease at a target site in the recipient (Szostak et al., 1983; Lambowitz and Belfort, 1993; Belfort and Perlman, 1995). The cleaved DNA strands of the recipient DNA are partially degraded, creating a gap that is filled in using the donor DNA as the template. If the flanking exon stretches in the donor and recipient differ, then coconversion will create a

"footprint" that can stay even after the intron itself is lost again (Cho and Palmer, 1999).

Much trust has been placed in coconversion tracts as historical evidence of intron presence. For example, an investigation of the evolutionary history of the *cox1* intron in the Araceae (Cho and Palmer, 1999) relied on the exonic coconversion tracts in the intron$^+$ species, coupled with the absence of any deletion footprints in the intron$^-$ species, to infer three to five intron gains via horizontal transfers. Reliance on the coconversion tracts here overrode the implication of a parsimony reconstruction, which would have been consistent with a vertical transmission history in Araceae, with one gain, followed by two losses (Cho and Palmer, 1999). Cho and Palmer (1999) also noticed that *Arisaema triphyllum* and *Pistia stratiotes* had identical coconversion tracts and grouped together in the intron phylogeny, suggesting that these two introns might be vertically inherited.

To test the hypotheses of Cho and Palmer (1999), namely that the *cox1* intron has been transferred horizontally in much of the Araceae family, but vertically in the *Arisaema*/*Pistia* clade, we analyzed the *cox1* gene in a dense sample of relevant Araceae, using available multi-gene phylogenetic frameworks (Renner and Zhang, 2004, this study). Surprisingly, with the larger taxon sample employed here, the distribution of the *cox1* intron in Araceae is more parsimoniously explained by ancestral presence, followed by independent losses, than by horizontal gene transfers.

Araceae are an early-branching lineage of flowering plants, and we therefore decided to investigate the distribution of the *cox1* intron among early and more recent lineages of angiosperms based on all angiosperm *cox1* sequences available in GenBank (plus new sequences generated in the course of this study). Comparison of the much larger intron phylogeny with the angiosperm phylogeny, the great sequence similarity among angiosperm *cox1* introns, the clustered distribution of exonic coconversion tract types, and hierarchical patterns of decay in the coconversion tracts suggest ancestral presence of the *cox1* intron, followed by numerous losses. This implies that the signal in coconversion tracts (the "footprint") is less reliable than previously thought. To infer a temporal framework for *cox1* intron turnover and the loss of coconversion tracts, we estimated maximal times over which the intron could have been gained or lost in Araceae, using angiosperm clade ages as a proxy for intron maximal ages. We also evaluate the hypothesis of

Seif et al. (2005) that the *cox1* intron in angiosperms originated in a fungus close to *Rhizopus oryzae.*

2.3 Materials and Methods

Taxon Sampling and Sequencing

Taxa selected for this study with their sources and herbarium vouchers (where applicable) are listed in supplementary table S2.1, which includes 179 angiosperms, 110 of them intron$^+$ and 69 intron$^-$. To deduce the evolutionary history of the *cox1* intron in the *Arisaema*/*Pistia* clade, we relied on two chloroplast loci (the *trnL* intron and adjacent spacer before the *trnF* gene and the *rpl20-5-rps12* intergenic spacer) and one mitochondrial locus (parts of exons b and c of *nad1* and the complete intron between them). We included 30 species of Araceae, many available from Renner and Zhang (2004). Newly generated sequences were produced with the same primers and PCR conditions as used in that study. The *cox1* exon and intron (where present) were sequenced for 36 Araceae, and in all, this study includes 56 newly generated sequences (36 *cox1* genes and 20 of other loci).

Total DNAs of silica-dried material were extracted with the NucleoSpin plant kit according to the manufacturer's protocol (Macherey-Nagel, Düren, Germany), and the complete *cox1* gene was directly amplified with the primer pair 82F (5' GGAGTGATGGGCACAT GCTTCT 3') and *cox1*1.6KR int (5' AAGGCTG-GAGGGCTTTGT AC 3'). Polymerase chain reactions (PCR) were performed with 10 mM primers in 25-μl reactions using BioTherm DNA polymerase (Genecraft, Lüdinghausen, Germany). The initial step of 5 min at 95 °C was followed by 35 cycles of 95 °C for 30 s for DNA denaturation, 60 °C for 60 s for primer annealing, and 72 °C for 2 min and 40 s for primer extension. PCR products were controlled by electrophoresis on an ethidiumbromide-stained 1% agarose gel with a 1 Kb Plus DNA ladder (Invitrogen, Karlsruhe, Germany). The amplified fragment was ca. 2340 nt long for intron$^+$, and about 1500 nt long for intron$^-$ taxa. Products were purified and quantified electrophoretically using Lambda DNA as standard. If multiple bands were detected, an additional electrophoresis was performed to excise and analyze them separately. Sequencing relied on Big Dye Terminator kits (Applied Biosystems, Warrington, UK) and the following primers (in different

combinations depending on the length of the sequences obtained, varying from 400 to 1000 nt): 42F (5' GGATCTTCTCCA CTAACCACAAA 3'), 82F (see above), 657R (5' GCGGGATCAGAAAAGGTTGTA 3'), IP53 (5' GGAGCAGTTGATTTAGC 3'), I589R (5' GGTAGTCGATGCTTCATAGC 3'), I361F (5' GTATTAAAATGCGATCAGGTGC 3'), I557F (5' AGGATTCTTTGATGCTGAGGG 3'), I942R (5' GGATGAATAGAAGAAAGGT 3'), Int1.2KF (5' AGCATGGCTAGCTTTCCTAGA 3'), 855F (5' TGGATTTCTTGTTTGGGCTCAT 3'), IP56 (5' GAGCAATGTCTAGCC C 3'), 1150F (5' TCTATGGGAGCCGTTTTTGC 3') and *cox1*.6KR (see above). The cycle sequencing products were cleaned by Sephadex G-50 Superfine gel filtration (Amersham, Uppsala, Sweden) on MultiScreen TM-HV membrane plates (Millipore, Bedford, USA) according to the manufacturers' protocols to remove unincorporated nucleotides. Fragments were separated on an ABI 3100 Avant capillary sequencer, assembled and edited using the software Sequencher (Gene Codes, Ann Arbor, MI, USA), and BLAST-searched in GenBank.

Alignments and Phylogenetic Analyses

Alignments were generated manually in MacClade (Maddison and Maddison, 1992) and adjusted by eye; all have been submitted to TreeBase. Amplification of the *cox1* exon of *Theriophonum dalzellii* failed, and the missing sequence for this species was coded with question marks. We analyzed four data matrices. The first comprised the four chloroplast and mitochondrial loci sequenced for the *Arisaema*/*Pistia* clade. The second consisted of 149 angiosperm *cox1* exon sequences including 11 newly sequenced Araceae species and 12 Araceae from Cho et al. (1998b). The third consisted of 106 angiosperm *cox1* intron sequences of which 38 were Araceae. The fourth matrix comprised the coconversion tracts of 179 angiosperms, 110 of them intron$^+$ and 69 intron$^-$. To assess the phylogenetic signal in the *cox1* gene and introns, we used the molecular phylogeny of angiosperms published by Qiu et al. (2005), the angiosperm phylogenetics database of Stevens et al. (2001 onwards onwards, version 8, June 2007), and an unpublished phylogeny of Araceae provided by S. Mayo (Royal Botanic Gardens Kew, personal communication, Feb. 2007).

DNA indels or missing data in the *cox1* intron and exon were excluded from

phylogenetic analyses. Phylogenetic inference relied on maximum likelihood searches (ML) as implemented in RAxML-VI-HPC version 2.2.3 (Stamatakis, 2006). Bayesian analysis relied on MrBayes version 3.1.2 (Ronquist and Huelsenbeck, 2003). The best-fitting model for the combined chloroplast and mitochondrial data (4682 characters, excluding the intron and 22 bp coconversion tract of the *cox1* gene) identified by Modeltest version 3.7 (Posada and Crandall, 1998) was the GTR + I + Γ model whether by hierarchical likelihood ratio testing or the Akaike Information Criterion. We therefore used this model in Bayesian analyses, while maximum likelihood analyses relied on the GTR + Γ model, this being the only model implemented in RAxML. The best-fitting model for the *cox1* intron matrix was the TvM + I + Γ model (5 substitution types). As the number of substitution types in MrBayes can only be set to 1, 2 or 6, we used the GTR + I + Γ model. Bayesian runs were started from independent random starting trees and repeated at least twice. Markov chain Monte Carlo (mcmc) runs extended for 1 million generations, with trees sampled every 100 generations. 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 1.0) for the gamma-shape parameter and all branch lengths. Convergence was assessed in several ways: by checking that final likelihoods and majority rule topologies in different runs were similar; that the standard deviations (SD) of split frequencies were <0.01; that the log probabilities of the data given the parameter values fluctuated within narrow limits; that the convergence diagnostic (the potential scale reduction factor given by MrBayes) approached 1; and by examining the plot provided by MrBayes of the generation number versus the log probability of the data. Trees saved prior to convergence were discarded as burn-in (2000-5000 trees) and a consensus tree was constructed from the remaining trees. Bootstrapping under ML used 1000 replicates performed in RAxML, with the initial rearrangement settings and the number of categories tested following the manual. Resulting bootstrap values as well as Bayesian posterior probabilities were plotted on the ML tree using the APE package (Paradis et al., 2004) in R (R Developmental Core Team, 2006).

The *cox1* exon data were analyzed under parsimony in PAUP version 4.0b10 (Swofford, 2002). Searches were heuristic, using 100 random taxon addition replicates, tree-bisection-reconnection (TBR) swapping, with the 'multiple trees' and the 'steepest descent' options in effect. Starting trees were obtained by stepwise

addition; the trees in memory were limited to 100.

Coconversion Tract Analysis

Coconversion tracts, i.e., the first 22 nucleotides downstream of the intron insertion site, in 179 angiosperms were compared with the exonic tract of an intron$^-$ Araceae, namely *Orontium aquaticum*, following Cho and Palmer (1999). For convenience, the *O. aquaticum* tract type is henceforth referred to as the unaltered, or 0, tract type, without this implying that it is an ancestral condition. The remaining tracts were categorized relative to the *O. aquaticum* type according to the number of alterations present in their 3^{rd} to 18^{th} position (whether 1, 2, 3, ..., 6 differences, all in the 3^{rd} position, and all silent), presence or absence of a T in the 20^{th} position (silent: C > U RNA editing), and presence or absence of an A in the 21^{st} position (silent). A coconversion tract that comprises six nucleotide differences compared with *Orontium aquaticum*, a T in position 20 and an A in position 21 is thus referred to as "6+T+A". Similarly, "4+T" refers to a coconversion tract with 4 substitutions and an T in position 20.

Divergence Time Estimation

For divergence time estimation, we relied on the combined chloroplast and mitochondrial data (4682 nt for 30 taxa) and the Bayesian relaxed clock approach implemented in multidivtime (Thorne et al., 1998; Thorne and Kishino, 2002). After calculating substitution model parameters for the DNA data under the F84 + Γ model (with five rate categories) on the ML topology obtained from the combined data, rooted on *Xanthosoma sagittifolium* and *Caladium bicolor*, we used Thorne's estbranches program to estimate branch lengths and their variance, given the specified evolutionary tree and model parameters. The a priori expected number of time units between the root and the tips was set to 0.9, with a standard deviation of 0.5; the prior on the mean root rate was set to 0.0128, by dividing the median distance from the ingroup root to the tips by the time unit. Thorne's manual recommends that the prior for brown mean (and its standard deviation) be set at values that, when multiplied by the approximate time from the root to the present, yield a value between 1 and 2, and we therefore set brown-mean to

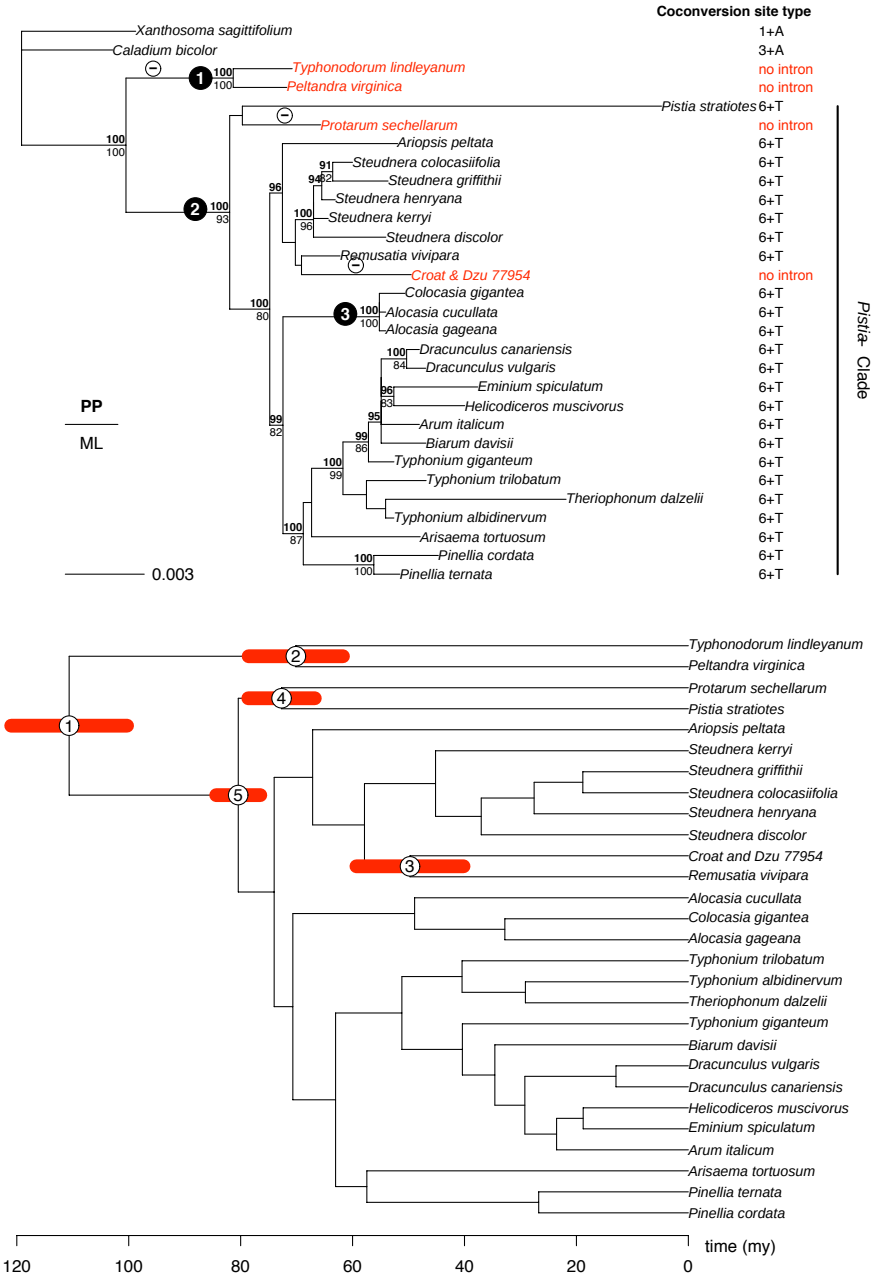


Figure 2.1: A. Maximum likelihood tree for the *Arisaema*/*Pistia* clade based on chloroplast and mitochondrial sequences (4360 bp) analyzed under a GTR + I + Γ model. Values above branches indicate a posterior probability >90, values below branches maximum likelihood bootstrap value >75. Numbered nodes (black) refer to the following minimal (Mi) or maximal (Ma) constraints based on fossils and a geological event: (1) Peltandreae, Mi = 60 MY; (2) Seychelles archipelago, Ma = 85 MY; (3) Colocasieae, Mi = 45 MY. Coconversion tract types (see Fig. 2.2A) of the respective taxa are given on the right. Three inferred intron loss events are marked by a circle-enclosed hyphen. **B.** Chronogram for the *Arisaema*/*Pistia* clade obtained under a Bayesian relaxed clock applied to the same data and constrained as shown in 2.1A. Nodes 1-5 are discussed in the text, the grey bars indicate standard deviations around estimates.

1.11. The Markov chain length was 1 million cycles, sampled every 100th cycle and with a burn-in of 100,000 cycles; analyses were repeated at least twice.

To obtain absolute times from genetic distances, we used the following constraints: (1) Peltandreae are first known from 60 million year (MY)-old leaves from Europe, Kazakhstan, North Dakota, and Tennessee (Wilde et al., 2005). This provides a minimal age of 60 MY for node 1 in Fig. 2.1A. (2) *Protarum sechellarum* is endemic to the Seychelles, and the age of this archipelago (Braithwaite, 1984) thus provides a maximal age of 85 MY for node 2. (3) Middle Eocene leaf impressions (*Caladiosoma messelense*; Wilde et al., 2005) that closely match modern Colocasieae provide a minimal age of 45 MY for node 3 in Fig. 2.1A. (4) The oldest fossils of Araceae are 110-120 MY old (Friis et al., 2004), and therefore 120 MY was used as a maximal age for the root node. The earliest angiosperms fossils are 141-132 MY old (Hughes, 1994).

2.4 Results

The *cox1* Intron and Exonic Coconversion Tracts in the *Arisaema/Pistia* Clade

The distribution of the *cox1* intron in the *Arisaema/Pistia* clade is shown in Fig. 2.1A (including the relevant outgroups). As predicted by the hypothesis of Cho and Palmer (1999) that the intron might be vertically inherited in this clade, most species are intron$^+$ and have the same coconversion tract. This tract, namely the 6+T type, comprises six nucleotide differences compared to intron$^-$ species and a T in position 20.

The *cox1* intron is lacking in *Typhonodorum lindleyanum* and *Peltandra virginica*, which form a clade, in *Protarum sechellarum*, and in the Vietnamese species *Croat and Dzu 77954*. For genera with more than one species, we checked at least one additional congeneric for the intron and the coconversion tract, and they all showed the same pattern, (figs. 2.3, 2.4). Of the intron$^-$ species, three have unaltered coconversion tracts, while *Croat and Dzu 77954* has an A, instead of a C, at position 21 of its tract (Fig. 2.1A), which may be circumstantial evidence that it once had an intron in its *cox1* gene (below). Of the outgroups, *C. bicolor* has the 3+A tract type, and *X. sagittifolium* and *X. mafaffa* the 1+A type (figs. 2.1A,

2.2A, and supplementary Fig. S1).

Cox1 Exonic Coconversion Tracts throughout Angiosperms

Analysis of all available angiosperm *cox1* sequences (GenBank, 1 March 2007), revealed 20 coconversion tract types of which 11 are only found in single species. Figure 2.2A summarizes the tract types of the 110 intron$^+$ and the 69 intron$^-$ *cox1* sequences. Overall, 112 tracts are of the +T type (101 in intron$^+$ *cox1* genes, 11 in intron$^-$ genes) and 19 are of the +A type (17 intron$^+$, 2 intron$^-$). In 15, the A occurs together with the T, while in four cases, all in the Araceae, the A occurs without the T (one of them in an intron$^-$ species and three in intron$^+$ species, see Fig. 2.2A and Fig. 2.3). Depending on the intron$^-$ angiosperm used for comparison, the C > T transversion in position 20 will be counted as part of the coconversion tract or not. Thus, comparison with *Zea mays* results in an apparent coconversion tract with six differences (Cho et al., 1998b), while comparison with *O. aquaticum* suggests a coconversion tract with seven differences (Cho and Palmer, 1999, this study).

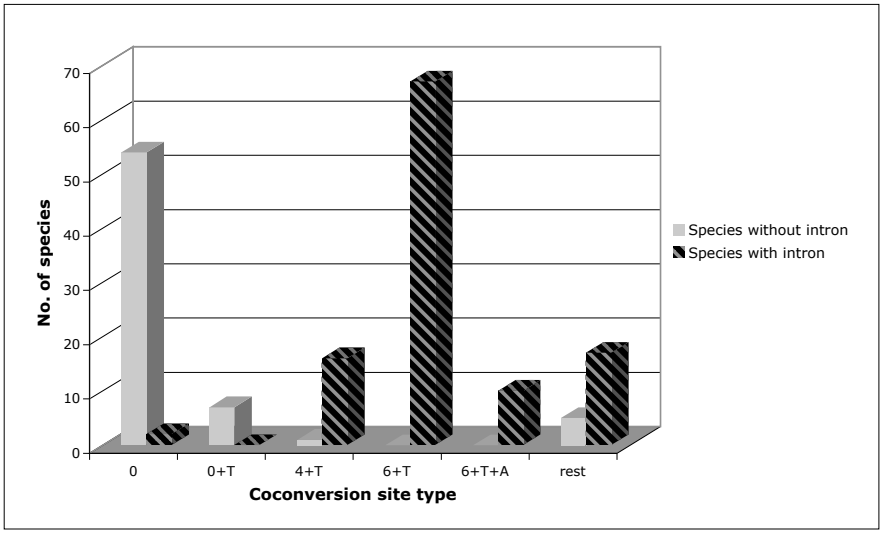
Of the intron$^+$ angiosperms, the majority (60%) has the 6+T tract type (34 of these are Araceae), while 9% (10 species) have the 6+T+A tract type (Fig. 2.2B). In other words, more than two thirds of the intron$^+$ angiosperms have all six third positions available in the coconversion tract changed (compared to intron$^-$ species). Five species (*Breynia nivosa, Hevea brasiliensis, Justicia americana, Pilea fontana, Plantago coronopus*) are of the X-types, which appear to have undergone a back mutation in the middle of a 6+T coconversion tract (Fig. 2.2A). Sixteen are of the 4+T type, four of the 3+T type, and the remaining ones belong to rare types, such as 1+A. Two of the 110 intron$^+$ angiosperms (*Coffea arabica, Rhamnus* sp.) exhibit no differences in their coconversion tract.

Of the intron$^-$ angiosperms, 78% have unaltered tracts like *O. aquaticum* (Fig. 2.2B) while 22% (15) show differences in their coconversion tract, i.e., have a deletion footprint. Of the footprints, one is of the 6+T type, one of the 4+T type, seven have only the T in tract position 20, and four have single substitutions in different positions. Two species (*Canella winterana, Crossosoma bigelovii*) have coconversion tracts that apparently underwent back mutations from 6+T and 6+T+A type tracts, and three are of the X* types (Fig. 2.2A), that is, have a G instead of an

Int	5' - Exon - 3' --> 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29	Coconversion site type	Number of species	% species without intron
Orontium aquaticum -	C A T C C A G A G G T G T A T A T T C C C A T T C T G C C	0	54	78,3
Euptelea polyandra -	C A T C C A G A A G T G T A T A T T C C C A T T C T G C C	X	1	1,4
Amborella trichopoda -	C A T C C A G A G G T T T A T A T T C C C A T T C T G C C	X	1	1,4
Zea mays -	C A T C C A G A G G T G T A T A T T C T C A T T C T G C C	+T	7	10,1
Croat & Dzu 77954 -	C A T C C A G A G G T G T A T A T T C C A A T T C T G C C	+A	1	1,4
Canella winterana -	C A T C C A G A G G T C T A T A T T C T C A T T C T G C C	X*+T	1	1,4
Crossosoma bigelovii -	C A T C C G A A G G T C T A T A T T C T A A T T C T G C C	X*+T+A	1	1,4
Plantago sericea -	C A C C C T G A A G T C T A T A T T C T C A T T C T G C C	X*+T	1	1,4
		Species without intron:	67	

Proteins	HIS	PRO	GLU	VAL	TYR	ILE	PRO			
										% species with intron
Rhamnus cathartica +	C A T C C A G A G G T G T A T A T T C C C A T T C T G C C							0	2	1,8
Barringtonia asiatica +	C A C C C A G A G G T G T A T A T T C C C A T T C T G C C							1	2	1,8
Xanthosoma mafaffa +	C A C C C A G A G G T G T A T A T T C C C A A T T C T G C C							1+A	2	1,8
Saranthe sp. +	C A C C C T G A G G T G T A T A T T C C C A T T C T G C C							2	1	0,9
Philodendron hederaceum +	C A C C C T G A A G T G T A T A T T C C C A T T C T G C C							3	2	1,8
Caladium bicolor +	C A C C C T G A A G T G T A T A T T C C C A A T T C T G C C							3+A	1	0,9
Plantago cynops +	C A C C C T G A A G T G T A T A T T C T C A T T C T G C C							3+T	4	3,6
Plantago coronopus +	C A C C C A G A A G T T T A T A T T C T C A T T C T G C C							X+T	1	0,9
Heliotropium arborescens +	C A C C C T G A A G T T T A T A T T C T C A T T C T G C C							4+T	16	14,5
Justicia americana +	C A T C C T G A A G T T T A C A T C C T A A T T C T G C C							X+T+A	1	0,9
Hevea brasiliensis +	C A C C C T G A A G T T T A C A T T C T A A T T C T G C C							5+T+A	3	2,7
Cucumis sativus +	C A C C C T G A A G T T T A C A T C C T A A T T C T G C C							6+T+A	10	9,1
Amorphophallus rivieri +	C A C C C T G A A G T T T A C A T C C T C A T T C T G C C							6+T	67	60,9
								Species with intron:	112	

A



B

Figure 2.2: Coconversion tract types of intron$^-$ and intron$^+$ angiosperm *cox1* exons. **A.** Top panel: Patterns of substitutions in the 22 nt downstream from the 3' end of the intron insertion site in 10 species selected to represent the tract types found in 69 intron$^-$ species. Bottom panel: Tract in 13 species selected to represent the tract types found in 110 intron$^+$ species. Most common patterns are given a specific name; the others are labeled as X. An asterisk following the X denotes substitutions in the third position that are different from the common nucleotides. Protein translation is given below the top panel. **B.** Histogram of the number of intron$^+$ and intron$^-$ possessing a certain type of coconversion tract. Only the five commonest types are explicitly shown (0, 0+T, 4+T, 6+T, 6+T+A).

A or a T in tract position 6 and/or a C instead of a G or a T in tract position 12. So far, *Plantago* is the only genus with both intron$^-$ and intron$^+$ species, and it also shows particularly variable coconversion tracts (Fig. 2.2A, supplementary Fig. S1).

Phylogenetic Analyses of Angiosperm *cox1* Exon and Intron Sequences

The hypothesis that the *cox1* intron in angiosperms was gained by multiple lateral transfers predicts incongruence between the angiosperm phylogeny and the intron phylogeny. By contrast, congruence among angiosperm and intron phylogenies points to vertical inheritance. We therefore performed phylogenetic analyses of *cox1* exons and introns and compared them with angiosperm relationships inferred from larger data sets (Stevens, 2001 onwards; Qiu et al., 2005).

A phylogeny based on *cox1* exon sequences (149 angiosperm species, 1288 characters, excluding the 22 nt of the coconversion tract), rooted on *Amborella*, is shown in Fig. 2.3. It recovers the monocots, core eudicos, and ordinal relationships in agreement with angiosperm phylogenies based on other data (Stevens, 2001 onwards; Qiu et al., 2005). Generic groupings within more densely sampled families such as Araceae (23 *cox1* sequences) agree with relationships obtained in larger data sets. For example, *Orontium* places as the first-diverging Araceae, followed by *Anthurium* and *Zamioculcas*. *Dieffenbachia* and *Zantedeschia*, as well as *Spathiphyllum* and *Scindapus* form sister pairs. The *Arisaema*/*Pistia* clade and its relatives also group together. All this fits with a molecular phylogeny of Araceae (S. Mayo, Royal Botanic Gardens Kew, personal communication, Feb. 2007). The earliest-diverging angiosperm lineages containing the *cox1* intron are the Magnoliales (*Asiminia, Knema, Myristica*) and Piperales (*Peperomia*).

A phylogeny based on *cox1* intron sequences (106 species, 967 characters, excluding 173 gapped positions) and rooted on Myristiaceae (Magnoliales) is shown in Fig. 2.4. Except for ten species that form a basal grade (*Erethia, Bursera, Lepionurus, Melia, Croton, Jasminum, Musella, Musa*), four large clades are apparent (labeled A-D in Fig. 2.4): Clade A includes all Araceae plus the two *Peperomia* species (probably reflecting long branch attraction). Within clade A, species with a 6+T coconversion tract cluster together. Clade B includes the re-

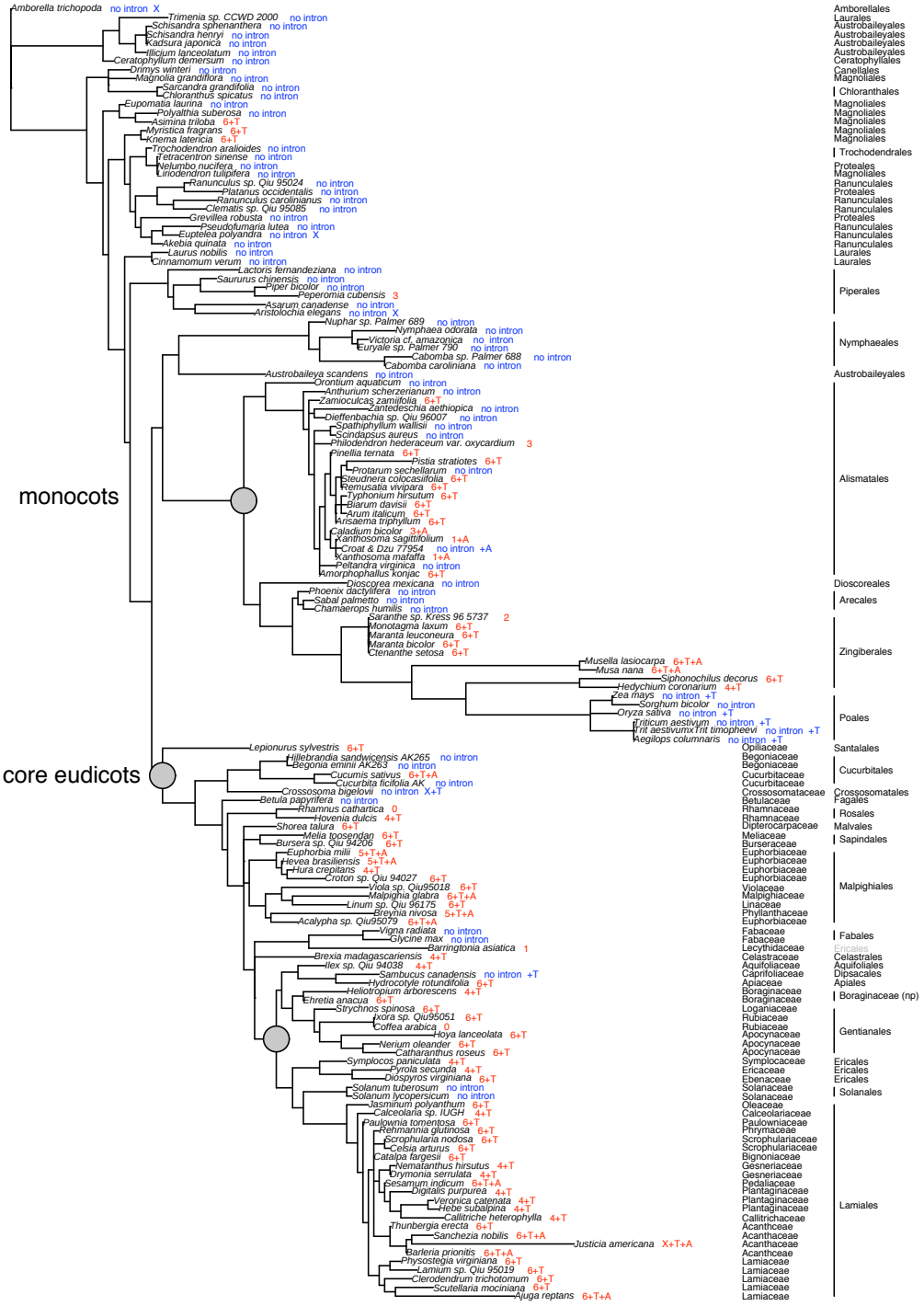


Figure 2.3: Parsimony tree for 148 angiosperms obtained from *cox1* exon sequences (1288 characters, 244 parsimony informative), rooted with *Amborella trichopoda*, intron absence (blue) and presence (red), coconversion tract types as well as plant orders are plotted. Orders with a vertical line before their name were recovered as monophyletic.

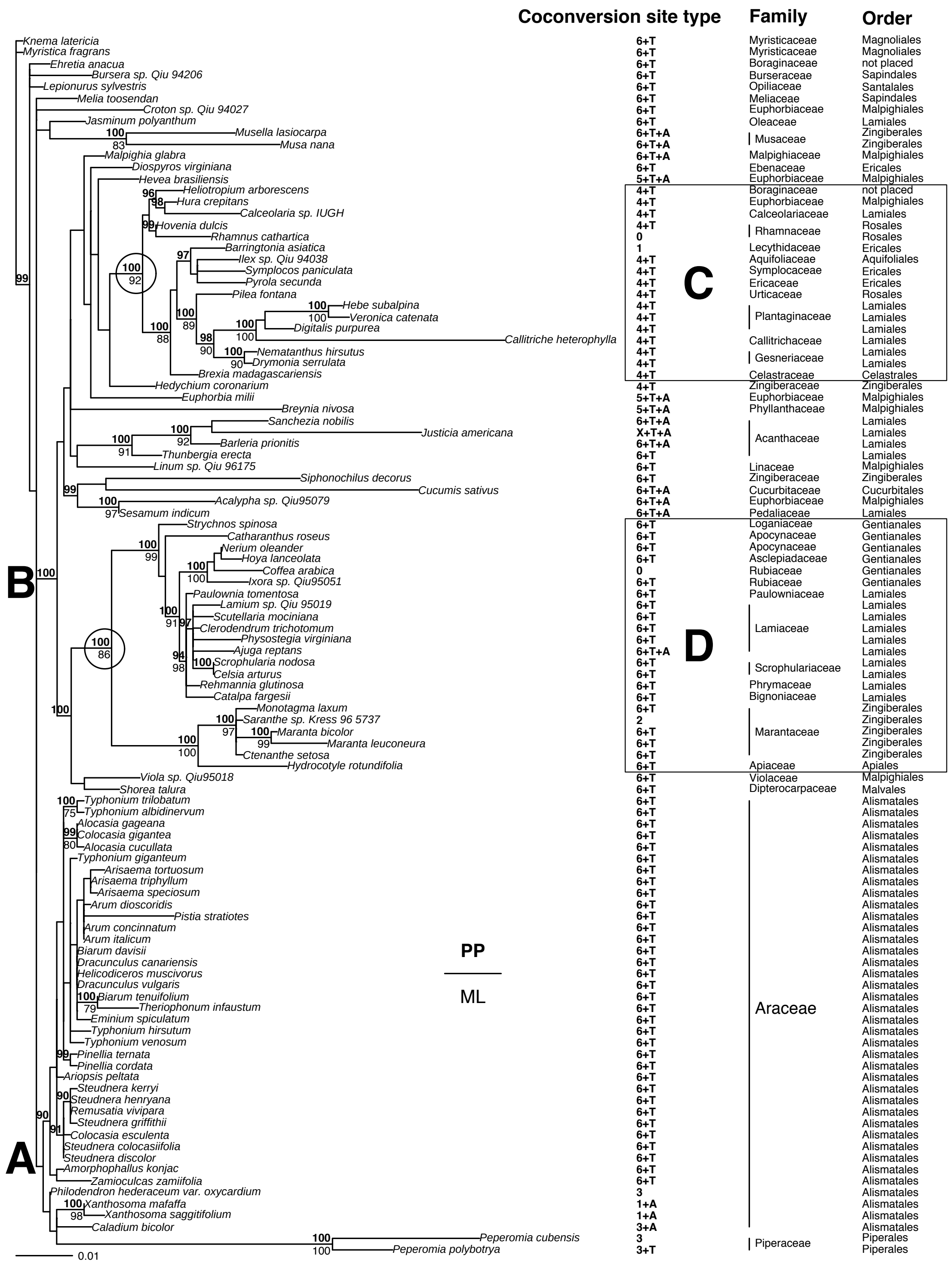


Figure 2.4: Maximum likelihood tree for 106 angiosperms based on mitochondrial *cox1* intron sequences (967 bp) analyzed under a TvM + I + Γ model. Values above branches indicate a posterior probability >90, values below branches a maximum likelihood bootstrap value >75. Coconversion tract types (see Fig. 2.2A) of the respective taxa are given on the right. Families with a vertical line before their name were recovered as monophyletic. A to D are clades of interest discussed in the text.

maining monocots as well as all eudicots. Clade C includes species with a 4+T coconversion tract, except for *Rhamnus* (no substitution in its coconversion tract) and *Barringtonia* (coconversion tract with 1 substitution). The closest relatives of clade C (not statistically supported) also have 4+T tracts or 5+T+A tracts. Clade D, finally, unites 22 species with mainly 6+T (one with 6+T+A) coconversion tracts, again with two exceptions: *Coffea* (no substitution in its coconversion tract), *Saranthe* (2 substitutions). Natural groups recovered within clade D are Lamiales (all Lamiaceae and Scrophulariaceae, amongst others), Gentianales and Zingiberales. Members of nine families form well-supported clades, namely Acanthaceae, Araceae, Gesneriaceae, Marantaceae, Musaceae, Piperaceae, Plantaginaceae, Rhamnaceae, and Scrophulariaceae.

Sequence similarity among the 110 *cox1* introns of the angiosperms ranges from 91% to identical (GenBank maximal identities with BLAST values of zero in each case), while the genetically closest non-angiosperms *cox1* introns (all in fungi) differ greatly from each other and from angiosperms (supplementary Fig. S2 and below, The Possible Origin of the *cox1* Intron from Fungi and Intron Functionality). Araceae appear to have especially low *cox1* intron mutation rates, judging from mean branch lengths of 0.012 (± 0.004) in clade A (excluding the fast-mutating and phylogenetically misplaced *Peperomia*, above), compared to 0.038 (± 0.013) in clade B (Fig. 2.4).

Hierarchical Distribution of *cox1* Exonic Coconversion Tracts and Time Frame of *cox1* Intron Loss in the Araceae

When plotting Araceae exonic tract types on an Araceae phylogeny (S. Mayo, personal communication, Feb. 2007) short tract types are found in derived positions, long tract types in basal positions. For example, *Xanthosoma* with a short coconversion tract of 1+A in both species sequenced is derived relative to *Amorphophallus* (6+T). Similarly, *Philodendron hederaceum* var. *oxycardium*, with a short coconversion tract (3 substitutions; Fig. 2.2A), is derived relative to *Zamioculcas* with 6+T.

A relaxed molecular clock applied to the 4360-nt matrix of combined chloroplast and mitochondrial data (*cox1* exon, *trnL* intron and spacer, *rpl20-rps12* intergenic spacer, and *nad1* b/c exon and intron) yielded an age of 111 (SD 91-131) MY for

the stem of the intron$^-$ *Typhonodorum lindleyanum*/*Peltandra virginica* clade and of 70 (SD 60-91) MY for the divergence between these two species (Fig. 2.1B, nodes 1 and 2). The divergence of the intron$^-$ Vietnamese species *Croat and Dzu 77954* from *Remusatia vivipara* is estimated as having occurred 49 (SD 30-67) MY ago (Fig. 2.1B, node 3). The fourth intron$^-$ species, *Protarum sechellarum*, is not securely placed by our data (Fig. 2.1A), but may have diverged from the remaining *Arisaema*/*Pistia* clade at about 73 (SD 60-82) MY ago (Fig. 2.1B, node 4).

The Possible Origin of the *cox1* Intron from Fungi and Intron Functionality

Vaughn et al. (1995) who first reported on the *cox1* intron in angiosperms assumed that its endonuclease was functional because of the presence of two LAGLI-DADG motifs (Belfort and Perlman, 1995). Experimental confirmation of homing ability is still lacking. Blasting of the hypothetical protein from *Arum concinnatum* (306 residues) yielded a BLAST value of 5 x e^{-120} (71% identical, 83% positives) with ORF 305 of the rice mould *Rhizopus oryzae* (a basal fungal lineage, formerly placed in Zycomycetes, family Mucoraceae), 3 x e^{-103} (63% identical, 78% positives) with an "unknown" region (fide GenBank) in the Oyster mushroom *Pleurotus ostreatus* (Agaricomycetes, Basidiomycota), 8 x e^{-75} (48% identical, 64% positives) with ORF 318 of *Monoblepharella* sp. (Chytridiomycota, Monoblepharidaceae), and of 3 x e^{-67} (46% identical, 63% positives) with the *cox1* aI4 intronic protein of *Saccharomyces cerevisiae* (Ascomycota, Saccharomycetaceae), which encodes site-specific DNA endonuclease and RNA maturase activities (Wenzlau et al., 1989). The putative *cox1* intron endonucleases of angiosperms have sequence similarities of 98% (*Philodendron oxycardium*) to 86% (*Peperomia grisoargentea*) with that of *Arum concinnatum*, and the entire *cox1* intron sequence of *A. concinnatum* has sequence similarities of 77% with *Rhizopus*, 71% with *Pleurotus* and *Cryptococcus* (both Basidiomycota), and 66% with *Monoblepharella* (supplementary Fig. S2).

2.5 Discussion

The *cox1* Intron in the Araceae – A Long History of Vertical Inheritance

Considering first the distribution of the *cox1* intron on a phylogeny of the *Arisaema*/*Pistia* clade, sampled for all its genera (Renner and Zhang, 2004), it is parsimoniously explained by vertical inheritance as suggested by Cho and Palmer (1999). All intron$^+$ species in this clade have the same 6+T coconversion tract type. An intron loss occurred in *Croat and Dzu 77954*, which is embedded among intron$^+$ relatives and has a coconversion tract with a single substitution (an A in position 21). Two further losses apparently occurred in *Protarum seychellarum* and in the common ancestor of the outgroup species *Typhonodorum* and *Peltandra* (Fig. 2.1A). Judging from the fossil-constrained relaxed molecular clock, the *cox1* intron has persisted in the genomes of the *Arisaema*/*Pistia* clade for at least 80 MY (Fig. 2.1B, node 5). If it is ancestral in the Araceae, not just the *Arisaema*/*Pistia* clade, as suggested by the intron's phylogenetic signal, which matches the Araceae family tree (Results), it may have persisted for 110 MY (oldest Araceae fossils, 110-120 MY; Friis et al., 2004). The timing of at least one intron loss can also be inferred. The stem lineage of the intron$^-$ *Typhonodorum*/*Peltandra* clade, which comprises just three species, is between 110 and 70 MY old (Results). Its sister clade consists of a similarly species-poor group (*Ambrosina*, with one species, *Arophyton* with three species, and *Arisarum* also with three species) that appears to have the intron (NC, unpublished data for *Arisarum vulgare*). Intron loss in the *Typhonodorum*/*Peltandra* clade could have occurred some 70 MY ago, with the coconversion tracts found in *Typhonodorum* and *Peltandra* persisting since then.

The Araceae have four exonic coconversion tract types, 6+T, 3+T, 3, or 1+A (Fig. 2.2A). Based on a small taxon sample, a tract with seven substitutions (6+T) appeared synapomorphic for the *Arisaema*/*Pistia* clade (Cho and Palmer, 1999), but when all angiosperm coconversion tracts are compared to the same reference Araceae, *O. aquaticum*, as used in Cho and Palmer (1999), it is clear that the 6+T type is the predominant *cox1* exonic coconversion tract of intron$^+$ angiosperms (figs. 2.2-2.4). It also appears that the *cox1* exonic coconversion tracts in Araceae may be hierarchically nested, with species having 3 or 1 difference(s) in their tracts

phylogenetically more derived than species with 6 differences. Such a pattern might be expected if the 6+T-type coconversion tract arose when the *cox1* intron first inserted itself into some ancestral Araceae (or angiosperm; see below) and was then passed on vertically, occasionally undergoing back mutation (which would lead to "shorter" coconversion tracts, viz. 6+T > 5+T > 4+T > 3+T, etc., Fig. 2.2A).

Based on the data available now, Araceae exonic coconversion tracts are less static than thought previously (Cho et al., 1998b; Cho and Palmer, 1999; Palmer et al., 2000), when it was argued that, "Regardless of how closely related they are, any two taxa whose coconversion tracts differ probably acquired their introns separately. For example, *Amorphophallus* and *Xanthosoma* are sister taxa with 85% bootstrap support, and thus are inferred to have received their introns by vertical transmission according to all parsimony models of intron distribution (Fig. 2.2B-E). However, because their coconversion tracts differ, and substantially so (Fig. 5 [compare our figs. 2.2A and 2.4]), we conclude that they most likely acquired their introns by two separate, and recent, horizontal transfers. By the same logic, we conclude that *Philodendron* and *Zamioculcas*, which cluster weakly in the shortest angiosperm tree (Fig. 2.2A), also acquired their introns separately (Fig. 6)." (p. 1161: Cho and Palmer, 1999).

The *cox1* Intron in the Angiosperms – Predominant Loss, not Horizontal Transfer

The hypothesis of multiple gains of the angiosperm *cox1* intron via horizontal gene transfer (Cho et al., 1998b; Cho and Palmer, 1999; Palmer et al., 2000; Richardson and Palmer, 2007) was based on relatively small taxon samples, making it appear that, "Given that we have still sampled only a tiny fraction of the >300,000 species of angiosperms, we are confident that the intron has been horizontally acquired at least hundreds of times during angiosperm evolution and probably over 1,000 times. Equally remarkably, all of these transfers seem to have occurred very recently, in the last 10 million years or so of angiosperm evolution." (p. 6965: Palmer et al., 2000). Evidence for independent gains came mostly from phylogenetic incongruence between intron and angiosperm phylogenies, patchy distribution of the intron, and analyses of exonic coconversion tracts, similar to the arguments used

in the case of Araceae (Cho and Palmer, 1999).

Considering the argument from phylogenetic incongruence between *cox1* intron and angiosperm phylogenies, the current data suggest a different interpretation. The three pairs of angiosperm genera (*Ilex/Hydrocotyle, Symplocus/Diospyros*, and *Maranta/Hedychium*) for which Cho et al.'s (1998b) data showed strong disagreement between the intron and the angiosperm phylogeny are not recovered with the current larger taxon sample (Fig. 2.4). And although the intron phylogeny contains many phylogenetically incorrect groups, it recovers an even larger number of correct clades at the species, genus, family, and even ordinal level (Fig. 2.4). The odd groupings found by Cho et al. (1998b) and in the current intron phylogeny (Fig. 2.4) are probably due to low sequence variability of the *cox1* intron leading to random groupings, and to a few taxa with higher mutation rates, causing long-branch-attraction. Regarding the coconversion tracts previously seen as evidence for or against a vertical or horizontal intron history, renewed analysis leads to the different conclusion. Several of the family-level clades recovered in the *cox1* intron tree (Fig. 2.4) include species that differ in their coconversion tracts (as shown in the figure). This is the case in Araceae, Marantaceae, Acanthaceae, and Rhamnaceae. The simplest explanation of this is that in each case the intron is inherited vertically, with the exonic tracts decaying stochastically over time. Conversely, in the densely sampled order Lamiales (Fig. 2.4), 6+T and 4+T coconversion tracts sort by family, and such slow exonic tract decay also predominates in the generally slowly evolving Araceae (at least in terms of their *cox1* sequences), which continue to pass on an ancient exonic tract. Taxa with high mitochondrial mutation rates, on the other hand, also undergo rapid changes in their coconversion tracts, as seen in *Plantago* (Cho et al., 1998a, 2004, our supplementary Fig. S1). There is also a slight positive correlation between a clade's species sampling density and its tract type diversity (supplementary Fig. S3).

The only finding suggestive of horizontal *cox1* intron transfer is a clade of phylogenetically unrelated taxa in the intron phylogeny that comprises many species with a 4+T tract type (clade C in Fig. 2.4; the clade also includes a few other tract types). We compared the *cox1* sequences of these taxa and found that they share four synapomorphic changes in loops L3 and L5 of the intron's predicted secondary structure (Vaughn et al., 1995). These substitutions, which do not seem to correlate with other changes in the intron or its coconversion tract, explain the

high bootstrap support of the 4+T clade (three non-homoplastic changes will lead to a bootstrap support of 95%; Felsenstein, 1985). The level of support for the 4+T clade is thus in fact not very high. A second observation arguing against horizontal transfer is that the subgroups inside the 4+T clade are monophyletic at family level (Rhamnaceae, Plantaginaceae, Gesneriaceae) or even the ordinal level (Lamiales). Vertical inheritance of the 4+T tract type, and insufficient phylogenetic signal in the *cox1* intron to recover relationships at hierarchical levels above the order, thus remain the simplest explanation for all groupings in Fig. 2.4.

Together these results suggest that differences in *cox1* coconversion tracts do not necessarily imply independent horizontal gene transfer and that phylogenetic evidence fits with a vertical history of the intron in angiosperms or at least fails to contradict it with statistical support. A largely vertical history also fits with the similar length of the intron across all angiosperms, its position at the same site in the *cox1* gene, and its generally high nucleotide similarity. Had there been thousands of horizontal transfers of the intron (perhaps over the past 10 million years; Palmer et al., 2000) the intron phylogeny would hardly recover as many natural groups as it does nor would one expect all angiosperm introns to be essentially equally distant from the closest fungal *cox1* intron (supplementary Fig. S2). The "high frequency angiosperm-to-angiosperm horizontal transfer" hypothesis for the *cox1* gene (Richardson and Palmer, 2007) also faces the difficulty of the still unknown transferring agent, although this is not a strong argument against lateral transfer.

Possible Mechanisms of *cox1* Intron Loss

One of the reasons why Cho et al. (1998b) preferred a hypothesis of multiple intron gains over multiple losses was that each plant cell contains thousands of mitochondrial genomes. Mitochondrial genes that have lost an intron should therefore suffer an onslaught of homing introns coming from other genomes in the same cell as long as the introns' homing endonucleases are intact. However, there are so far no experimental data showing that the ORF-encoded protein in angiosperms *cox1* introns functions as endonuclease. Conceivably, the angiosperm *cox1* intron ORF long ago lost its endonuclease function and now acts only as maturase for the splicing process (Delahodde et al., 1989; Wenzlau et al., 1989; Haugen et al.,

2005). If this were the case, intron reinsertion by homing would not longer be possible.

Molecular mechanisms for intron loss are either recombination between an intron$^+$ an intron$^-$ gene or recombination between the genomic copy of an intron$^+$ gene and a reverse transcribed copy of spliced mRNAs (Dujon, 1989; Roy and Gilbert, 2005; Roy and Penny, 2007); another mechanism is genomic deletion as in the intron presence-absence polymorphism in *Drosophila* (Llopart et al., 2002). The *cox1* intron is always gained or lost in one step, because it is self-splicing and can only function if the entire intron is inserted. For the angiosperms, we assume that the intron is lost by gene conversion (i.e., by one of the above two recombination mechanisms). That the distribution of the most common coconversion tract types is so biased, with most intron$^-$ angiosperms having the 0 tract type (Fig. 2.1A, top panel), most intron$^+$ angiosperms the 6+T tract (Fig. 2.1A, bottom panel), suggests that one reflects an event during intron insertion, the other an event correlated with intron loss. (It is also possible that the original intron donor and the first angiosperm recipient had identical *cox1* tracts and there was no coconversion. The 6+T tract would then simply be an ancestral angiosperm *cox1* sequence and the 0 type would be the "footprint" of intron loss.) In the long run, selection on the host should favor intron loss. We suggest that the *cox1* coconversion tract is usually lost during the intron excision process, which would explain the similar coconversion tract in most intron$^-$ angiosperms. A stage in the angiosperm life cycle at which such loss might logically occur is during megaspore or zygote formation, where number of mitochondria is reduced and changes in the mitochondrial genome could spread more easily. (Only maternal mitochondria closest to the egg cell become part of the zygote.)

Fungi as Donors of the *cox1* Intron in Angiosperms

Regarding the possible donor of the angiosperm and/or Araceae *cox1* intron, the current hypothesis is that it came from a fungus (Vaughn et al., 1995; Adams et al., 1998; Cho et al., 1998b; Seif et al., 2005). This idea is based on two observations. First, the *cox1* intron is the only group I intron in vascular plant mtDNAs, while in fungi, group I introns in the *cox1* gene are common. Second, most angiosperms have symbiotic interactions with fungi, providing a conceivable way of

intron transfer from a fungus to an angiosperm. A recent study that analyzed fungi group I introns with ORFs, including eight in the *cox1* gene, found a *cox1* intron in *Rhizopus oryzae* (*cox1*-i1-ORF 305) that was similar to the *cox1* intron of angiosperms (Seif et al., 2005), leading to the suggestion that the angiosperm *cox1* intron "originated in a zygomycete close to *Rhizopus*." Renewed BLAST searching of angiosperm *cox1* introns (28 August 2007) still yields the widespread mould *R. oryzae* and the Oyster mushroom *Pleurotus ostreatus* as the closest relatives outside of angiosperms (supporting material Fig. S2). However, sampling in the fungi is extremely sparse and sequence homology low.

Regardless of how many times and from which fungus the *cox1* intron entered the angiosperms, such entry was hardly a straightforward process because of differences in the genetic code used by fungi and angiosperms (Fox, 1987). For intron homing to function, the encoded endonuclease must be translated, and differences in codes may cause difficulties in translation. Further difficulties are the existence of C > U RNA editing in plant but not fungal mitochondria (Gray, 1996) and differences in promoter sequences recognized by the fungal and plant mitochondrial transcriptional apparatus (Tracy and Stern, 1995). Nevertheless, there is indirect evidence that angiosperm-to-fungus intron transfer can occur (Nishida and Sugiyama, 1995). The *cox1* introns of other spermatophytes, e.g., *Marchantia polymorpha* (Ohta et al., 1993), are more distant from angiosperm *cox1* introns than are fungal *cox1* introns.

2.6 Conclusion

For Araceae, the fit between the *cox1* intron and the Araceae phylogeny, and the highly conserved coconversion tracts together suggest vertical intron inheritance over 110 MY, with several independent losses within Araceae. Current data for all angiosperms likewise point to a history dominated by vertical intron inheritance followed by repeated intron loss. The alternative hypothesis of numerous horizontal acquisitions has difficulties explaining the observed congruence between the intron and the angiosperm phylogeny as well as the evidence from the 20 coconversion tract types found across angiosperms. Coconversion tracts can no longer be regarded as static footprints. Instead their analysis in a phylogenetic framework provides evidence of their gradual decay and loss, most likely at the excision

stage and by RT-mRNA-mediated coconversion. The hypothesis that fungi are the source of the angiosperm *cox1* intron fits with current data, but sampling in fungi is still extremely sparse, and specific donor lineages can therefore not yet be named.

2.7 References

Adams, K. L., M. J. Clements, and J. C. Vaughn. 1998. The *Peperomia* mitochondrial *coxI* group I intron: Timing of horizontal transfer and subsequent evolution of the intron. J. Mol. Evol. 46:689–696.

Belfort, M. 2003. Two for the price of one: a bifunctional intron-encoded DNA endonuclease-RNA maturase. Genes Dev. 17:2860–2863.

Belfort, M. and P. Perlman. 1995. Mechanisms of intron mobility. J. Biol. Chem. 270:30237–30240.

Bell-Pedersen, D., S. M. Quirk, M. Aubrey, and M. Belfort. 1989. A site-speciic endonuclease and co-conversion of flanking axons associated with the mobile *td* intron of phage T4. Gene 82:119–126.

Braithwaite, C. J. R. 1984. Geology of the Seychelles. Pages 17–38 *in* Biogeography and ecology of the Seychelles Islands (D. Stoddart, ed.). Junk, The Hague.

Chevalier, B. S. and B. L. Stoddard. 2001. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. Nucleic Acids Res. 29:3757–3774.

Cho, Y., K. L. Adams, Y. L. Qiu, P. Kuhlman, J. C. Vaughn, and J. D. Palmer. 1998a. A highly invasive group I intron in the mitochondrial *cox1* gene. Pages 19–23 *in* Plant mitochondria: from gene to function (M. I.-M., G. P., G. K., and G. E., eds.). Backhuys Publishers, Leiden, The Netherlands.

Cho, Y., J. P. Mower, Y.-L. Qiu, and J. D. Palmer. 2004. Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. Proc. R. Soc. Lond. B Biol. Sci. 101:17741–17746.

Cho, Y. and J. D. Palmer. 1999. Multiple acquisitions via horizontal transfer of a group I intron in the mitochondrial *cox1* gene during evolution of the Araceae family. Mol. Biol. Evol. 16:1155–1165.

Cho, Y., Y. L. Qiu, P. Kuhlman, and J. D. Palmer. 1998b. Explosive invasion of plant mitochondria by a group I intron. Proc. R. Soc. Lond. B Biol. Sci. 95:14244–14249.

Delahodde, A., V. Goguel, A. M. Becam, F. Creusot, J. Perea, J. Banroques, and C. Jacq. 1989. Site-specific DNA endonuclease and RNA maturase activities of two homologous intron-encoded proteins from yeast mitochondria. Cell 56:431–441.

Dujon, B. 1989. Group I intron as mobile genetic elements: facts and mechanistic speculations—a review. Gene 82:91–&.

Felsenstein, J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.

Fox, T. D. 1987. Natural variation in the genetic code. Annu. Rev. Genet. 21:67–91.

Friis, E. M., K. R. Pedersen, and P. R. Crane. 2004. Araceae from the Early Cretaceous of Portugal: Evidence on the emergence of monocotyledons. Proc. Natl. Acad. Sci. USA 101:16565–16570.

Goddard, M. R. and A. Burt. 1999. Recurrent invasion and extinction of a selfish gene. Proc. Natl. Acad. Sci. USA 96:13880–13885.

Gray, M. W. 1996. RNA editing in plant organelles: A fertile field. Proc. Natl. Acad. Sci. USA 93:8157–8159.

Haugen, P., D. M. Simon, and D. Bhattacharya. 2005. The natural history of group I introns. Trends Genet. 21:111–119.

Hughes, N. F. 1994. The Enigma of Angiosperm Origins. Cambridge: Cambridge Univ. Press.

Lambowitz, A. M. 1989. Infectious introns. Cell 56:323–326.

Lambowitz, A. M. and M. Belfort. 1993. Introns as mobile genetic elements. Annu. Rev. Biochem. 62:587–622.

Lang, B. F. 1984. The mitochondrial genome of the fission yeast *Schizosaccharomyces pombe*: highly homologous introns are inserted at the same position of the otherwise less conserved *coxI* genes in *Schizosaccharomyces pombe* and *Aspergillus nidulans*. Embo J. 3:2129–2136.

Llopart, A., J. M. Comeron, F. G. Brunet, D. Lachaise, and M. Long. 2002. Intron presence-absence polymorphism in Drosophila driven by positive Darwinian selection. Proc. Natl. Acad. Sci. USA 99:8121–8126.

Maddison, W. P. and D. R. Maddison. 1992. MacClade: analysis of phylogeny and character evolution. Version 3.0. Sinauer, Sunderland, Massachusetts, USA.

Nishida, H. and J. Sugiyama. 1995. A Common Group I Intron between a Plant Parasitic Fungus and Its Host. Mol. Biol. Evol. 12:883–886.

Ohta, E., K. Oda, K. Yamoto, Y. Nakamura, M. Takemura, N. Nozato, K. Akashi, K. Ohyama, and F. Michel. 1993. Group I introns in the liverwort mitochondrial genome: the gene coding for subunit 1 of cytochrome oxidase shares five intron positions with its fungal counterparts. Nucleic Acids Res. 21:1297–1305.

Palmer, J. D., K. L. Adams, Y. R. Cho, C. L. Parkinson, Y. L. Qiu, and K. M. Song. 2000. Dynamic evolution of plant mitochondrial genomes: Mobile genes and introns and highly variable mutation rates. Proc. Natl. Acad. Sci. USA 97:6960–6966.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290.

Posada, D. and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817–818.

Qiu, Y. L., O. Dombrovska, J. Lee, L. B. Li, B. A. Whitlock, F. Bernasconi-Quadroni, J. S. Rest, C. C. Davis, T. Borsch, K. W. Hilu, S. S. Renner, D. E. Soltis, P. S. Soltis, M. J. Zanis, J. J. Cannone, R. R. Gutell, M. Powell, V. Savolainen, L. W. Chatrou, and M. W. Chase. 2005. Phylogenetic analyses of basal angiosperms based on nine plastid, mitochondrial, and nuclear genes. Int. J. Plant Sci. 166:815–842.

R Developmental Core Team. 2006. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Renner, S. S. and L.-B. Zhang. 2004. Biogeography of the *Pistia* clade (Araceae): Based on chloroplast and mitochondrial DNA sequences and Bayesian divergence time inference. Syst. Biol. 53:422–432.

Richardson, A. O. and J. D. Palmer. 2007. Horizontal gene transfer in plants. J. Exp. Bot. 58:1–9.

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

Rot, C., I. Goldfarb, M. Ilan, and D. Huchon. 2006. Putative cross-kingdom horizontal gene transfer in sponge (Porifera) mitochondria. BMC Evol. Biol. 6.

Roy, S. W. and W. Gilbert. 2005. The pattern of intron loss. Proc. Natl. Acad. Sci. USA 102:713–718.

Roy, S. W. and D. Penny. 2007. Patterns of intron loss and gain in plants: Intron loss-dominated evolution and genome-wide comparison of O-sativa and A-thaliana. Mol. Biol. Evol. 24:171–181.

Seif, E., J. Leigh, Y. Liu, I. Roewer, L. Forget, and B. F. Lang. 2005. Comparative mitochondrial genomics in zygomycetes: bacteria-like RNase P RNAs, mobile elements and a close source of the group I intron invasion in angiosperms. Nucleic Acids Res. 33:734–744.

Stamatakis, A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

Stevens, P. F. 2001 onwards. Angiosperm Phylogeny Website. `http://www.mobot.org/MOBOT/research/APweb/` Version 9, June 2008.

Swofford, D. L. 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Version 4.0b. Sinauer Associates, Sunderland, Massachusetts.

Szostak, J. W., T. L. Orrweaver, R. J. Rothstein, and F. W. Stahl. 1983. The double-strand-break repair model for recombination. Cell 33:25–35.

Thorne, J. L. and H. Kishino. 2002. Divergence time and evolutionary rate estimation with multilocus data. Syst. Biol. 51:689–702.

Thorne, J. L., H. Kishino, and I. S. Painter. 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. 15:1647–1657.

Tracy, R. L. and D. B. Stern. 1995. Mitochondrial transcription initiation: promoter structures and RNA polymerases. Curr. Genet. 28:205–216.

Vaughn, J. C., M. T. Mason, G. L. Sperwhitis, P. Kuhlman, and J. D. Palmer. 1995. Fungal origin by horizontal transfer of a plant mitochondrial group I intron in the chimeric *coxI* gene of *Peperomia*. J. Mol. Evol. 41:563–572.

Wenzlau, J. M., R. J. Saldanha, R. A. Butow, and P. S. Perlman. 1989. A latent intron-encoded maturase is also an endonuclease needed for intron mobility. Cell 56:421–420.

Wilde, V., Z. Kvacek, and J. Bogner. 2005. Fossil leaves of the Araceae from the European Eocene and notes on other aroid fossils. Int. J. Plant Sci. 166:157–183.

2.8 Online Supporting Material

Supplementary Table S2.1: Sources and GenBank accession numbers of newly generated sequences as well as sequences downloaded from GenBank.

Species	Source	*TrnL* intron	*TrnL-trnF* spacer	*Rpl20-rps12* spacer	*Nad1* b/c intron	*Cox1* exon	*Cox1* intron
***Pistia* clade**							
Alocasia cucullata (Lour.) G. Don	MO acc. 751658	AY248983	AY248945	AY248908	AY243116	EF517193	EF517193
Alocasia gageana Engl. & K. Krause	MO acc. 78364	AY248984	AY248946	AY248909		EF517194	EF517194
Ariopsis peltata J. Graham	J. Murata s.n. 16 Oct 2001	AY248985	AY248947	AY248910	AY243120	EF517198	EF517198
Arisaema speciosum (Wall.) Mart.	W. Hetterscheid s.n. Jul 2002	AY248995	AY248957	AY248920	AY243115	EF517176	EF517176
Arisaema tortuosum (Wall.) Schott	Anaimudi 20/5	AY248995	AY248957	AY248920	AY243115	EF517177	EF517177
Arisaema triphyllum (L.) Torr.	Cho and Palmer 1999					AY009454	AY009454
Arum concinnatum Schott	B. W. Magrys s.n., 15. Mar. 02	AY248997	AY248959	AY248922	AY243121	EF517179	EF517179
Arum dioscoridis Sibth. & Sm.	B. W. Magrys s.n., 15. Mar. 02	AY248997	AY248959	AY248922	AY243121	EF517180	EF517180
Arum italicum Mill.	BG Mainz, 20 Jul 2001	AY248997	AY248959	AY248922	AY243121	EF517181	EF517181
Biarum davisii Turrill	MO acc. 78231	AY248998	AY248960	AY248923	AY243122	EF517182	EF517182
Biarum tenuifolium (L.) Schott	BG Bonn 16014	AY248999	AY248961	AY248924		EF517183	EF517183
Colocasia esculenta (L.) Schott	Bogner, 18.Jul. 2001, BG Munich					EF517196	EF517196
Colocasia gigantea (Blume) Hook. zf.	T. Croat & Dzu 78014 (MO)	AY249000	AY248962	AY248925	AY243117	EF517195	EF517195
Dracunculus canariensis Kunth	BG Bonn 13049	AY249001	AY248963	AY248926	AY243123	EF517184	EF517184
Dracunculus vulgaris Schott	T. Croat 78286 (MO)	AY249002	AY248964	AY248927		EF517185	EF517185
Eminium spiculatum (Blume) Schott	BG Bonn 15031	AY249003	AY248965	AY248928	AY243124	EF517186	EF517186
Helicodiceros muscivorus (L. f.) Engl.	MO acc. 71821	AY249004	AY248966	AY248929	AY243125	EF517187	EF517187
Pinellia cordata N. E. Brown	J. McClements s.n., 30 Jul 2001	AY249005	AY248967	AY248930	AY243111	EF517175	EF517175
Pinellia ternata (Thunb.) Breit.	J. McClements s.n., 30 Jul 2001	AY249006	AY248968	AY248931	AY243112	EF517178	EF517178
Pistia stratiotes L.	J. Bogner, 18 Jul 2001, BG Munich	AY249007	AY248969	AY248932	AY243126	EF517204	EF517204
Protarum sechellarum Engl.	J. Bogner 2545 (M)	AY249008	AY248970	AY248933	AY243127	EF517203	no intron
Remusatia vivipara (Lodd.) Schott	MO acc. 69705b	AY249009	AY248971	AY248934	AY243118	EF517197	EF517197
Croat & Dzu 77954	T. Croat & Dzu 77954 (MO)	AY249010	AY248972	AY248935	AY243119	EF517201	no intron
Steudnera colocasiifolia K. Koch	J. Bogner 1891 (M)	EF517218	EF517213	EF517223	EF517172	EF517208	EF517208
Steudnera discolor Bull	J. Bogner 1582 (M)	EF517216	EF517211	EF517221	EF517170	EF517199	EF517199
Steudnera grifithii (Schott) Hook f.	J. Bogner 2588 (M)	EF517219	EF517214	EF517224	EF517173	EF517210	EF517210
Steudnera henryana Engl.	J. Bogner 2619 (M)	EF517217	EF517212	EF517222	EF517171	EF517200	EF517200
Steudnera kerrii Gagnep.	J. Bogner 2291 (M)	EF517220	EF517215	EF517225	EF517174	EF517209	EF517209
Theriophonum dalzelii Schott	J. Murata s.n., 21 Aug 2002	AY249011	AY248973	AY248936	AY243128		
Theriophonum infaustum N.E.Br.	P. Bruggemann, PB 099, Apr. 05					EF517202	EF517202
Typhonium albidinervum Tang & Li	J. Murata 1	AY249012	AY248974	AY248937	AY243129	EF517192	EF517192
Typhonium giganteum Engl.	J. W. Waddick s.n., 20 Aug. 2001	AY249013	AY248975	AY248938	AY243130	EF517189	EF517189
Typhonium hirsutum (S. Y. Hu) Murata & Mayo	W. Hetterscheid H.AR 036	AY249014	AY248976	AY248939		EF517190	EF517190
Typhonium trilobatum (L.) Schott	J. Murata 5	AY249016	AY248978	AY248941	AY243131	EF517188	EF517188
Typhonium venosum (Dryand. ex Aiton) Hett. & P.C.Boyce	J. Bogner s.n. (M), 27. Jun. 02					EF517191	EF517191
Araceae outgroups							
Caladium bicolor (Aiton) Vent.	T. Croat 60868 (MO)	AY249018	AY248980	AY248943	AY243134	EF517207	EF517207
Peltandra virginica Raf.	J. Bogner 2119 (M)	AY249017	AY248979	AY248942	AY243132	AJ007550	no intron
Typhonodorum lindleyanum Schott	J. Bogner s.n. (M)	AY249019	AY248981		incomplete	EF517205	no intron
Xanthosoma sagittifolium (L.) Schott & Endl.	MO acc. 850652b	AY249020	AY248982	AY248944	AY243133	EF517206	EF517206
Xanthosoma mafaffa Schott	Cho and Palmer 1999					AJ223807	AJ223807
Amorphophallus rivieri Durieu ex Riviere	Cho and Palmer 1999					AJ007548	AJ007548

Supplementary Table S2.1 continued

	Anthurium scherzerianum Schott	Cho and Palmer 1999	AJ007551	no intron
	Dieffenbachia sp.	Cho and Palmer 1999	AJ007548	no intron
	Orontium aquaticum L.	Cho and Palmer 1999	AJ007551	no intron
	Philodendron hederaceum var. oxycardium Schott	Cho and Palmer 1999	AJ223438	AJ223438
	Scindapsus aureus Engl.	Cho and Palmer 1999	AJ007552	no intron
	Spathiphyllum wallisii Hort.	Cho and Palmer 1999	AJ007553	no intron
	Zamioculcas zammiifolia Engl.	Cho and Palmer 1999	AJ007547	AJ007547
	Zantedeschia aethiopica (L.) Spreng..	Cho and Palmer 1999	AJ007555	no intron
Other Angiosperms				
Escalloniaceae	*Brexia madagascariensis*	Cho et al., 1998	AJ223413	AJ223413
Burseraceae	*Bursera sp.*	Cho et al., 1998	AJ223412	AJ223412
Bignoniaceae	*Catalpa fargesii*	Cho et al., 1998	AJ223411	AJ223411
Apocynaceae	*Catharanthus roseus*	Cho et al., 1999	AJ223423	AJ223423
Lamiaceae	*Clerodendrum trichotomum*	Cho et al., 1998	AJ223414	AJ223414
Cucurbitaceae	*Cucumis sativus*	Cho et al., 1998	AJ223416	AJ223416
Scrophulariaceae	*Digitalis purpurea*	Cho et al., 1998	AJ223415	AJ223415
Ebenaceae	*Diospyros virginiana*	Cho et al., 1998	AJ223417	AJ223417
Euphorbiaceae	*Euphorbia milii*	Cho et al., 1998	AJ223418	AJ223418
Scrophulariaceae	*Hebe subalpina*	Cho et al., 1998	AJ223419	AJ223419
Boraginaceae	*Heliotropium arborescens*	Cho et al., 1998	AJ223425	AJ223425
Euphorbiaceae	*Hevea brasiliensis*	Cho et al., 1998	AJ223436	AJ223436
Apiaceae	*Hydrocotyle rotundifolia*	Cho et al., 1998	AJ223424	AJ223424
Aquifoliaceae	*Ilex sp.*	Cho et al., 1998	AJ223429	AJ223429
Myristicaceae	*Knema latericia*	Cho et al., 1998	AJ223430	AJ223430
Lamiaceae	*Lamium sp.*	Cho et al., 1998	AJ223428	AJ223428
Opiliaceae	*Lepionurus sylvestris*	Cho et al., 1998	AJ223439	AJ223439
Malpighiaceae	*Malpighia glabra*	Cho et al., 1998	AJ223433	AJ223433
Meliaceae	*Melia toosendan*	Cho et al., 1998	AJ223420	AJ223420
Myristicaceae	*Myristica fragrans*	Cho et al., 1998	AJ223434	AJ223434
Apocynaceae	*Nerium oleander*	Cho et al., 1998	AJ223421	AJ223421
Rhamnaceae	*Rhamnus cathartica*	Cho et al., 1998	AJ223422	AJ223422
Acanthaceae	*Sanchezia nobilis*	Cho et al., 1998	AJ223437	AJ223437
Symplocaceae	*Symplocos paniculata*	Cho et al., 1998	AJ223435	AJ223435
Scrophulariaceae	*Veronica catenata*	Cho et al., 1998	AJ223427	AJ223427
Euphorbiaceae	*Acalypha sp. Qiu95079*		AJ247597	AJ247597
Poaceae	*Aegilops columnaris*		ACU46764	no intron
Lamiaceae	*Ajuga reptans*		AJ247595	AJ247595
Lardizabalaceae	*Akebia quinata*		AY009429	no intron
Amborellaceae	*Amborella trichopoda*		AF193953	no intron
Aristolochiaceae	*Aristolochia elegans*		AY009431	no intron
Aristolochiaceae	*Asarum canadense*		AY009432	no intron
Annonaceae	*Asimina triloba*		AY009433	no intron
Austrobaileyaceae	*Austrobaileya scandens*		AF193954	no intron
Acanthceae	*Barleria prionitis*		AJ247601	AJ247601
Lecythidaceae	*Barringtonia asiatica*		AJ247581	AJ247581
Betulaceae	*Betula papyrifera*		U77620	no intron
Capparaceae	*Breynia nivosa*		AJ247605	AJ247605
Cabombaceae	*Cabomba caroliniana*		AY009435	no intron
Cabombaceae	*Cabomba sp. Palmer 688*		AF193949	no intron
Scrophulariaceae	*Calceolaria sp. IUGH*		AJ247585	AJ247585
Callitrichaceae	*Callitriche heterophylla*		AJ247577	AJ247577
Canellaceae	*Canella winterana*		AY009437	no intron

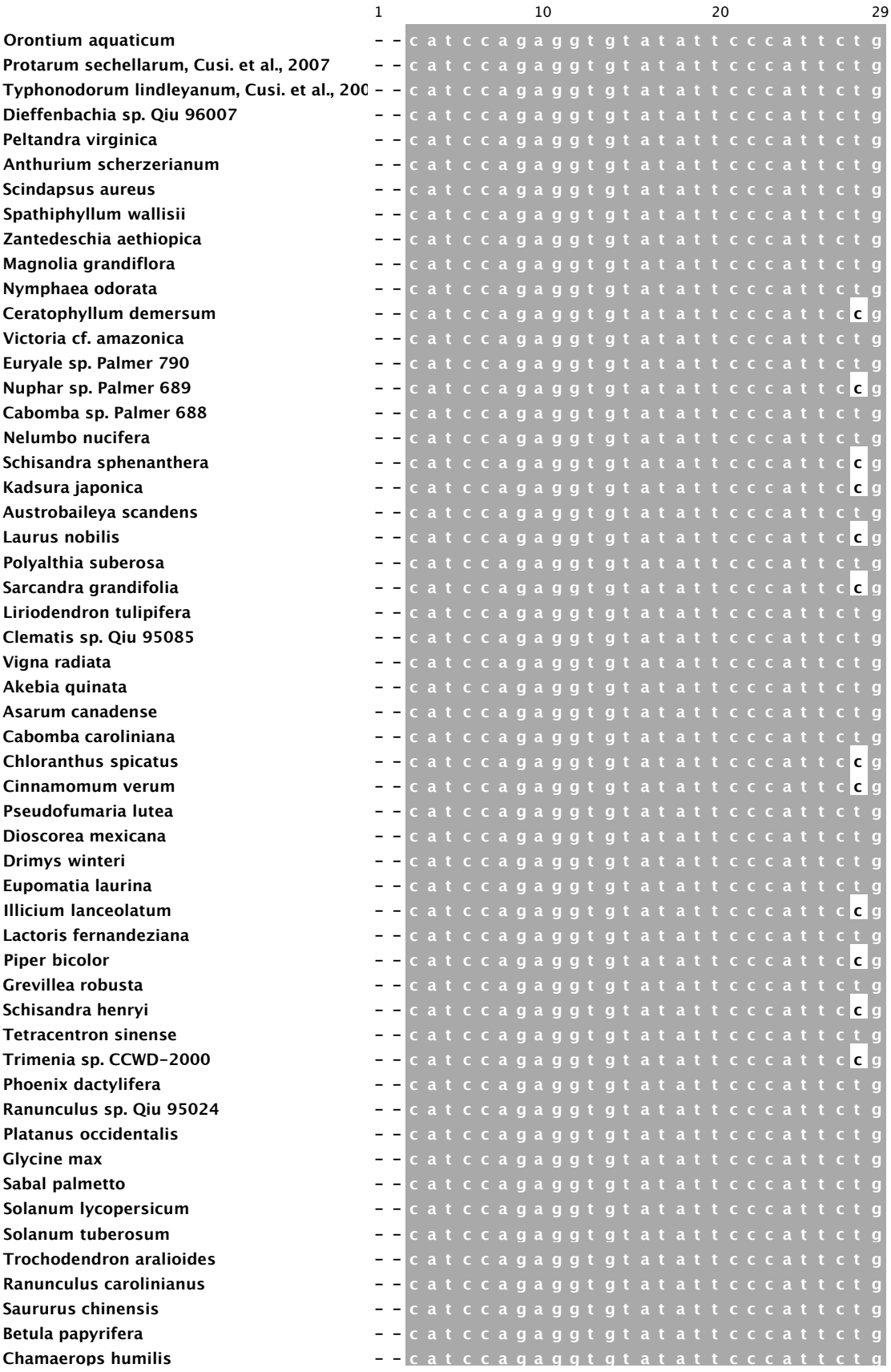
Supplementary Table S2.1 continued

Scrophulariaceae	*Celsia arturus*	AJ247590	AJ247590
Ceratophyllaceae	*Ceratophyllum demersum*	AF193945	no intron
Arecaceae	*Chamaerops humilis*	U77621	no intron
Chloranthaceae	*Chloranthus spicatus*	AY009439	no intron
Lauraceae	*Cinnamomum verum*	AY009440	no intron
Ranunuculaceae	*Clematis sp. Qiu 95085*	AF193960	no intron
Rubiaceae	*Coffea arabica*	AJ247586	AJ247586
Crossosomataceae	*Crossosoma bigelovii*	DQ317034	no intron
Euphorbiaceae	*Croton sp. Qiu 94027*	AJ247608	AJ247608
Marantaceae	*Ctenanthe setosa*	AY673019	AY673019
Dioscoreaceae	*Dioscorea mexicana*	AY009442	no intron
Winteraceae	*Drimys winteri*	AY009443	no intron
Gesneriaceae	*Drymonia serrulata*	AJ247579	AJ247579
Boraginaceae	*Ehretia anacua*	AJ247606	AJ247606
Eupomatiaceae	*Eupomatia laurina*	AY009444	no intron
Eupteleaceae	*Euptelea polyandra*	AF193963	no intron
Nymphaeaceae	*Euryale sp. Palmer 790*	AF193947	no intron
Fabaceae	*Glycine max*	M16884	no intron
Proteaceae	*Grevillea robusta*	AY009449	no intron
Zingiberaceae	*Hedychium coronarium*	AJ223426	AJ223426
Rhamnaceae	*Hovenia dulcis*	AJ247583	AJ247583
Asclepiadaceae	*Hoya lanceolata*	AJ247588	AJ247588
Euphorbiaceae	*Hura crepitans*	AJ247584	AJ247584
Illiciaceae	*Illicium lanceolatum*	AY009445	no intron
Rubiaceae	*Ixora sp. Qiu95051*	AJ247587	AJ247587
Oleaceae	*Jasminum polyanthum*	AJ247607	AJ247607
Acanthaceae	*Justicia americana*	AJ247602	AJ247602
Schisandraceae	*Kadsura japonica*	AF193952	no intron
Lactoridaceae	*Lactoris fernandeziana*	AY009446	no intron
Lauraceae	*Laurus nobilis*	AF193956	no intron
Linaceae	*Linum sp. Qiu 96175*	AJ247604	AJ247604
Magnoliaceae	*Liriodendron tulipifera*	AF193959	no intron
Magnoliaceae	*Magnolia grandiflora*	AF020568	no intron
Marantaceae	*Maranta bicolor*	AY673024	AY673024
Marantaceae	*Maranta leuconeura*	AJ223432	AJ223432
Marantaceae	*Monotagma laxum*	AY673026	AY673026
Zingiberaceae	*Musa nana*	AJ247609	AJ247609
Zingiberaceae	*Musella lasiocarpa*	AY673040	AY673040
Nelumbonaceae	*Nelumbo nucifera*	AF193950	no intron
Gesneriaceae	*Nematanthus hirsutus*	AJ247578	AJ247578
Nympheaceae	*Nuphar sp. Palmer 689*	AF193948	no intron
Nympheaceae	*Nymphaea odorata*	AF020570	no intron
Poaceae	*Oryza sativa*	X15990	no intron
Scrophulariaceae	*Paulownia tomentosa*	AJ247592	AJ247592
Piperaceae	*Peperomia cubensis*	AF029782	AF029783
Piperaceae	*Peperomia polybotrya*	X87335	X87336
Arecaceae	*Phoenix dactylifera*	AY166800	no intron
Lamiaceae	*Physostegia virginiana*	AJ247594	AJ247594
Urticaceae	*Pilea fontana*	AJ247580	AJ247580
Piperaceae	*Piper bicolor*	AY009448	no intron
Platanaceae	*Platanus occidentalis*	DQ317031	no intron
Annonaceae	*Polyalthia suberosa*	AF193957	no intron
Fumariaceae	*Pseudofumaria lutea*	AY009441	no intron
Ericaceae	*Pyrola secunda*	AJ247582	AJ247582

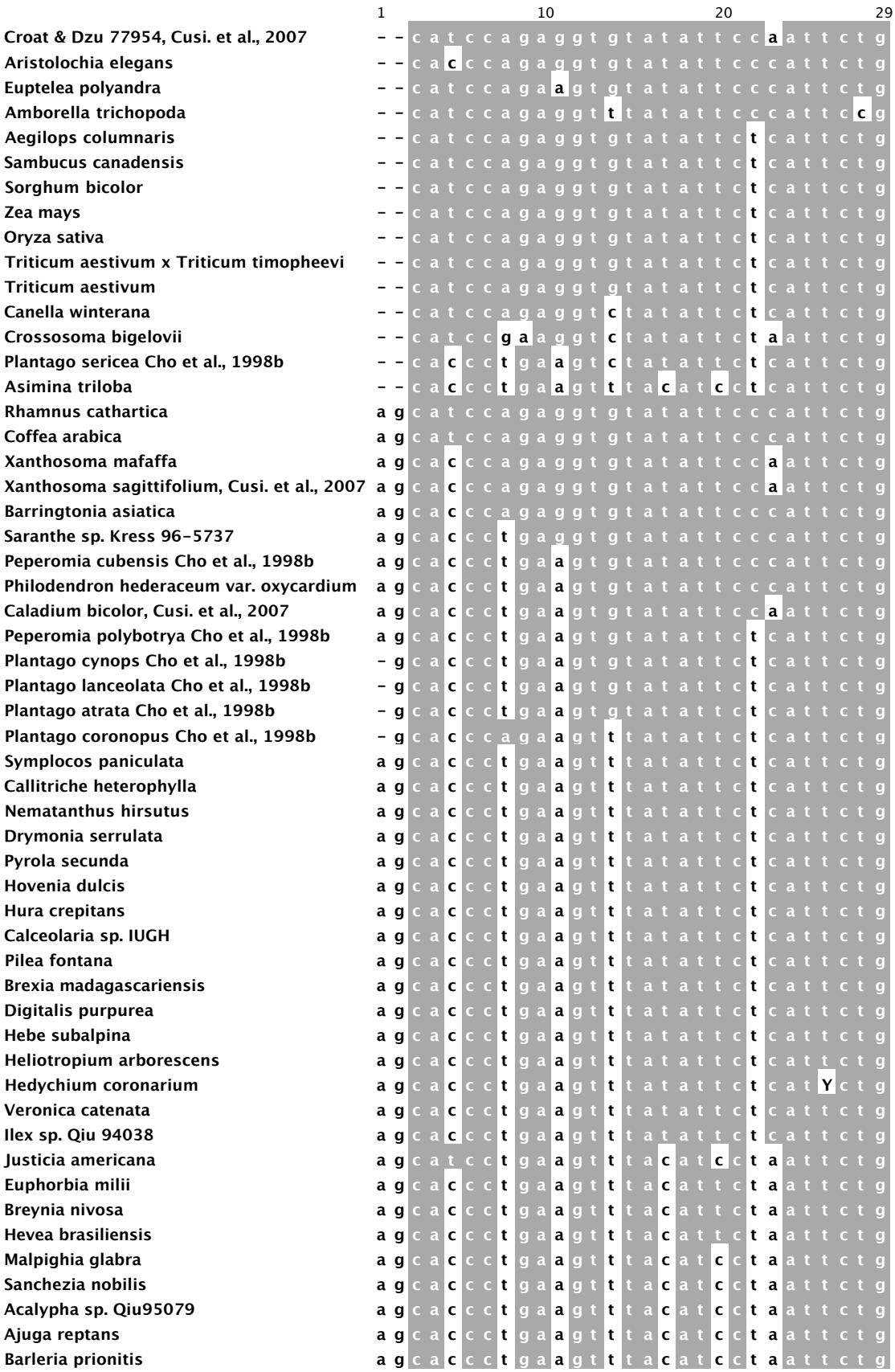
Supplementary Table S2.1 continued

Ranunculaceae	*Ranunculus carolinianus*	AY009451	no intron
Ranunculaceae	*Ranunculus sp. Qiu 95024*	DQ317030	no intron
Scrophulariaceae	*Rehmannia glutinosa*	AJ247589	AJ247589
Arecaceae	*Sabal palmetto*	U77624	no intron
Caprifoliaceae	*Sambucus canadensis*	AF193965	no intron
Marantaceae	*Saranthe sp. Kress 96-5737*	AY673030	AY673030
Chloranthaceae	*Sarcandra grandifolia*	AF193958	no intron
Piperaceae	*Saururus chinensis*	AY009452	no intron
Schisandraceae	*Schisandra henryi*	AY009453	no intron
Schisandraceae	*Schisandra sphenanthera*	AF193951	no intron
Scrophulariaceae	*Scrophularia nodosa*	AJ247591	AJ247591
Lamiaceae	*Scutellaria mociniana*	AJ247593	AJ247593
Pedaliaceae	*Sesamum indicum*	AJ247598	AJ247598
Dipterocarpaceae	*Shorea talura*	AJ247599	AJ247599
Zingiberaceae	*Siphonochilus decorus*	AY673043	AY673043
Solanaceae	*Solanum lycopersicum*	X54738	no intron
Solanaceae	*Solanum tuberosum*	X83206	no intron
Poaceae	*Sorghum bicolor*	M14454	no intron
Loganiaceae	*Strychnos spinosa*	AJ247596	AJ247596
Tetracentraceae	*Tetracentron sinense*	AY009455	no intron
Acanthceae	*Thunbergia erecta*	AJ247603	AJ247603
Monimiaceae	*Trimenia sp. CCWD-2000*	AY009456	no intron
Poaceae	*Triticum aestivum*	Y00417	no intron
Poaceae	*Triticum aestivum x Triticum timopheevi*	X56186	no intron
Trochodendraceae	*Trochodendron aralioides*	AF020581	no intron
Nympheaceae	*Victoria cf. amazonica*	AF193946	no intron
Fabaceae	*Vigna radiata*	AF338446	no intron
Violaceae	*Viola sp. Qiu95018*	AJ247600	AJ247600
Poaceae	*Zea mays*	X02660	no intron
Fungi			
Basidiomycota	*Cryptococcus neoformans*		AY560609
Mono-blepharidaceae	*Monoblepharella sp.*		AY182007
Basidiomycota	*Pleurotus ostreatus*		EF204913
Mucoraceae	*Rhizopus oryzae*		AY863212
Ascomycota	*Saccharomyces cerevisiae*		S76641

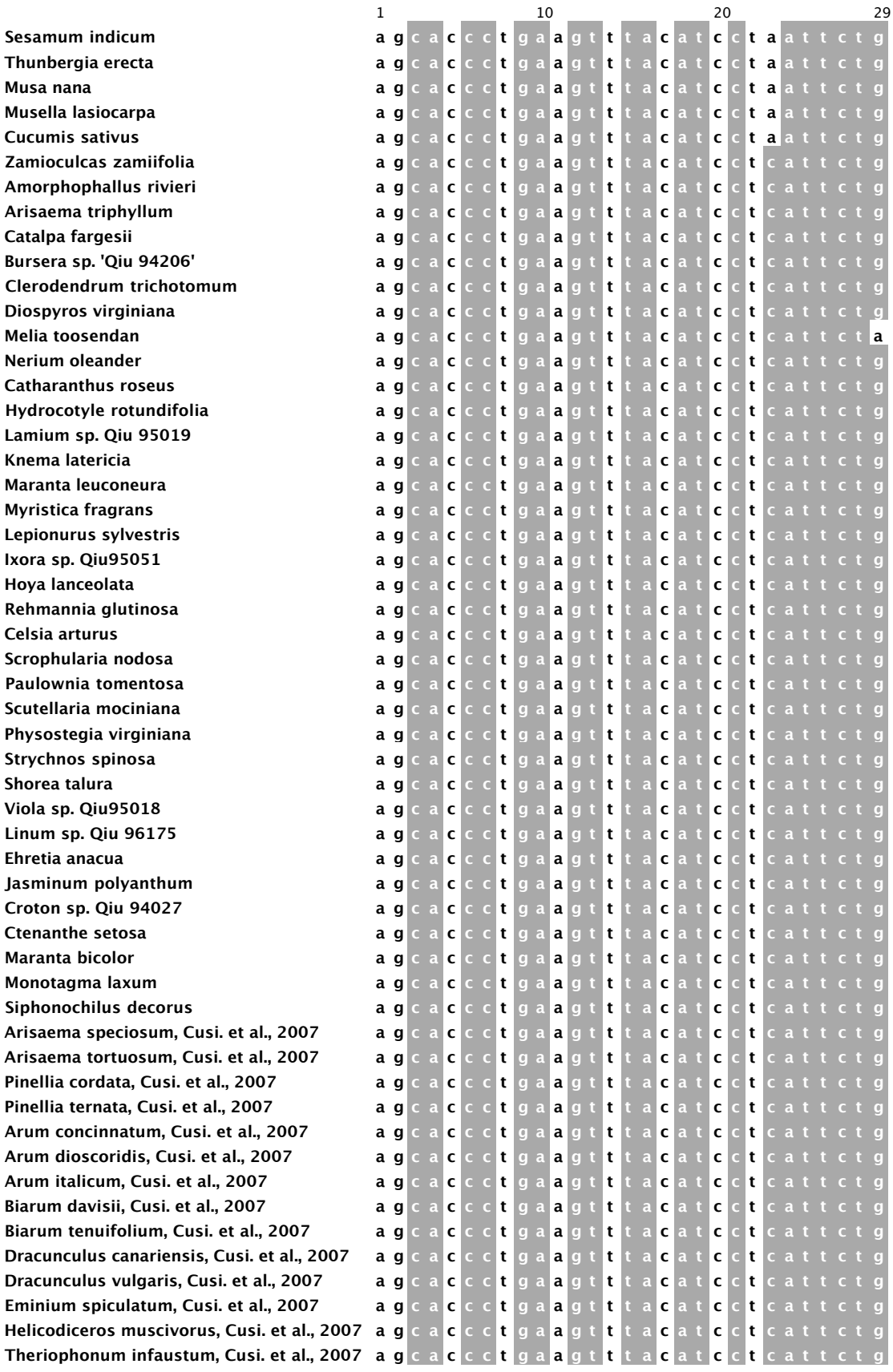
Supplementary Figure S2.1: Alignment of 179 angiosperm *cox1* exon coconversion tracts. The marked differences refer to the reference sequence Orontium aquaticum (Cho and Palmer 1999). The first two nucleotides are the last two of the intron (if present), and nucleotides 3 to 24 are the coconversion tract.



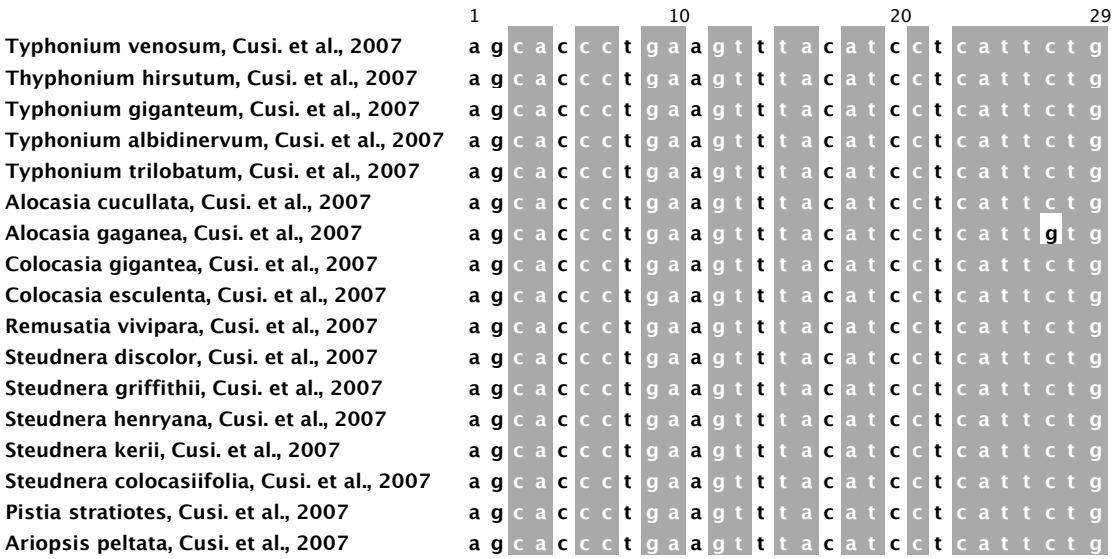
Supplementary Figure S2.1 continued



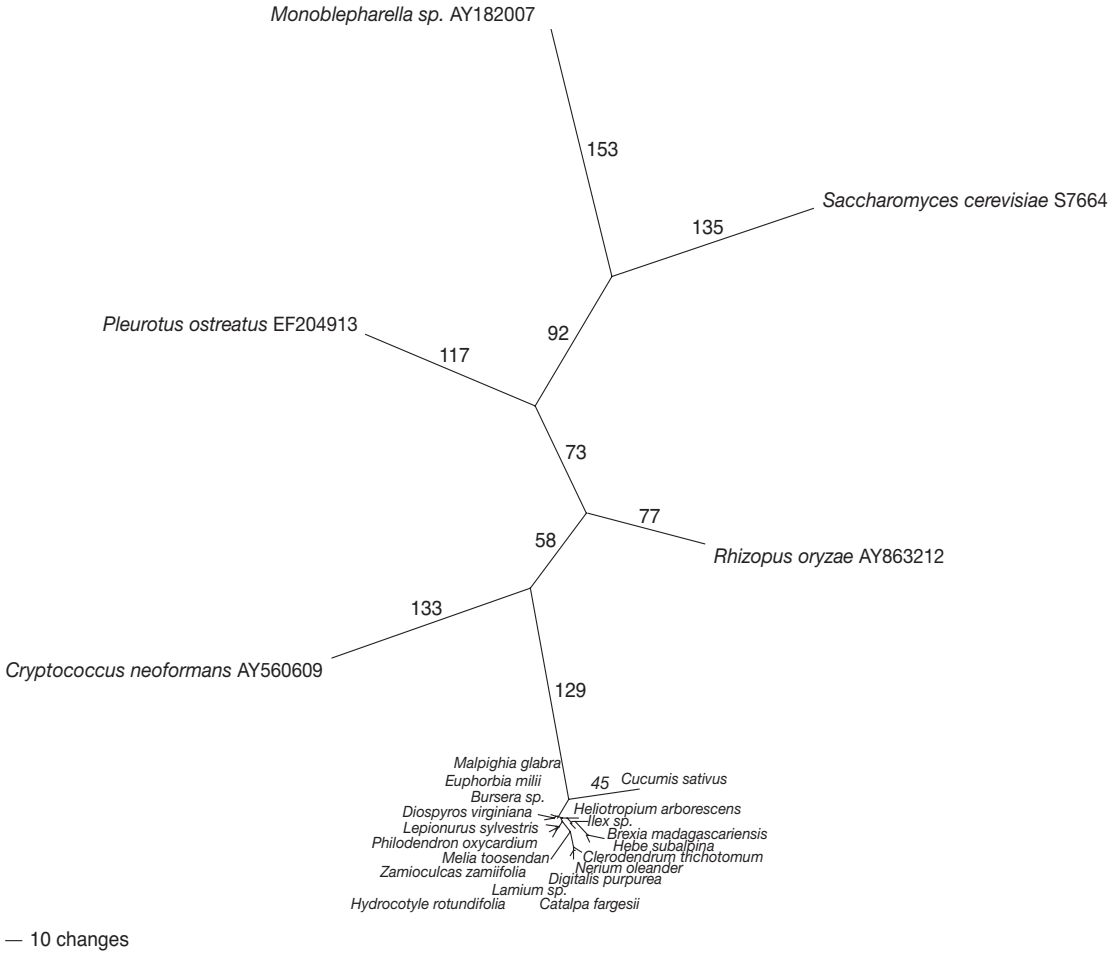
Supplementary Figure S2.1 continued



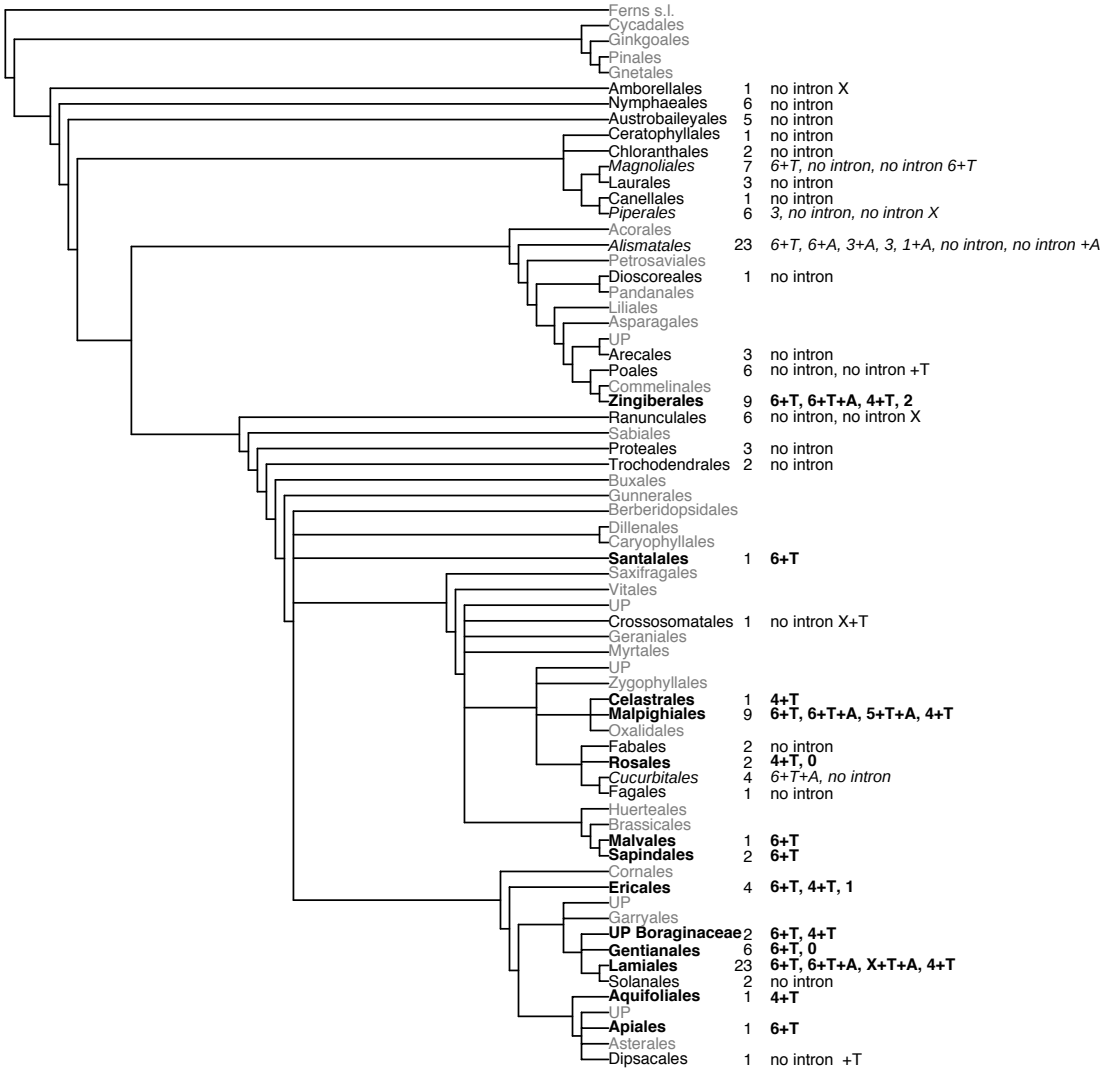
Supplementary Figure S2.1 continued



Supplementary Figure S2.2: Parsimony phylogram of angiosperm *cox1* intron sequences plus the five most similar fungi *cox1* intron sequences obtained from a GenBank BLAST search on August 28, 2007. Numbers on branches are branch lengths.



Supplementary Figure S2.3: *Cox1* exonic tract types plotted on the current phylogeny of angiosperm orders (Stevens 2001 onwards, version 8, June 2007). To the right of each order, the number of species sequenced for *cox1* and the coconversion tract types found in them. Grey: Orders not yet investigated for the cox1 intron. Black: Orders investigated for the *cox1* intron. Font styles: regular: only intron- species found, italics: intron$^-$ and intron$^+$ species found; bold face: only intron$^+$ species found.



Chapter 3

A phylogeny of the Areae (Araceae) implies that *Typhonium*, *Sauromatum*, and the Australian species of *Typhonium* are distinct clades

Cusimano, N., Barrett, M., Hetterscheid, W. L. A. and S. S. Renner. *Taxon* 59(2): 439-447.

3.1 Abstract

With in excess of 70 species, the Southeast Asian/Australian genus *Typhonium* is the largest genus of the Areae, a tribe that includes up to nine smaller genera, of which *Sauromatum* and *Lazarum* have recently been reduced to the synonymy of *Typhonium*. To test circumscriptions and relationships of *Typhonium* to the other Areae, we useed chloroplast and nuclear DNA sequences (4319 aligned nucleotides) for 86 of their 153 species, including representatives of all relevant genera. In the resulting phylogeny, *Typhonium* species fall into three well-supported clades: the first comprises most *Typhonium* species, including the type, *T. trilobatum*; the second clade consists of the type of *Sauromatum*, *S. guttatum* and other species formerly placed in that genus; the third includes only Australian endemics. Each of the remaining six genera of Areae are monophyletic. *Sauromatum* and *Typhonium* are not sister groups, requiring the recognition of *Sauromatum*. The Australian clade also needs to be ranked as a genus to achieve similar levels of morphological, geographic, and genetic distinctness among the genera of Areae. However, since only ten of the 16 described Australian endemics currently placed in *Typhonium* have so far been sequenced, not including the type of the name of the Australian genus *Lazarum*, we refrain from applying this name to the Australian clade. Among the nomenclatural and taxonomic results of this study are a key to the nine species of *Sauromatum*, and five new combinations. We also report two new chromosome counts and discuss the implications of the molecular phylogeny for the evolution of *Sauromatum* karyotypes.

Key Words

Areae, chromosome numbers, *Lazarum*, molecular phylogenetics

3.2 Introduction

The tribe Areae (Araceae) in its current circumscription comprises seven genera (Hay, 1997; Mayo et al., 1997; Hetterscheid and Boyce, 2000) and at least 153 species (including 14 awaiting description). The Areae range from Australia to Europe, with one species in Africa, and a center of diversity in SE Asia, another

in the Mediterranean region and the Near East. Two genera, the SE Asian *Sauromatum* and the Australian *Lazarum*, have recently been sunk into *Typhonium*, the largest genus of the tribe. The monophyly of this broadly circumscribed *Typhonium*, however, appears doubtful. An early DNA restriction fragment analysis that included eight species of *Typhonium* (five of these now belonging to *Sauromatum*) and representatives of two other Areae genera found *Typhonium* paraphyletic (Sriboonma et al., 1993). A concurrent morphological cladistic analysis of 36 species of *Typhonium* (including six here shown to belong in *Sauromatum*) and one Arum species yielded the same result (Sriboonma et al., 1994). That *Typhonium* and *Sauromatum* might be distinct clades was suggested by the family wide restriction fragment analysis of French et al. (1995) in which the single species of *Typhonium* and *Sauromatum* included did not form a clade, but instead a grade with *Theriophonum* falling in between (*Lazarum* was not included). A more recent chloroplast phylogeny for Areae that included five species of *Typhonium* (three of these here shown to belong in *Sauromatum*), and numerous other Areae, yielded the same result (Renner and Zhang, 2004). None of these molecular studies, however, sampled a sufficient number of species to properly test the monophyly of *Typhonium*.

The genus *Sauromatum* was erected by Heinrich Willem Schott (1832) to accommodate *S. guttatum* (Ait.) Schott and *S. pedatum* (Link & Otto) Schott (*Arum pedatum* Link & Otto), two obviously related entities from the understory of monsoon forests in India. Kunth (1841) transferred *Arum venosum* Dryand ex Ait., a species described from a cultivated specimen, to *Sauromatum* (this species is now considered conspecific with *S. guttatum* and *S. pedatum*). Shortly thereafter, Miquel (1855, 1864) added *S. pulchrum* from Sumatra and *S. horsfieldii* from Central Java. The last was transferred to *Typhonium* by van Steenis (1948) who considered it to be the same as *T. fallax* N.E. Br. and *T. pedatum* Engl. Another species moved between *Sauromatum* and *Typhonium* is *T. brevipes* (Hooker, 1893) from Sikkim in the southern Himalayas (Brown, 1903). Of these doubtfully assigned species, two are widely cultivated, *Sauromatum venosum*, the commercial curiosity marketed as the "voodoo lily", and *S. giganteum*, in spite of the vile scent they produce at peak flowering.

The most important characters used by Schott (1832) to distinguish his *Sauromatum* from *Typhonium* were the connate spathe tube, ovaries with two (rather than one) ovules, scattered staminodes and a short peduncle in *Sauromatum*. In

addition, Schott noted that *Sauromatum* produced leaves after flowering, whereas *Typhonium* produced them before or during flowering. The later discovery of intermediate forms, however, diluted these generic differences. Thus, partly fused spathe tubes are found in *S. hirsutum* (see below) and a fused spathe base, but leaves occurring during flowering, in *S. brevipes*, resulting in the above described transfers (e.g., Brown, 1903; van Steenis, 1948; Hetterscheid and Boyce, 2000). The problem of generic assignments was exacerbated by incompletely known species. For example, a fruiting specimen of *T. hirsutum* was discovered in 1958 in dry evergreen forest in Chiang Mai (Thailand) at an altitude of 1130 m and described as *Arisaema hirsutum* by Hu (1968). Some 30 years later, Murata and Mayo (1991) realized that a flowering specimen represented the same species; however, they thought it better placed in *Typhonium*, with "some interesting resemblances to its neighboring genus *Sauromatum*," such as a spathe tube fused for up to one quarter of its length and an inflorescence intermediate between *T. giganteum* and *S. venosum*. A fuller understanding of many species of *Typhonium* and *Sauromatum* was only achieved once they were brought into cultivation (Hetterscheid et al., 2001). Nevertheless, the apparently impossible morphological separation of *Sauromatum* and *Typhonium* caused Hetterscheid and Boyce (2000) to sink *Sauromatum* into *Typhonium*.

The other genus whose relationship to *Typhonium* has been difficult to assess from morphology alone is *Lazarum*. This is a monotypic entity based on an Australian species discovered on Melville Island near Darwin (Northern Territory), in 1984 (Hay, 1992). *Lazarum mirabile* resembles *Sauromatum*, *Typhonium*, and *Biarum*. It differs from these genera in its connate spathe tube with two chambers divided by an annular septum. After the discovery of another species from Darwin, *T. praetermissum*, with spathe tube characters intermediate between *Lazarum* mirabile and "typical" *Typhonium*, *L. mirabile* was transferred to *Typhonium* (Hay, 1997).

Here we test the monophyly and relationships of *Typhonium* based on a dense species sampling consisting of 86 species of Areae selected to represent all geographically and morphologically distinct groups of *Typhonium* and its relatives; we specifically included most of the species ever placed in *Sauromatum*. It turns out that the *Typhonium* in its broad circumscription is a polyphyletic group composed of three distinct clades. We present the phylogenetic results and formalize

some of the nomenclatural changes required by our findings. We also report new chromosome numbers and discuss the implications of the molecular phylogeny for the evolution of *Sauromatum* karyotypes. Finally, we present a key to the species of our redefined *Sauromatum*.

3.3 Materials and Methods

Taxon Sampling and Sequencing

We sampled 86 of the 153 species of Areae. The Areae comprise *Arum* with 29 species (Boyce, 1993, 2006; Lobin et al., 2007), *Biarum* with 21 (Boyce, 2008), *Dracunculus* with two (Mayo et al., 1997), *Eminium* with nine (Mayo et al., 1997; Bogner and Boyce, 2008), *Helicodiceros* with one (Mayo et al., 1997), *Theriophonum* with five (Sivadasan and Nicolson, 1982; Mayo et al., 1997), *Typhonium* s.l. with 72 species (Hay, 1993; Sriboonma et al., 1994; Hay and Taylor, 1996; Hay, 1997; Sookchaloem and Murata, 1997; Hetterscheid and Boyce, 2000; Hetterscheid and Nguyen, 2001; Hetterscheid et al., 2001; Murata et al., 2002; Wang et al., 2002; Hetterscheid and Galloway, 2006; Dao et al., 2007; Nguyen, 2008), however, this number includes species here shown to belong in *Sauromatum* or the Australian clade), and at least 14 as yet undescribed species belonging to various of these genera. For this study, we included 52 species of *Typhonium* s.l., 18 of *Arum*, 9 of *Biarum*, 2 each of *Eminium*, and *Theriophonum*, both *Dracunculus* species, and the single species of *Helicodiceros*. Several species of *Typhonium* are represented by two or three accessions. As outgroups, we used two species of Arisaema based on the chloroplast phylogeny of Renner and Zhang (2004). This totals to 88 species, represented by 98 accessions. All species with their sources and herbarium vouchers (where applicable) are shown in the Appendix.

To deduce phylogenetic relationships, we relied on a nuclear locus, the phytochrome C gene (*PhyC*), and two chloroplast loci, the *rpl20-rps12* intergenic spacer and the tRNA^{Lys} gene (*trnK*), which contains a group I intron that encodes the maturase K (*matK*) open reading frame. Total DNA from silica-dried leaves was extracted with the NucleoSpin plant kit according to the manufacturer's protocol (Macherey-Nagel, Düren, Germany). Sequencing of the circa 2300 nucleotide (nt)-long *trnK* marker, amplified in one piece with the primer pair *trnK*-3914F

(dicot) – *trnK*-16R (Johnson and Soltis, 1994), was problematic. Consequently, we designed new internal primers and amplified the section in two pieces: *trnK*-3914F – *trnK*-RM and *trnK*-FM – *trnK*-16R. Primer sequences are as follows:
trnK-RM 5'-AAGATGTTGATCGTAAATAAGAGG-3' and
trnK-FM 5'-GTTTTGCTGTCATTATGGAAATTCC-3'.
PhyC was also amplified in two pieces with the newly designed primers: A20F – 750R and 430F – AR: A20F: 5'-CACTCAATCCTACAAACTGGC-3',
750R: 5'-AAGATCCATAACATTTGGTGATTGT-3',
430F: 5'-CTCGTGATGTCTGTCACAATAAG-3' and
AR: 5'-GAATAGCATCCATTTCAACATC-3'.
The *rpl20 – rps12* intergenic spacer was amplified using the primers and PCR conditions described in Renner and Zhang (2004).

Polymerase chain reactions (PCR) were performed with 10 μM primers in 25-μl reactions, using BioTherm DNA polymerase (Genecraft, Lüdinghausen, Germany). The initial step of 5 min at 95 °C was followed by 35 cycles of 95 °C for 30 s for DNA denaturation, 60 °C for 60 s for primer annealing, and 72 °C for 2 min and 40 s for primer extension. PCR products were controlled by electrophoresis on an ethidium bromide-stained 1% agarose gel with the Lambda DNA size marker. PCR products were controlled by electrophoresis on an ethidium bromide-stained 1% agarose gel with the Lambda DNA size marker. PCR products were purified using either Promega Wizard®SV Gel and PCR Clean-Up System or Agencourt AMPure®PCR purification kit and quantified electrophoretically, using Lambda DNA as standard. If multiple bands were detected, an additional electrophoresis was performed to excise and analyze them separately. Sequencing relied on Big Dye Terminator kits (Applied Biosystems, Warrington, UK) and the amplification primers. The cycle sequencing products were cleaned by Sephadex G-50 Superfine gel filtration (Amersham, Uppsala, Sweden) on MultiScreen TM-HV membrane plates (Millipore, Bedford, USA) according to the manufacturers' protocols to remove unincorporated nucleotides. Fragments were separated on an ABI 3100 Avant capillary sequencer, assembled and edited using the software Sequencher (Gene Codes, Ann Arbor, MI, USA), and BLAST-searched in GenBank. Sequences are deposited in GenBank (for accession numbers see Appendix).

Alignments and Phylogenetic Analyses

Alignments (Table 1) were generated in MacClade (Maddison and Maddison, 1992) and continuously adjusted manually. The three data partitions were first analyzed separately, and in the absence of statistically supported topological contradictions (>80%), they were then combined.

Phylogenetic inference of the combined data (4319 aligned nucleotides) relied on maximum likelihood (ML) as implemented in the RAxML BlackBox (Stamatakis et al., 2008, http://phylobench.vital-it.ch/raxml-bb/) and on Bayesian analysis as implemented in MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). Bootstrapping under ML used 1000 replicates performed in RAxML. All searches relied on the GTR + Γ model, with model parameters estimated during runs. (RAxML uses the GTRCAT approximation of the GTR + Γ model, with the gamma shape parameter having 25 rate categories.)

Bayesian runs were started from independent random starting trees and repeated four times. Markov chain Monte Carlo (mcmc) runs extended for 2 million generations, with trees sampled every 2000th generations (resulting in 1001 trees for each run). 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. Convergence was assessed by checking that (1) final likelihoods and majority rule topologies among runs were similar; (2) the standard deviations (SD) of split frequencies were < 0.01; (3) the convergence diagnostic (the potential scale reduction factor given by MrBayes) approached 1; and (4) by examining the plot provided by MrBayes of the generation number versus the log probability of the data. TRACER (Rambaut and Drummond, 2007) was used to assess whether runs had reached convergence. Trees saved prior to convergence were discarded as burn-in (100 trees) and a consensus tree constructed from the remaining 3604 trees.

Chromosome Counts

Chromosome numbers were obtained for *S. tentaculatum* (living plant number H.AR.042) and *S. hirsutum* (H.AR.036) from individuals cultivated by the third author in the botanical garden of Wageningen. Offspring of the two species are now (2009) in cultivation at the Munich Botanical Garden. Each count is based

on ten to 15 nuclei from one individual. To obtain good-quality chromosomal spreads, root tip meristems were collected in the morning, pretreated for 3-5 hours with colchicine at 4 °C, fixed with ethanol- acetic acid (3:1), and stored at −20 °C until use. For basic karyotype assessment, hydrolyzed meristems were stained with Schiff"s reagent, squashed under a cover slip, analyzed under a light microscope, and documented using a digital image capture system.

3.4 Results

Phylogeny

The combined sequence matrix (4319 characters, 98 accessions, Table 3.1) yielded a well-supported Areae clade in which all genera with more than one species were monophyletic, except for *Typhonium*, which split into three well-supported clades (Fig. 3.1). The largest of these, sister to the remaining genera, contains 31 of the 52 sampled species of *Typhonium*, including the type *T. trilobatum*. The second "*Typhonium*" clade includes only Australian endemics (Fig. 3.1); we refer to this clade as the Australian clade. Within the Australian clade, there are two statistically supported subclades, one containing *T. alismifolium, T. wilbertii* and *T. angustilobium* (the latter not monophyletic), one containing *T. nudibaccatum, T. praetermissum* and the undescribed species *T.* sp. Kununurra, *T.* sp. Prince Regent and *T.* sp. Morgan River. The phylogenetic positions of other Australian taxa are not well resolved. The next branching clade is *Theriophonum*, followed by the third "*Typhonium*" clade, which includes all species at one time placed in, or morphologically similar to, *Sauromatum* (Fig. 3.1). Within *Sauromatum*, there

Table 3.1: Sizes of the individual and combined chloroplast and nuclear data matrices.

	Aligned nucleotides			**Accessions**		
DNA locus	**Total**	**Excluded**	**Included**	**Total**	**Ingroup**	**Outgroup**
trnK	2719	169	**2550**	**94**	92	2
rpl20-rps12	872	122	**750**	**58**	56	2
phyC	1192	173	**1019**	**51**	49	2
Combined data			**4319**	**98**	96	2

are three statistically supported subclades, one of *S. hirsutum* and *S. tentaculatum*, one of *S. brevipes* and *S. venosum*, and one comprising *S. diversifolium, S. gaoligongense*, and *S. horsfieldii*. The sister relationship between the latter two is also well supported. The entire *Sauromatum* clade is sister to a well-supported clade of the remaining five genera of Areae, which are centered in the Mediterranean region.

Chromosome Numbers

For *Sauromatum*, we obtained new counts of $2n = 26$ for *S. hirsutum* and *S. tentaculatum*. Together with the other counts available for the genus (Fig. 3.2 and Discussion), a base number of $x = 13$ can be inferred, under the assumption that higher numbers in the genus are tetraploid or octoploid.

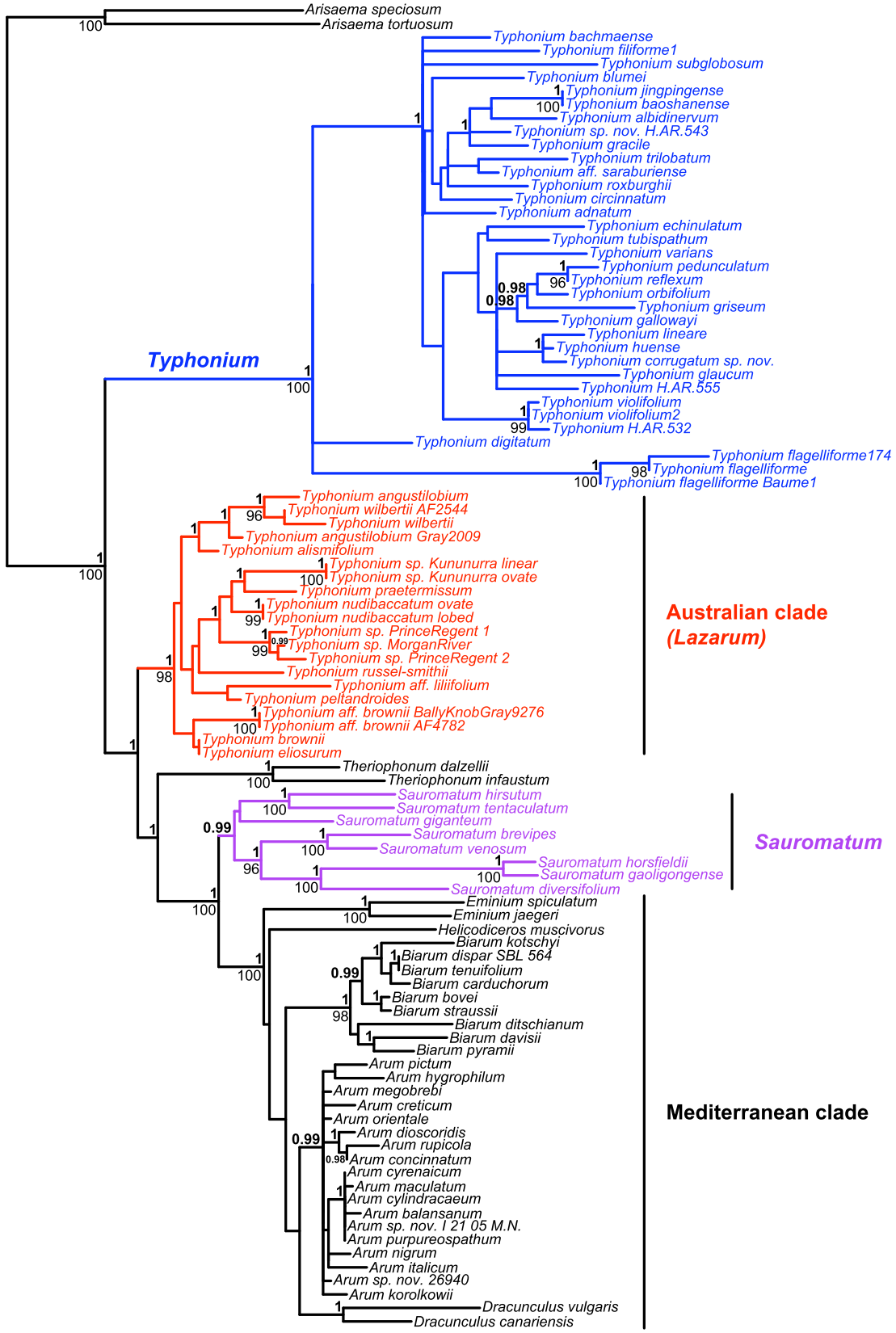
3.5 Discussion

Phylogeny

This study elucidates the phylogenetic relationships among *Typhonium* based on nuclear and chloroplast sequence data obtained for 86 of the 153 species of Areae. The results show that *Typhonium* splits into three clades that are not closely related. Fortunately, the type species, *Typhonium trilobatum* (L.) Schott, falls into the largest clade, so that only a small number of names need to be transferred to achieve monophyletic genera (see Taxonomic Conclusions).

The Australian species all fall in a clade (Fig. 3.1) except for *T. flagelliforme*, which is not endemic to Australia but also occurs in India and Southeast Asia (*T. blumei* and *T. roxburghii* are introduced to Australia). The arrival of *T. flagelliforme* in Australia appears to be quite recent, whereas the Australian clade is a more ancient lineage that has diversified within Australia. This clade will

Figure 3.1 ***(facing page)*:** Maximum likelihood phylogeny for 96 accessions representing 86 species of the nine genera of Areae based on nuclear and chloroplast sequences (4319 aligned nucleotides). Values above branches refer to posterior probabilities from Bayesian inference (2 Mio. generations), those below branches to bootstrap support (percentages of 1000 replicates) under maximum likelihood.


Arisaema speciosum
Arisaema tortuosum
Typhonium bachmaense
Typhonium filiforme1
Typhonium subglobosum
Typhonium blumei
Typhonium jingpingense
Typhonium baoshanense
Typhonium albidinervum
Typhonium sp. nov. H.AR.543
Typhonium gracile
Typhonium trilobatum
Typhonium aff. saraburiense
Typhonium roxburghii
Typhonium circinnatum
Typhonium adnatum
Typhonium echinulatum
Typhonium tubispathum
Typhonium varians
Typhonium pedunculatum
Typhonium reflexum
Typhonium orbifolium
Typhonium griseum
Typhonium gallowayi
Typhonium lineare
Typhonium huense
Typhonium corrugatum sp. nov.
Typhonium glaucum
Typhonium H.AR.555
Typhonium violifolium
Typhonium violifolium2
Typhonium H.AR.532
Typhonium digitatum
Typhonium flagelliforme174
Typhonium flagelliforme
Typhonium flagelliforme Baume1
Typhonium
Typhonium angustilobium
Typhonium wilbertii AF2544
Typhonium wilbertii
Typhonium angustilobium Gray2009
Typhonium alismifolium
Typhonium sp. Kununurra linear
Typhonium sp. Kununurra ovate
Typhonium praetermissum
Typhonium nudibaccatum ovate
Typhonium nudibaccatum lobed
Typhonium sp. PrinceRegent 1
Typhonium sp. MorganRiver
Typhonium sp. PrinceRegent 2
Typhonium russel-smithii
Typhonium aff. liliifolium
Typhonium peltandroides
Typhonium aff. brownii BallyKnobGray9276
Typhonium aff. brownii AF4782
Typhonium brownii
Typhonium eliosurum
Australian clade
(Lazarum)
Theriophonum dalzellii
Theriophonum infaustum
Sauromatum hirsutum
Sauromatum tentaculatum
Sauromatum giganteum
Sauromatum brevipes
Sauromatum venosum
Sauromatum horsfieldii
Sauromatum gaoligongense
Sauromatum diversifolium
Sauromatum
Eminium spiculatum
Eminium jaegeri
Helicodiceros muscivorus
Biarum kotschyi
Biarum dispar SBL 564
Biarum tenuifolium
Biarum carduchorum
Biarum bovei
Biarum straussii
Biarum ditschianum
Biarum davisii
Biarum pyramii
Arum pictum
Arum hygrophilum
Arum megobrebi
Arum creticum
Arum orientale
Arum dioscoridis
Arum rupicola
Arum concinnatum
Arum cyrenaicum
Arum maculatum
Arum cylindracaeum
Arum balansanum
Arum sp. nov. I 21 05 M.N.
Arum purpureospathum
Arum nigrum
Arum italicum
Arum sp. nov. 26940
Arum korolkowii
Dracunculus vulgaris
Dracunculus canariensis
Mediterranean clade

need to be accorded genus rank to achieve a balanced classification of Areae, and if *Lazarum mirabile* A. Hay (*Typhonium mirabile* (A. Hay) A. Hay) turns out to belong to this clade, *Lazarum* will become the correct name for this ninth genus of Areae. A revised morphological definition of *Lazarum* will be necessary, because the characters initially used to define *Lazarum* (Hay, 1997), viz., the connate spathe base, marcescent spathe (meaning that the withered spathe persists on the plant), annular septum at the spathe constriction, and shoot architecture are shared in only three species, *T. mirabile, T. praetermissum*, and *T. taylori* (Hay, 1997). An undescribed species with some of these characters is *T.* (*Lazarum*) sp. Kununurra, which however lacks the annular septum. *T. praetermissum* and *T.* sp. Kununurra are deeply nested within the Australian clade (Fig. 3.1), suggesting that these characters evolved independently within the Australian clade, rather than indicating a relationship with other genera, such as *Sauromatum*, as initially suggested by (Hay, 1992). Another six Australian species currently placed in *Typhonium*, viz. *T. cochleare* A. Hay, *T. johnsonianum* A. Hay & S. M. Taylor, *T. jonesii* A. Hay, *T. mirabile* A. Hay, *T. taylorii* A. Hay, and *T. weipanum* A. Hay, remain to be sequenced.

Within the Australian clade, species with marcescent spathe bases were recovered as a monophyletic group with strong statistical support (*T.* sp. Kununurra, *T. praetermissum*, *T. nudibaccatum*, *T.* sp. Prince Regent, *T.* sp. Morgan River). This clade is nearly equivalent to the "*nudibaccati* group" of Hay (1993), with the inclusion of several species discovered since that publication (Hay, 1997). The expanded "*nudibaccati* group" is restricted to the Kimberley region of West Australia and the northern part of the Northern Territory. The remainder of the Australian species falls into several groups showing geographic structuring, with variable levels of support. A strongly supported clade containing *T. angustilobium, T. wilbertii* and *T. alismifolium* appears restricted to tropical Queensland. Specimens from the Northern Territory previously assigned to *T. angustilobium* requires molecular confirmation of its identity, but could not be sampled for this study. Collections from central arid Australia attributed to *T. alismifolium* in Hay (1993) belong to an undescribed species.

T. peltandroides and *T.* aff. *liliifolium* are sister taxa in Fig. 3.1, but without strong support. Both species occur in the Kimberley region of Western Australia and are may be related to *T. liliifolium* s.str. from the Northern Territory because

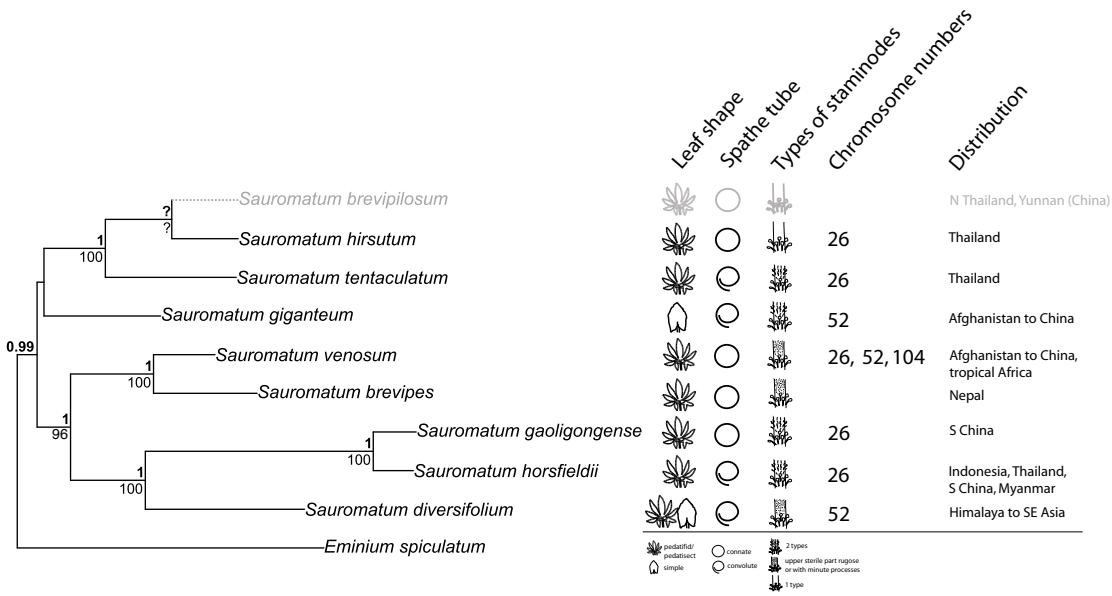


Figure 3.2: Evolution of selected traits in *Sauromatum* as inferred on the relevant part of the Areae phylogeny shown in Figure 3.1. The position of *S. brevipilosum* (grey) is inferred from morphology.

all three species share entire leaves with dense venation. *Typhonium eliosurum, T. browni* and *T.* aff. *brownii* also form a clade in Fig. 3.1, albeit without support. These species are distributed along the east coast of Australia and share laterally elongated rather than depressed-globose corms as present in most other Australian taxa.

The newly revealed *Sauromatum* clade includes six species that were at one time placed in, or thought similar to, *Sauromatum*. Only *S. diversifolium* and *S. tentaculatum* where never before compared to *Sauromatum*, but their morphology fits the genus well (below). Additionally, both species have a chromosome number based on $x = 13$, which appears to be the base number of *Sauromatum*, a characteristic of the genus first reported here.

Sauromatum can be separated from all other Areae genera by a combination of four characters (Fig. 3.2; Fig. 3.3 shows the inflorescences of all nine species of *Sauromatum*): *Sauromatum* has (1) pedatisect leaves; (2) a spathe tube with fused margins, (3) clavate lower staminodes; and (4) upper staminodes that are differently shaped from the lower ones and/or longitudinal ridges on the spadix between the lower staminodes and the stamens. The ridges on the spadix may

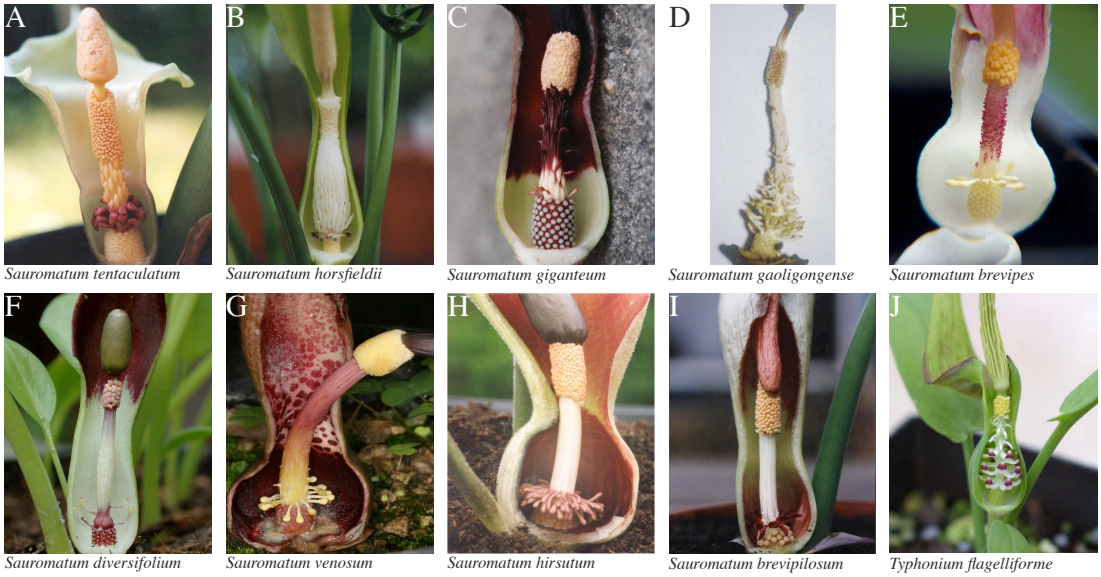


Figure 3.3: Longitudinally-opened spathes of the nine species of *Sauromatum* (A-I) and of *T. flagelliforme* (J), showing the different types of staminodes. Photos A-C, E-F, and H-J by W. Hetterscheid from plants cultivated in his greenhouse; photo D by H. Li, Kunming Institute of Botany; photo E by Arno Clement; photo G by B.O. Schlumpberger, Systematic Botany, University of Munich.

represent vestigial staminode bases, as is visible in *S. giganteum* (Fig. 3.3C). When the last two characters apply and a species has pedatisect leaves and/or a fused spathe tube, then it belongs in *Sauromatum*. On these morphological grounds, we here also transfer *T. brevipilosum* (as yet unsequenced) to *Sauromatum*. Two other species that Sriboonma et al. (1994) revealed as closely related to the investigated *Sauromatum* species are *T. omeiense* H. Li and *T. alpinum* C.Y. Wu ex H. Li, Y. Shiao & S.L. Tseng. However, these names will be synonymized with, respectively, *S. horsfieldii* and *S. diversifolium* in a forthcoming treatment for the *Flora of China* (of which WH is a coauthor).

As here circumscribed, *Sauromatum* has 5-13 foliate pedatisect leaves, except for *S. giganteum* and some forms of *S. diversifolium*, which have simple leaves. *S. brevipes, S. brevipilosum, S. hirsutum, S. gaoligongense* and *S. venosum* have a a spathe tube with fused margins. This character is never found within *Typhonium* (s.str.) but occurs in its sister clade (Fig. 3.1), namely in *Biarum, Eminium jaegeri, T. mirabile, T. taylori, T. praetermissum*, and the undescribed *T.* sp.

Kununurra. The only *Typhonium* s.str. species that has two types of staminodes is *T. flagelliforme* (Fig. 3.3J), but here, the lower ones are not clavate, nor are the leaves pedatisect or the spathe tube fused.

The phylogenetic relationships found here (Figs. 3.1, 3.2) suggest that the fused spathe tube has evolved three times independently in *Sauromatum* (once in *S. hirsutum/S. brevipilosum*, once in *S. venosum/S. brevipes*, and once in *S. gaoligongense*). In contrast, the clavate-shaped lower staminodes seem to have evolved once, along with the distinct upper staminodes (Fig. 3.3A, B), later reduced to varying extents (Fig. 3.3C-I). If *S. brevipilosum* is indeed the sister of *S. hirsutum* (Fig. 3.2), the complete loss of the upper staminodes happened once, their reduction twice (in *S. venosum/S. brevipes* and *S. diversifolium*). The pedatisect leaves seem to be the plesiomorphic condition in the genus that has been lost twice (*S. giganteum* and *S. diversifolium*).

Chromosome Numbers in *Typhonium*, *Sauromatum*, and the Australian clade

In *Sauromatum* chromosome numbers are available for *S. horsfieldii* ($2n = 26$), *S. diversifolium, S. giganteum* ($2n = 52$), *S. venosum* ($2n = 26$, 52, 104; Petersen, 1989; Bogner and Petersen, 2007), and *S. gaoligongense* ($2n = 26$; Li Heng, Kunming Institute of Botany, pers. comm. on 20 July 2009). Hence, including our new counts for *S. hirsutum* and *S. tentaculatum* of $2n = 26$, the base chromosome number of Sauromatum is $x = 13$. From these counts, the two most widespread species, *S. giganteum* (Afghanistan to China) and *S. venosum* (Afghanistan to China, Africa) appear to be tetraploid, and *S. venosum* appears to comprise also diploid and octoploid forms. The five remaining, more narrowly distributed species all have $2n = 26$, viz. *S. tentaculatum* (Thailand), *S. gaoligongense* (S China), *S. hirsutum* (Thailand) and *S. horsfieldii* (Indonesia, Thailand, S China, Myanmar). The switch to polyploidy appears to have happened three times, but chromosome numbers are still lacking for two species (*S. brevipilosum* and *S. brevipes*).

In *Typhonium*, by contrast, chromosome numbers are extremely variable: the early-diverging *T. flagelliforme* has $2n = 16$, while other species have $x = 5$, 7, 8, 9, 10 or 13 and most probably also $x = 4$, 6, 11 (N. Cusimano, unpublished data). Diploid chromosome numbers range from $2n = 8$ to $2n = 78$. Aneuploidization

and subsequent polyploidization events (or vice versa), or other complex chromosome rearrangements, seem to have played an important role in the evolution of the *Typhonium*, leading to the existing variety of chromosome numbers. The drastic reduction of chromosome number could have happened through chromosome fusion, translocations of chromosome parts and/or loss of DNA.

For the Australian clade, only two chromosome numbers have been reported, both extremely high: *T. eliosurum* ($2n > 100$) and *T. brownii* ($2n = 160$; Briggs in Evans, 1961). Along with an undescribed species from New South Wales, these species may belong to a polyploid complex.

Taxonomic Conclusions

Our results show that the broadly circumscribed *Typhonium* of Hetterscheid and Boyce (2000) is polyphyletic and comprises three distinct clades, which should be recognized at generic rank to achieve a balanced classification of the Areae. The Asian and Malesian species form a clade (*Typhonium* s.str.) that is sister to the other genera of Areae. The Australian species of *Typhonium* so far sequenced form a distinct clade. *Theriophonum*, from southern India, diverges next, followed by *Sauromatum*, now composed of nine species, and sister to a clade including the Mediterranean genera of Areae (*Arum, Biarum, Dracunculus, Eminium, Helicodiceros*). The genus *Sauromatum* can be circumscribed not only genetically, but also morphologically as we have shown here. Application of the name *Lazarum* for the Australian clade (Fig. 3.1) awaits the sequencing of the type species of this generic name. At this stage, therefore, *Typhonium* s.str. and the Australian clade are not defined morphologically, but on our molecular evidence are clearly distinct.

In the following, we resurrect the genus *Sauromatum*, make the necessary five new combinations, and present a key to the nine species of *Sauromatum*. Detailed descriptions of all species and up-to-date information on their geographic distribution are available elsewhere (Sriboonma & al., 1994; Wang and Li, 1999; Hetterscheid & Boyce, 2000; Hetterscheid & al., 2001).

Sauromatum SCHOTT in H.W. Schott et Endlicher, Melet. Bot.: 17. 1832 – Lectotype (designated by D. Nicolson in Taxon 16: 518. 1967): *S. guttatum* SCHOTT (*Arum guttatum* WALLICH 1831, NON SALISBURY 1796).

Sauromatum brevipes (HOOK.F.) N.E. BROWN in Gard. Chron. 3, 34(2): 93. 1903 ≡ *Typhonium brevipes* HOOK.F. in Fl. Brit. India 6: 511. 1893 – Syntypes: Darjeeling, 7500 ft., Clarke 26708 (K); Jore Pokri, 7600 ft., Gammie s.n. (K).

Sauromatum gaoligongense Z.L. WANG & H. LI in Acta Bot. Yunnan, suppl. 11: 61. 1999 ≡ *Typhonium gaoligongense* (Z.L. WANG & H. LI) HETT. & P.C. BOYCE in Aroideana 23: 51. 2000 – Holotype: China, Yunnan prov., Baoshan Xianm, Li Heng & G. Ruckert 11309A (KUN).

Sauromatum horsfieldii MIQ. in Fl. Ned. Ind. 3: 196. 1855 ≡ *Typhonium horsfieldii* (MIQ.) STEENIS in Bull. Jard. Bot. Buitenzorg Ser. 3(17): 403. 1948 – Holotype: Java, Oenagaran, Horsefield s.n. (K). [For full synonymy see Sriboonma et al., 1994.]

Sauromatum venosum (DRYAND. ex AITON) KUNTH in Enum. Pl. 3: 28. 1841 ≡ *Arum venosum* DRYAND. EX AITON in Hort. Kew. 3: 315. 1789 ≡ *Desmesia venosum* (DRYAND. ex AITON) RAF. in Fl. Tellur. 3: 63. 1837 ≡ *Sauromatum guttatum* (AIT.) SCHOTT var. *venosum* (AIT.) ENGL. in Pflanzenr. IV23F (Heft 73): 125. 1920 ≡ *Typhonium venosum* (DRYAND. ex AITON) HETT. & P.C. BOYCE in Aroideana 23: 51. 2000 – Holotype: Plant of unknown origin introduced into cultivation at Kew by William Malcom in 1774 (BM). [For full synonymy see Hetterscheid and Boyce (2000).]

Sauromatum brevipilosum (HETT. & M. SIZEMORE) CUSIMANO & HETT., **comb. nov.** ≡ *Typhonium brevipilosum* HETT. & M. SIZEMORE in Aroideana 23: 52. 2000 – Holotype: Indonesia, Sumatera, West Sumatera, near Aeksah, Hetterscheid H.AR.097-T (orig. coll. Sizemore s.n.) flowered in cult. in Leiden Bot. Gard., 29 Nov 1999 (L, spirit coll.).

Sauromatum diversifolium (WALL. EX SCHOTT) CUSIMANO & HETT., **comb. nov.** ≡ *Typhonium diversifolium* WALL. in Wallich's Numer. List n. 8933. 1949, nom. nud, ex SCHOTT in Aroideae 13: 20. 1855 ≡ *Heterostalis diversifolia* SCHOTT in Oesterr. Bot. Wochenbl. 7: 267. 1857 – Holotype: Nepal, Wallich's

Numer. List. no. 8933a in 1821 (K). [For full synonymy see Sriboonma et al., 1994.]

Sauromatum hirsutum (S.Y. Hu) Cusimano & Hett., **comb. nov.** ≡ *Arisaema hirsutum* S.Y. Hu in Dansk Bot. Ark. 23(4): 454. 1968 ≡ *Typhonium hirsutum* (S.Y.Hu) J. Murata & Mayo in Kew Bull. 46(1): 129. 1991 – Holotype: Thailand, Payap 3939 (C).

Sauromatum giganteum (Engl.) Cusimano & Hett. **comb. nov.** ≡ *Typhonium giganteum* Engl. in Bot. Jahrb. Syst. 4: 66. 1883 – Holotype: China, Beijing, Skatschkow s.n. (LE). [For full synonymy see Sriboonma et al., 1994.]

Sauromatum tentaculatum (Hett.) Cusimano & Hett., **comb. nov.** Basionym: *Typhonium tentaculatum* Hett. in Aroideana 24: 49. 2001 – Holotype: Thailand SW 37, Kanchanaburi, Sangklaburi Distr., Lai Wo Subdistr., Toong Yai Wildlife Reserve, Ban Saneh Pawng, West side of Paneh limestone mountain, 300m Hetterscheid H.AR.042-T (orig. coll. J. F. Maxwell 93-647), flowered in cult. in the Leiden Bot. Gard. 9. Apr 1997 (BKF, spirit coll.).

Key to *Sauromatum* Species

1 Two types of staminodes, upper clearly of different shape or reduced compared to the lower ones **2**

- Only one type of staminodes, upper sterile part longitudinally grooved and roughened with projections or processes, or naked **5**

2 Spathe tube fused; 3-5 bulbils in the lower part of the petiole ***S. gaoligongense***

- Spathe tube convolute **3**

3 Lower staminodes clearly clavate **4**

- Lower staminodes clavate/spathulate, clavate part anvil shaped or dorso-ventrally flattened, dark purple; upper staminodes narrowly spindle shaped, whitish ***S. tentaculatum***

4 Leaves simple, ovate, cordate to hastate; upper staminodes smaller than the lower ones, subulate or occasionally absent . . . . . ***S. giganteum***

- Leaves pedatisect; upper staminodes filiform, gradually shorter towards the male zone . . . . . ***S. horsfieldii***

5 (1) Spathe tube free; leaf blade simple and ovate-lanceolate, cuneate to hastate

3-5 lobed or 5-9 foliate pedatisect; upper sterile part of the spadix naked or with apiculate projections . . . . . ***S. diversifolium***

\- Spathe tube fused, at least basal part **6**

6 Petiole and leaf blade with hairs; upper sterile part of the spadix axis completely naked **7**

\- Petiole and leaf blade without hairs, upper sterile part of the spadix axis rough, shaggy or with minute processes (highly reduced staminodes) **8**

7 Hairs distinct and long; also outside surface of spathe covered with hairs ***S. hirsutum***

\- Hairs short (ca. 0.5 mm); outside surface of spathe smooth . ***S. brevipilosum***

8 Leaflets oblong-lanceolate, apex acuminate; inflorescence up to 40 cm long; spathe tube dark purple inside, spathe blade inside maculate . . . . . ***S. venosum***

\- Leaflets linear-lanceolate, apex long acuminate; lower clavate staminodes white, upper minute processes purple; inflorescence max. 7.5 cm long; spathe tube inside greenish to white, spathe blade inside not maculate, dull purple basally and pink above . . . . . ***S. brevipes***

Acknowledgements

For plant material, we thank Josef Bogner, Li Heng, Clare Herscovitch, Wolfram Lobin and Allan Galloway. We also thank Josef Bogner and two anonymous reviewers for their comments. This study was supported by DFG grant RE 603/7-1.

3.6 References

Bogner, J. and P. C. Boyce. 2008. *Eminium jaegeri* (Araceae), a new species from northwestern Iran. Willdenowia 38:149–153.

Bogner, J. and G. Petersen. 2007. The chromsome numbers of the aroid genera. Aroideana 30:82–90.

Boyce, P. C. 1993. The genus *Arum*. The Bentham-Moxon Trust, Royal Botanical Gardens, Kew, UK.

Boyce, P. C. 2006. *Arum* - A decade of change. Aroideana 29:132–137.

Boyce, P. C. 2008. A monograph of the genus *Biarum* - special part. Curtis's Botanical Magazine 25:1–119.

Brown, N. E. 1903. New or noteworthy plants – *Sauromatum brevipes*. The Gardener's Chronicle 3:93–94.

Dao, Z.-L., S.-T. Chen, Y.-H. Ji, and H. Li. 2007. *Typhonium baoshanense* Z. L. Dao & H. Li, a new species of Araceae from western Yunnan, China. Acta Phytotaxon. Sin. 45:234–238.

Evans, O. 1961. Araceae. In: Flora of New South Wales. Contrib. New South Wales Nat. Herb. Flora, Series Nos. 21-22:6–13.

French, J. C., M. G. Chung, and Y. K. Hur. 1995. Chloroplast DNA phylogeny of the Ariflorae. Pages 255–275 *in* Monocotyledons: Systematics and Evolution (P. J. Rudall, P. Cribb, D. F. Cutler, and C. J. Humphries, eds.). Royal Botanic Gardens, Kew, England.

Hay, A. 1992. A new Australian genus of Araceae, with notes on generic limits and biogeography of the Areae. Bot. J. Linnean Soc. 109:427–434.

Hay, A. 1993. The genus *Typhonium* (Araceae-Areae) in Australasia. Blumea 37:345–376.

Hay, A. 1997. Two new species and a new combination in Australian *Typhonium* (Araceae tribe Areae). Edinb. J. Bot. 54:329–336.

Hay, A. and S. M. Taylor. 1996. A new species of *Typhonium* Schott (Araceae - Areae) from the Northern Territory, with notes on the conservation status of two Areae endemic to the Tiwi Islands. Telopea 6:563–568.

Hetterscheid, W. L. A. and P. C. Boyce. 2000. A reclassification of *Sauromatum* Schott and new species of *Typhonium* Schott (Araceae). Aroideana 23:48–55.

Hetterscheid, W. L. A. and A. Galloway. 2006. New *Typhonium* (Araceae) species from Thailand. Aroideana 29:80–85.

Hetterscheid, W. L. A. and V. D. Nguyen. 2001. Three new species of *Typhonium* (Araceae) from Vietnam. Aroideana 24:24–29.

Hetterscheid, W. L. A., D. Sookchaloem, and J. Murata. 2001. *Typhonium* (Araceae) of Thailand: New species and a revised key. Aroideana 24:30–55.

Hooker, J. D. 1893. Flora of British India, vol 6: Orchideae to Cyperaceae. L. Reeve & Co., London.

Hu, S. Y. 1968. Araceae. Studies in the Flora of Thailand 41. Dansk. Bot. Ark. 23:411–457.

Huelsenbeck, J. P. and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755.

Johnson, L. A. and D. E. Soltis. 1994. *MatK* DNA sequences and phylogenetic reconstruction in Saxifragaceae s.str. Syst. Bot. 19:143–156.

Kunth, C. S. 1841. Enumeration Plantarum 3. J.G. Cottae, Stuttgart, Tübingen.

Lobin, W., M. Neumann, J. Bogner, and P. C. Boyce. 2007. A new *Arum* species (Areae, Araceae) from NE Turkey and Georgia. Willdenowia 37:445–449.

Maddison, W. P. and D. R. Maddison. 1992. MacClade: analysis of phylogeny and character evolution. Version 3.0. Sinauer, Sunderland, Massachusetts, USA.

Mayo, S. J., J. Bogner, and P. C. Boyce. 1997. The genera of Araceae. The Trustees, Royal Botanical Gardens, Kew, UK.

Miquel, F. A. W. 1855. Flora van Nederlandsch Indie, vol. 3. C.G. van der Post, Amsterdam.

Miquel, F. A. W. 1864. Mantissa Araliacearum, Ericacearum, Cupuliferarum et Aracearum. Ann. Mus. Bot. Lugd. Bat. vol. 1 1:219–221.

Murata, J. and S. J. Mayo. 1991. A new combination in *Typhonium* (Araceae). Kew Bull. 46:129–131.

Murata, J., D. Sookchaloem, and W. L. A. Hetterscheid. 2002. *Typhonium watanabei* (Araceae), a new species from Thailand. J. Jap. Bot. 77:163–166.

Nguyen, V. D. 2008. *Typhonium stigmatilobatum* (Araceae tribe Areae), a new species from Vietnam. Kew Bull. 93:491–493.

Petersen, G. 1989. Cytology and systematics of Araceae. Nord. J. Bot. 9:119–166.

Rambaut, A. and A. J. Drummond. 2007. Tracer v1.4. `http://beast.bio.ed.ac.uk/Tracer` .

Renner, S. S. and L.-B. Zhang. 2004. Biogeography of the *Pistia* clade (Araceae): Based on chloroplast and mitochondrial DNA sequences and Bayesian divergence time inference. Syst. Biol. 53:422–432.

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

Schott, H. W. 1832. Araceae. Pages 16–22 *in* Meletemata Botanica (H. W. Schott and S. Endlicher, eds.) vol. 17. C. Gerold, Vienna.

Sivadasan, M. and D. H. Nicolson. 1982. A revision of *Theriophonum* (Araceae. Kew Bull. 37:277–290.

Sookchaloem, D. and J. Murata. 1997. A new species of *Typhonium* from Thailand. Thai Forest Bull. (Bot.) 25:57–59.

Sriboonma, D., M. Hasebe, N. Murakami, J. Murata, and K. Iwatsuki. 1993. Phylogeny of *Typhonium* (Araceae) inferred from restriction fragment analysis of chloroplast DNA. J. Plant Res. 106:11–14.

Sriboonma, D., J. Murata, and K. Iwatsuki. 1994. A revision of *Typhonium* (Araceae). J. Fac. Sci. Univ. Tokyo, III 14:255–313.

Stamatakis, A., P. Hoover, and J. Rougemont. 2008. A rapid bootstrap algorithm for the RAxML web-servers. Syst. Biol. 57:758–771.

van Steenis, C. G. G. J. 1948. Miscellaneous botanical notes I. Bull. Jard. Bot. Buitenzorg, ser. 3 17:383–411.

Wang, Z., H. Li, and F. Bian. 2002. *Typhonium jinpingense*, a new species from Yunnan, China, with the lowest diploid chromosome number in Araceae. Novon 12:286–289.

Wang, Z.-L. and H. Li. 1999. *Sauromatum gaoligongense* (Araceae), a new species from Gaoling mountains. Acta Bot. Yunnnan suppl. 11:61–64.

Appendix

Voucher information for studied species. GenBank numbers are given for the three markers sequenced, *trnK, PhyC* and *rpl20-rps12*; an n-dash denotes a missing marker.

Outgroups: *Arisaema speciosum* (Wall.) Mart., Hetterscheid H.AR.294 (L, spirit coll.), EU886502, EU886470, AY279168; *Arisaema tortuosum* (Wall.) Schott, Anaimudi 20/5, EU886577, EU886469, AY248920; ***Arum*** L.: *A. balansanum* R. Mill, V. Haller & M. Koenen 1*91*TR H*K (BONN), EU886512, –, EU886624; *A. concinnatum* Schott, B.W. Magrys s.n., cult. 15. Mar. 02, EU886516, –, GU255991; *A. creticum* Boiss. & Heldr., Tillich 4881 (M), EU886504, –, EU886595; *A. cyrenaicum* Hruby, LY-0-BONN-6425 (BG Bonn), EU886515, –, EU886623; *A. dioscoridis* Sibth. & Sm., B. W. Magrys s.n., cult. 15. Mar. 02, EU886505, –, GU255992; *A. hygrophilum* Boiss., CY-0-BONN-6427 (BG Bonn), EU886509, EU886471, EU886620; *A. italicum* Mill., BG

Mainz, cult. 20 Jul 2001, EU886517, EU886472, AY248922; *A. korolkowii* Regel, S. Volz 20 (M), EU886589, –, EU886598; *A. maculatum* L., Cusimano06–3 (M), EU886506, –, EU886593; *A. megobrebi* Lobin, M. Neumann & al. 24219 (BONN), EU886513, –, EU886625; *A. nigrum* Schott, Cusimano06–1 (M), EU886507, EU886473, EU886597; *A. orientale* Bieb., A. Groeger 06/1845w, EU886510, –, EU886621; *A. pictum* L. f., xx-0-BONN-273 (BG Bonn), EU886518, –, EU886596; *A. purpureospathum* Boyce, Tuber from E. Walton, 15 Apr 2002, EU886508, –, EU886594; *A. rupicola* Boiss., J. Bogner 1790 (M), EU886519, –, EU886592; *A.* sp. nov., M. Neumann I 21 05 M.N., EU886514, –, EU886622; *A.* sp. nov., W. Lobin 26940 (BONN), EU886511, –, EU886626; ***Biarum*** Schott: *B. bovei* Blume, T.F. Hewer H7951 (M), EU886529, –, EU886601; *B. carduchorum* (Schott) Engl., M. Jaeger JLMS-60 (BG Giessen), EU886521, EU886478, EU886618; *B. davisii* Turrill, MO living acc. 78231, EU886525, EU886479, AY248923; *B. dispar* (Schott) Talavera, M. Jaeger SBL 564 (BG Giessen), EU886522, –, EU886619; *B. ditschianum* Bogner & Boyce, BG Bonn 4695, rec., EU886526, EU886477, EU886600; *B. kotschyi* (Schott) B. Mathew ex H. Riedl, TR-0-BONN-8431 (BG Bonn), EU886527, –, EU886599; *B. pyramii* (Schott) Engler, J. Mayr s.n. (BG Gies- sen), EU886523, –, EU886617; *B. straussii* Engler, M. Jaeger JZZ-54 (BG Giessen), EU886524, –, EU886615; B. tenuifolium (L.) Schott, ES-0-BONN-16014 (BG Bonn), EU886528, –, AY248924; ***Dracunculus*** P. Miller: *D. canariensis* Kunth, ES-0-BONN-13049 (BG Bonn), EU886531, EU886475, AY248926; *D. vulgaris* Schott, T. Croat 78286 (MO), EU886532, EU886476, AY248927; ***Eminium*** (Blume) Schott: *E. jaegeri* Bogner & P.C. Boyce, M. Jaeger JJMZ-67a (M), EU886520, –, EU886616; *E. spiculatum* (Blume) Schott, M. Neumann 27/96 (BONN), EU886530, EU886474, AY248928; ***Helicodiceros*** Schott: *H. muscivorus* (L. f.) Engl., MO living acc. 71821, EU886533, EU886480, AY248929; ***Sauromatum*** Schott: *S. brevipes* (Hook. f.) N.E. Brown, J. McClements cult., EU886539, EU886484, EU886608; *S. diversifolium* (Wall.) Cusimano & Hett., Hetterscheid H.AR.484 (L, spirit coll.), EU886540, EU886482, EU886605; *S. gaoligongense* Wang & H. Li, Chen YM 024 (KUN), EU886590, EU886487, –; *S. giganteum* (Engl.) Cusimano & Hett., J.W. Waddick cult., 20 Aug. 2001, EU886536, EU886490, AY248938; *S. hirsutum* (S.Y. Hu) Cusimano & Hett., Hetterscheid H.AR.036 (L, spirit coll.), EU886542, EU886489, AY248939; *S. horsfieldii* Miq., J. Murata 3 (TI), EU886541, EU886483, EU886604; *S. tentaculatum* (Hett.) Cusimano & Hett., Hetterscheid H.AR.042 (L, spirit coll.), EU886543, EU886488, EU886612; *S. venosum* (Dryand. ex Ait.) Kunth, J. Bogner s.n. (M), 27. Jun. 02, EU886544, EU886481, EU886603; ***Theriophonum*** Blume: *T. dalzelii* Schott, J. Murata s.n., 21 Aug. 2002, –, –, AY248936; *T. dalzelii*, P. Bruggemann PB 168, cult., EU886534, EU886486, –; *T. infaustum* N.E. Br., P. Bruggemann PB 099, cult., EU886535, EU886485, EU886602; ***Typhonium*** Schott: *T. adnatum* Hett. & Sookchaloem, A. Galloway AGA-1095-17, EU886547, –, –; *T. albidinervum* C.Z. Tang & H. Li, J. Murata 1 (TI), EU886548, EU886497, AY248937; *T. bachmaense* Nguyen Van Dzu & Hett., Nguyen Van Dzu 185 (HN), EU886549, –, –; *T. baoshanense* Z.L. Dao & H. Li, Chen YM 017 (KUN), EU886591, –, EU886629; *T. blumei* Nicolson & Sivadasan, G. Hausner 5 (M), EU886553, –, –; *T. circinnatum* Hett. & J. Mood, Hetterscheid H.AR.258 (L, spirit coll.), EU886551, –, –; *T. corrugatum* sp. nov, Bogner 2962 (M), GU255984, –, –; *T. digitatum* Hett. & Sookchaloem, Hetterscheid H.AR.215 (L, spirit coll.), EU886552, –, –; *T. echinulatum* Hett. & Sookchaloem, Hetterscheid H.AR.225 (L, spirit coll.), EU886554, EU886499, –; *T. filiforme* Ridl., Hetterscheid H.AR.128 (L, spirit coll.), EU886555, –, –; *T. flagelliforme* (Lodd.) Blume, SE Asia, Hetterscheid H.AR.028 (L, spirit coll.), EU886556, –, –; *T. flagelliforme* (Lodd.) Blume, Cape York, QLD, Australia, Baume 1 (CNS), –, GU255955, –; *T. flagelliforme* (Lodd.) Blume, Michel CR 2016 (DNA), GU255983, –, –; *T. gallowayi* Hett. & Sookchaloem, A. Galloway AGA-0516-01, EU886558, –, –; *T. glaucum* Hett. & Sookchaloem, Hetterscheid H.AR.535 (L, spirit coll.), EU886559, –, –; *T. gracile* (Roxb.) Schott, J. Murata 2 (TI), EU886563, EU886495, –; *T. griseum* Hett. & Sookchaloem, Hetterscheid H.AR.044

(L, spirit coll.), EU886561, –, –; *T. huense* V.D. Nguyen & Croat, Hetterscheid H.AR.178 (L, spirit coll.), EU886557, –, –; *T. jingpingense* Z.L. Wang, H. Li & F.H. Bian, Chen YM 023 (KUN), EU886564, EU886498, EU886614; *T. lineare* Hett. & V.D. Nguyen, Hetterscheid H.AR.244 (L, spirit coll.), EU886565, –, –; *T. orbifolium* Hett. & Sookchaloem, Hetterscheid H.AR.227 (L, spirit coll.), EU886566, –, –; *T. pedunculatum* Hett. & Sookchaloem, Hetterscheid H.AR.559 (L, spirit coll.), EU886567, –, –; *T. reflexum* Hett. & Sookchaloem, A. Galloway AGA-1547-01, EU886568, –, –; *T. roxburghii* Schott, J.C. Wang 11621 (TNU), EU886569, –, EU886613; *T. saraburiense*? Sookchaloem, Hett. & Murata, Hetterscheid H.AR.538 (L, spirit coll.), EU886570, –, –; *T. sp.* nov., Hetterscheid H.AR.555 (L, spirit coll.), EU886550, –, –; *T. sp.* nov., Hetterscheid H.AR.532 (L, spirit coll.), EU886572, –, –; *T. sp.* nov., Hetterscheid H.AR.543 (L, spirit coll.), EU886573, –, –; *T. subglobosum* Hett. & Sookchaloem, A. Galloway AGA-1006-01, EU886546, –, –; *T. trilobatum* (L.) Schott, J. Murata 5 (TI), EU886571, EU886496, AY248941; *T. tubispathum* Hett. & A.Galloway, Hetterscheid H.AR.469 (L, spirit coll.), EU886574, –, –; *T. varians* Hett. & Sookchaloem, Hetterscheid H.AR.560 (L, spirit coll.), EU886575, EU886494, –; *T. violifolium* Gagnep., Hetterscheid H.AR.461 (L, spirit coll.), EU886560, –, –; *T. violifolium* 2 Gagnep., Hetterscheid H.AR.168 (L, spirit coll.), EU886562, –, EU886611; Australian Typhonium (Lazarum A. Hay): *T. alismifolium* F. Muell. s.str., B. Gray 9146 (CNS), GU255975, GU255961, –; *T. angustilobium* F. Muell., B. Gray 9277 (CNS), GU255974, GU255960, –; *T. angustilobium* F. Muell., H. Schaefer 2007/32 (M), EU886576, EU886491, EU886609; *T. brownii* Schott, Hetterscheid H.AR.043 (L, spirit coll.), EU886538, EU886492, EU886607; *T. eliosurum* (F. Muell. ex Benth.) O.D. Evans, Hetterscheid H.AR.364 (L, spirit coll.), EU886537, EU886493, EU886606; *T. nudibaccatum* A. Hay (linear), R.L. Barrett 3957 (PERTH), GU255981, GU255968, –; *T. nudibaccatum* A. Hay (ovate), R.L. Barrett 3957 (PERTH), –, GU255969, GU255990; *T. peltandroides* A. Hay, M.D. Barrett & R.L. Barrett, M.D. Barrett 599 (PERTH), GU255973, GU255958, GU255986; *T. praetermissum* A. Hay, Hay s.n. 16.10.1996 (NSW), GU255982, GU255970, –; *T. russell*-smithii A. Hay, I. Cowie 104311 (DNA), GU255985, –, –; *T. sp.* aff. brownii Schott, B. Gray 9276 (CNS), GU255971, GU255956, –; *T. sp.* aff. brownii Schott, A. Ford 4782 (CNS), GU255972, GU255957, –; *T. sp.* aff. liliifolium sp. Theda, M.D. Barrett & R.L. Barrett MDB 1504 (PERTH), –, GU255959, GU255987; *T. sp.* Kununurra, M.D. Barrett & R.L. Barrett MDB 2264 (PERTH), –, GU255965, –; *T. sp.* Kununurra, R.L. Barrett 3069 (PERTH), GU255979, GU255966, GU255989; *T. sp.* Morgan River, M.D. Barrett & R.L. Barrett MDB 2265 (PERTH), GU255980, GU255967, –; *T. sp.* Prince Regent, R.L. Barrett & M.D. Barrett RLB 1716 (PERTH), GU255977, GU255963, GU255988; *T. sp.* Prince Regent, M.D. Barrett 1033 (PERTH), GU255978, GU255964, –; *T. wilbertii* A. Hay, A. Ford 2544 (CNS), GU255976, GU255962, –; *T. wilbertii* A. Hay, Hetterscheid H.AR.033 (L, spirit coll.), EU886545, –, EU886610.

Chapter 4

The handling of missing species in diversification analyses – with empirical examples illustrating a new approach

Cusimano, N. and S. S. Renner.
In review at *Systematic Biology*.

4.1 Abstract

Chronograms from molecular dating are increasingly being used to infer rates of diversification and their change over time. A major limitation in such analyses is incomplete species sampling because estimators of diversification rates assume the complete sampling of all extant species. A review of plant diversification studies shows that most are based on incomplete phylogenies and that the handling of missing species is highly inconsistent. Here we use two non-nested clades of Araceae with a species sampling of 57% (*Typhonium*) and 48% (*Arum* clade) to study different approaches for handling missing species and to illustrate a new approach, CorExS, which involves simulating missing splits under an exponential model, including information about their ages where available. Specifically, we apply the γ statistic and MCCR test, birth-death-likelihood analysis with the ΔAIC_{rc} test, survival analysis, and CorExS correction, and evaluate the advantages and disadvantages of each method. A weakness of the MCCR and ΔAIC_{rc} test is that simulating trees and randomly pruning them to the sample size (the number of species sequenced out of all those comprising the clade) does not create a proper null distribution because species sequenced for phylogenies generally are not sampled randomly. The CorExS approach has the advantages of not requiring detailed phylogenetic knowledge about the placements of missing species and of creating a proper null distribution, which has the effect of reducing type I error. Additionally, it provides information about the impact of missing data on parameter estimation. It does, however, require an *a priori* diversification rate, the minimal value of which is constrained by the clade age and extant species number. For the *Typhonium* and *Arum* examples, the CorExS method yielded different inferences than obtained with the MCCR test and BDL analysis.

Key Words

Diversification rates, missing-species-problem, γ statistic, survival analysis, birth-death likelihood analysis, model fitting, tree simulation

4.2 Introduction

Large time-calibrated phylogenies are now readily obtained and are increasingly being used to infer diversification patters (Hey, 1992; Nee et al., 1992; Sanderson and Bharathan, 1993; Sanderson and Donoghue, 1994; Harvey et al., 1994; Paradis, 1997; Baldwin and Sanderson, 1998; Paradis, 1998; Magallón and Sanderson, 2001; Nee, 2006; Rabosky, 2006b; Rabosky et al., 2007; McPeek, 2008; Phillimore and Price, 2008). However, inferring rates of diversification is statistically challenging, and the sensitivity of methods when their underlying assumptions are not met is poorly understood. A major problem in diversification analysis is incomplete species sampling (Pybus and Harvey, 2000, this study). This is a common problem when clades are species-rich and access to samples is problematic and costly. Several methods have been proposed that attempt to correct for the bias introduced by incomplete sampling of a clade's species (Table 4.1). Some of them attempt the correction before the analysis; others attempt correction after the analysis (Nakagawa and Freckleton, 2008, for a review of methods for handling missing data). Of the methods that try to correct for missing (not sequenced) species before the analysis, survival analysis (SA; Paradis, 1997) adds them as censored events. Alternatively, missing species have been added halfway along the branch where they are thought to belong (Barraclough and Vogler, 2002) or to the stem of their clade (Purvis et al., 1995). Another approach is to add missing species to random locations within their clade, using an MCMC tree chain (Day et al., 2008, the legend of figure S1 in this study is misleading; T. Barraclough, Imperial College, personal communication, 18 Aug. 2009). All these *a priori* corrections require knowledge about the relationships of the missing species; censoring moreover requires knowing the missing species' minimum ages. This greatly limits the application of these methods. Another problem is that these approaches do not quantify (statistically) the impact of the missing data on parameter estimation.

Approaches that correct for missing species after the analysis, that is, after diversification models have been fit to the topology/branching times, involve the creation of a null distribution. Thus, one carries out numerous simulations of trees under the Yule model, with the number of tips corresponding to the complete number of species in the focal clade. Trees are then randomly pruned to the sample size (the number of species actually sequenced) and, like the empirical data, tested

Table 4.1: Approaches for handling missing species in diversification analyses.

Approach	Comments	Reference
Adding missing species to a phylogeny where they are thought to belong	• Requires knowledge of species' relationships	
• half way along the branch	• Not statistical (only 1 complete tree)	Barraclough & Vogler (2002)
• randomly along the branches of the clade where they are thought to belong, using a Bayesian MCMC chain of trees	• Creates a pseudoreplication problem because species will be distributed in proportion to nodes' (and branches') presence in the respective part of the Bayesian trees	Day et al. (2008)
Adding missing species as censoring events in survival analyses	• Requires estimate of species' minimum ages • Not statistical (only 1 complete tree)	Paradis (1997, 1998)
Simulating phylogenies with the true number of species and then pruning them to the size of the sample phylogeny to create a null distribution under a constant-rates model	• Requires diversification rate • Valid distribution as long as the sequenced species are a random sample of the clade's total species • Does not take into account knowledge about likely ages of missing species • Combines method of testing for rate-constancy with correction for incomplete species sampling	Pybus & Harvey (2000); Rabosky (2006a)
Adding missing species as simulated splitting times under an exponential (or other) model (CorExS)	• Requires diversification rate and model to estimate the splitting time distribution • Creates a statistically valid distribution of missing splits • Does not depend on information about the ages or relationships of the missing species • Can take into account information about the age of missing species if available	This study

for rate constancy, using either the Monte Carlo constant-rates (MCCR) test for the γ statistic (Pybus and Harvey, 2000) or the, henceforth, ΔAIC_{rc} test in birth-death likelihood analysis (BDL; Rabosky, 2006a). An assumption underlying this approach is that the missing species represent a random sample.

Here, we first assess the magnitude of the missing species problem based on a review of all plant studies that have used statistical methods for inferring diversification rates from molecular trees. We tabulated species sampling density, handling of missing species, and which parameters were inferred under which models. It turned out that more than half of the plant clades had <60% of their species sampled, down to as little as 3.2%. We then apply the most commonly used approaches for handling missing species to two closely related plant clades with a similarly incomplete species sampling (48 and 57%). The investigated clades belong to the Araceae family and occur in the Mediterranean basin and Southeast Asia, regions with different geological histories and present day climates, which was the reason we suspected different diversification patterns. We also introduce a new method for handling missing species, which involves simulations of missing splits under an exponential model, essentially using model-based data augmentation and multiple imputation (Nakagawa and Freckleton, 2008). The new method makes use of information that may be available about the likely ages of the missing

species, but does not require knowledge about their precise placement. Its main advantage, however, is that it provides information (i.e., statistical parameters) regarding the impact of missing data on parameter estimation, which none of the other methods do.

While the focus of this study is the handling of missing species in diversification analysis, a comparison of relevant methods is only possible in conjunction with the inference of diversification itself. We thus apply the three most widely used approaches for diversification modeling, namely the γ statistic, SA, and BDL, together with the above mentioned methods of handling missing species, namely tree simulation and pruning, and our new method, the simulation of missing splits.

4.3 Materials and Methods

Study Systems, Taxon Sampling, and Sequencing

The Areae comprise 153 species in nine genera (Cusimano et al., in review) and are a tribe of the phylogenetically well-understood Araceae (4000 species in 113 genera, Bogner and Petersen, 2007, Cusimano et al., submitted). All Areae are geophytes with a seasonal life cycle. Within Areae, our focal groups are *Typhonium* with 54 species and the *Arum* clade with 64 species in five genera (*Arum, Biarum, Dracunculus, Eminium, Helicodiceros*). *Typhonium* occurs in the Southeast Asian mainland tropics and subtropics, the *Arum* clade in the Mediterranean basin and the Near East, a few species also occur in cold temperate regions of the Himalayas and in Northern Europe. For the *Arum* clade, we sampled 31 of 64 species (48%); we lack 14 species of *Arum*, 12 of *Biarum* and 7 of *Eminium*. For *Typhonium*, we sampled 31 of its 54 species (57%). Outgroup choice was based on Renner and Zhang (2004) and influenced by the need to include taxa with a fossil record for calibration of genetic distances. Online supporting material (OSM) Table S1 lists 14 outgroup taxa with their herbarium vouchers (where applicable) and Genbank numbers; information about all other sequenced species is provided in Cusimano et al. (in review).

We sequenced a nuclear locus, the phytochrome C gene (*PhyC*), and two chloroplast loci, the *rpl20-rps12* intergenic spacer and the tRNA$_{Lys}$ gene (*trnK*), which contains a group I intron that encodes the maturase K (*matK*) open reading frame.

Total DNA extraction, amplification, and sequencing followed the procedure described in Cusimano et al. (in review). Alignments were generated in MacClade (Maddison and Maddison, 1992) and continuously adjusted by eye. The chloroplast and nuclear data partitions were first analyzed separately, and in the absence of statistically supported (>80%) topological contradictions, they were then combined.

Phylogenetic inference relied on maximum likelihood (ML) and ML bootstrapping as implemented in the RAxML BlackBox (Stamatakis et al., 2008) and on Bayesian analysis as implemented in MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003). The GTR + Γ + I model was used for all analyses, with model parameters estimated directly during runs. Markov chain Monte Carlo (mcmc) runs extended for at least 1 million generations, with trees sampled every 100th generations. Convergence was assessed as recommended in the MrBayes manual, and trees saved prior to convergence were discarded as burn-in (1000 trees). Bootstrap support and Bayesian posterior probabilities were plotted on the ML tree using the APE package (Paradis et al., 2004) in R (R Developmental Core Team, 2006).

Divergence Time Estimation

We carried out two sets of Bayesian relaxed clock calculations; one relied on multidivtime (Thorne et al., 1998; Thorne and Kishino, 2002), the other on BEAST (Drummond et al., 2006; Drummond and Rambaut, 2007) , the main difference being that the first approach assumes rate autocorrelation, while the second does not. With 14 additional outgroups, the matrix comprised 4349 nt and 92 species, from which we calculated a ML tree, rooted on *Xanthosoma sagittifolium* and *Caladium bicolor*. This became the input topology used for multidivtime. Time estimation (but not the input tree topology) relied on the chloroplast data matrix of 3333 nt for 92 species because the *PhyC* data were incomplete for the outgroups.

Multidivtime dating used LAGOPUS (Heibl and Cusimano, 2008), an R package that checks the input data for consistency, automates the assignment of constraints to nodes, and connects the executables of the mentioned software packages in a pipeline. Model parameters for the 92-taxon-3333 nucleotides matrix were estimated in baseml, and branch lengths and their variance then calculated in estbranches, all under the F84 + Γ model (the only model implemented in mul-

tidivtime). Based on outgroup fossils (above), the prior on the mean time unit between root and tips (rttm) was set to 120 or 90 Myr (above), with an equally large standard deviation. The priors on the substitution rate at the root were set to 0.00021 or 0.00028 substitutions per site and million years obtained by dividing the mean distance between the root and the tips in the estbranches phylogram by rttm. The prior for the Brownian motion parameter (ν), which controls the magnitude of autocorrelation along the descending branches of the tree, was set to 0.01 (nearly strict clock), 0.1, or 1 (highly relaxed clock), with standard deviations of the same sizes. This had nearly no effect on the node ages, indicating that the signal in the data overwrote these priors; for the final run ν was set to 1. The Markov chain length was 1 million generations, sampled every 100th generation and with a burn-in of 100,000 generations; analyses were repeated several times to explore the effects of different priors as described.

BEAST dating was carried out with version 1.4.8. Analyses used a speciation model that followed a Yule tree prior, with rate variation across branches uncorrelated and log-normally distributed; the substitution model used was the same as in the remaining analyses, GTR + Γ + I. Four groups were constrained to be monophyletic, the Peltandreae, the Pistia clade, the clade of Colocasia and Alocasia, and the Areae. MCMC chains were run for 1 million generations (burnin 10%), with parameters sampled every 1000th generation. Results from individual runs were combined and effective sample sizes were then well above 100. BEAST accommodates calibration uncertainty by applying a prior probability distribution (defined in terms of means and standard deviations) on the age of nodes to which calibration fossils are assigned. We chose normal distributions for all fossil constraints (below), which places less prior probability on a relatively narrow time frame than do exponential or lognormal distributions. Dating runs used the same constraints as described in Renner and Zhang (2004) and Cusimano et al. (2008). The root node constraint of maximally 120 Ma based on the oldest fossils of Araceae (Friis, 1985; Friis et al., 2004) is likely too old, given that the tribe Areae is nested high up in the Araceae family tree (Cabrera et al., 2008). To test the effects of the root constraint, we also ran analyses in which the root node age was constrained to maximally 90 Myr.

Analyses of Diversification and Handling Missing Species

Missing species in plant diversification studies

To assess whether missing (i.e., not sequenced) species are a common problem in plant diversification studies and to compare how they have been handled, we compiled diversification studies in vascular plants over the past 15 years. We tabulated total species, sampling density, inferred ages, whether stem or crown groups had been used, results of different estimators, such as Kendall (1949) and Moran (1951) or maximum-likelihood estimators, diversification and/or speciation rates, the extinction fraction used, and whether the γ statistic with the MCCR test, survival analysis (SA), and/or BDL analysis had been used. In a few cases we recalculated rates after consultation with the authors of the studies.

Comparison of methods in an empirical study system

The two focal clades, *Typhonium* and the *Arum* clade, are not nested and have approximately similar numbers of species (54 and 64, respectively) of which we sampled 48 and 57% (above, *Study Systems, Taxon Sampling, and Sequencing*). Diversification analyses relied on the ultrametric tree obtained under the Bayesian autocorrelated-rates relaxed clock model (above). Lineage-through-time (LTT) plots depict the logarithmic number of lineages ancestral to contemporary species over time. In the next sections we first describe the three methods used for diversification analysis (carried out separately for each of the focal clades), then the methods used to correct for missing species.

γ statistic

The γ statistic (Pybus and Harvey, 2000) tests for departure from a constant-rates pure-birth model. For completely sampled phylogenies, Pybus and Harvey (2000) found that $\gamma = -1.645$ ($+1.645$) represents the critical value of the constant rates test. Values below this cut-off reject the pure birth model ($\gamma = 0$). We relied on the implementation of this statistic (`gamStat`) in the Laser 2.2 package for R (Rabosky, 2006a).

Birth-death likelihood analysis (BDL)

We also applied five likelihood models implemented in the `fitdAICrc` function in Laser, namely two constant-rate models of diversification (a pure-birth model with constant speciation rate and a birth-death model with constant speciation and extinction rates) and three variable-rate models (logistic density dependence, exponential density dependence, and a two-rates variant of the pure-birth model with a rate shift at a certain time point). The best model is judged with the ΔAIC_{rc} statistic using differences in the Akaike Information Criterion between the best-fitting rate-constant and rate-variable model. Additionally, we checked for Type I errors in the inference of rate upswings using the criteria proposed by Rabosky (2006b).

Survival analysis

To conduct SA we used the `diversi.time` function of the APE package 2.2 (Paradis, 1997; Paradis et al., 2004). Three survival models are fit to the data to detect significant departures from a constant-rates (CR) model of diversification (model A). Model B assumes a gradually changing diversification rate and model C an abrupt change in rate at some breakpoint time in the past. None of the three models incorporates extinction. Since the CR model is nested in both variable-rate models, it can be compared with each in a likelihood ratio test, while the non-nested models B and C cannot strictly be evaluated against each other. Best-fitting breakpoint times (T_c) for model C were assessed with an adaptation of the `diversi.time` function written by the first author (`batch.dt`); breakpoint times that gave the highest likelihood score were preferred.

Correction for missing species by tree simulation and pruning

All the above methods are sensitive to incomplete species sampling. To correct for this bias, the γ statistic as well as the BDL analysis rely on repeated (1000 replicates) simulation of trees under the Yule model with the number of tips corresponding to the total species of the focal clades, here 54 and 64. The speciation and extinction rates used for tree simulation was obtained by the best-fitting constant-rates model to our data (Table 4.2). Simulated trees were then randomly pruned to the sample sizes, here 31 and 31. This yields null distribution of γ values against

which the empirical γ value is compared with the Laser function `mccrTest` (MCCR test, Pybus and Harvey, 2000). In BDL analysis, the five diversification models are fit to the thousands of pruned trees and the resulting AIC values compared to that of the empirical tree with the Laser function `fitdAICrc.batch` (ΔAIC$_{rc}$ test, Rabosky, 2006a). Widely used programs for tree simulation, such as Phyl-O-Gen (Rambaut, 2002), have been found to introduce particular biases that concern splits near the present (Hartmann et al., 2008, Results and Discussion). Here we rely on Phyl-O-Gen and Cass (Gernhard, 2008), and additionally two functions implemented in Laser, `birthdeathSim` for the BLD analyses and `mccrTest` for the MCCR analysis.

Table 4.2: Results of BDL analyses. Models: YULE: constant-rate pure birth model; BD: constant-rates birth-death model; DDL: logistic density dependent model; DDX: exponential dd model; Yule2rates: pure birth model with a shift in diversification rate at a certain time point, T$_c$. r = net diversification rate ($r1$ before, $r2$ after a breakpoint time T$_c$); a = extinction fraction; T$_c$ = time of rate shift in Myr; k = carrying capacity parameter; x = rate change parameter; LH = Log-Likelihood; AIC = Akaike information criterion; ΔAIC = difference in AIC values from the best rate-constant model; in bold the value used for the statistic. The last two columns show the p values resulting from the ΔAIC$_{rc}$ test statistic, which evaluated the scores of the empirical data against null distributions obtained from tree simulations (n = 1000) under the Yule model with the estimated diversification rate r (bold), using the simulation programs *Cass*, *Phyl-O-Gen*, and *Laser*. ***significant at the 99% level; **significant at 95%.

	YULE	BD	DDL	DDX	Yule2rates	Program for tree simulation	ΔAIC$_{rc}$ p =
***Typhonium* clade**							
r	**0.0383**	0.0383	0.111	0.399	***r1*** = 0.131	Phylo-O-Gen	0.012**
		a =0	***k*** = 32.900	***xp*** = 0.838	***r2*** = 0.029	Laser	0.001***
					T$_c$ = 32.384	Cass	0***
LH	-48.983	-48.983	**-42.180**	-43.542	-43.428		
AIC	99.965	101.965	**88.36**	91.085	92.856		
ΔAIC	**11.606**	13.606	**0**	2.725	4.496		
***Arum* clade**							
r	**0.058**	0.058	0.122	0.198	***r1*** = 0.098	Phylo-O-Gen	0.236
		a = 0	***k*** = 37.728	***xp*** = 0.453	***r2*** = 0.0406	Laser	0.355
					T$_c$ = 14.657	Cass	0.106
LH	-36.843	-36.843	**-33.903**	-35.162	-34.118		
AIC	75.686	77.686	**71.807**	74.323	74.235		
ΔAIC	**3.879**	5.879	**0**	2,517	2.429		

The CorExS approach for simulating missing species

This new method involves the repeated simulation of missing branching times under an exponential model and their addition to the empirical data, yielding numerous completed data sets. We refer to this approach as CorExS, which stands for "correcting for missing splits by combining exponentially simulated splits with empirical splits." Splitting times (t) were generated from the exponential distribution $P = e^{-mt}$, where P is the cumulative probability density of a split occurring t time units in the past (where the present $t = 0$), and m is the net diversification rate per unit of time. For each missing split, we chose a random number for P from a 0,1 uniform distribution and calculated the time for that splitting event as $t = -ln(P)/m$. We generated 23 values of P for *Typhonium* and 33 for the *Arum* clade, corresponding to the number of missing taxa in each clade. Thus, the simulated additional splitting times (t) for each clade were exponentially distributed with average age $1/m$. Simulations were repeated a 1000 times to obtain a distribution of missing splits, and the γ statistic, SA, and BDL analysis were then applied to the numerous completed data sets (many 1000 since we also used different tree simulation programs; above). Fitted model parameters, likelihoods, and AIC values were compared with a Wilcoxon signed-rank test (a non-parametric statistical hypothesis test for the case of two related samples or repeated measurements on a single sample); nested survival models were compared with a likelihood ratio test.

Diversification rates

For comparability with other plant diversification studies (above, *Missing species in plant diversification studies*) we also calculated diversification rates using the Kendall (1949) and Moran (1951) estimators, and a ML estimator (Sanderson and Donoghue, 1994). None of these consider extinction. It is possible, however, to include extinction by introducing a constant extinction-to-speciation ratio κ (Magallón and Sanderson, 2001; Ricklefs, 2006, for equations see footnotes of Table 4.3). Magallón and Sanderson (2001) used $\kappa = 0$ and $\kappa = 0.9$, which they considered an upper limit. However, Linder (2008) using a broad range of angiosperm clades found that a higher value for κ (namely 0.999) fit much better. We therefore tried both, $\kappa = 0.9$ and $\kappa = 0.999$. Using this approach, we calculated the

mean speciation rate λ (in species/Myr) under a constant-rates birth-death model (CR-BD).

4.4 Results

Missing Species and their Handling in Plant Diversification Studies

Table 4.3* summarizes the main parameters of 23 studies (including the present one) that have investigated diversification patterns in 34 plant clades. Two studies involved meta-analyses (Magallón and Sanderson, 2001; Linder, 2008). Early studies typically report diversification rates under the Yule model, whereas more recent (post-2003) studies have inferred rate changes with the help of the γ statistic and/or SA. Clade ages range from 0.76 and 96 Myr, and clade size from 5 to 4418

*__Footnotes for Table 4.3:__

[1] Kendall (1949) and Moran (1951): Diversification rate $D = \frac{N_t - N_0}{B}$, where B is the summed duration of all branches, N_0 the number of species at time 0, and N_t the number of species at time t (Yule model).

[2] Kendall's estimator for the variance (Kend. var.) for D: $var(D) = \frac{D^2}{2(e^{DT}-1)}$

[3] Maximum-likelihood estimator of diversification rates (Stanley, 1979; Sanderson and Donoghue, 1994): If $N_t = N_0 e^{D(t-t_0)}$ under the pure-birth, or Yule model, then $D = \frac{\ln(N_t - N_0)}{t - t_0}$. The number of species at t_0 is 1 if the calculation begins at the stem, but 2 if the calculation uses only the crown group.

[4] Method-of-moments estimator (mom) of diversification rates (Rohatgi, 1976). Assuming some extinction-to-speciation ratio κ (or ϵ in the notation of Magallón and Sanderson, 2001), one can calculate the diversification rate of the stem or the crown group (Magallón and Sanderson, 2001, equation 6, equation 7; Ricklefs 2006, equation 4)

Table 4.3 ***(facing page)*****:** Studies of plant diversification rates that have used molecular phylogenies combined with diversification modeling. Abbreviations stand for: K & M = Kendall (1949) and Moran (1951) estimators for diversification rates; ML = maximum-likelihood estimator for diversification rates; κ = extinction/speciation ratio; D = diversification rate (species/Myr); λ = speciation rate (species/Ma); SA = survival analysis; dec. = decrease in diversification rate; inc. = increase in diversification rate; BDL = dirth-death likelihood analysis; M = models involved. Results from the present paper in bold. An asterisk marks significant results tested with the Monte Carlo constant-rates test (MCCR).*

Year	Studied Group	Spp. sampled / total spp.	% of spp. sampled	Age in Myr / time interval	Group	K & M[1]	ML[3]	κ	*D*	λ	γ statistic +MCCR	SA	BDL	Authors
1998	Hawaiian silversword alliance (Compositae)	11 / 28	39%	5.2 ± 0.8	crown	+ Kend. var.[2]		0	0.56 ± 0.17					Baldwin & Sanderson
2001	average of 50 angiosperm				crown and stem		mom[4]	0	0.09 (0 - 0.33)					Magallón & Sanderson (Table 1)
							mom[4]	0.9	0.0767 (0 - 0.27)					
2001	*Gentianella* Australia	32 / 256	12.5%	3			+	0	1.73					von Hagen & Kadereit
	Gentianella s.str.	202 / 256	79%	3			+	0	1.77					
	Gentianella S America	2x85 / 256 (2 dispersal events)	33.2%	3			+	0	1.48					
	Gentianella S America (Gentianaceae)	170 / 256 (1 dispersal event)	66.4%	1.06 - 3			+	0	3.21/1.71					
2002	*Gaertnera* (Rubiaceae)	28 / 68	41.2%	5.89-5.07	crown, but not N/2		+	0	0.72 - 0.83 (corrected: 0.6-0.65)					Malcomber
				54	stem		+		0.085					
	SE Asian clade	9 / 16	56.3%	4.1 ± 0.17	stem		+		0.65 - 0.71					
	Sri Lanka clade	4 / 5	80%	4.1 ± 0.17	stem		+		0.37 - 0.41					
2003	*Ehrharta* (Poaceae)	27 / 37	73%	9.82			+	0	0.12 - 0.39					Verboom et al.
				9.82-8.74 (time of radiation)			+	0	0.87 - 4.18					
2003	Hookeriales (Bryophytes)	71 / 743	9.6%	rel. age				0.5				A		Shaw et al.
	Hypnales (Bryophytes)	141 / 4418	3.2%	rel. age				0.5				C dec.		
2003	*Halenia*	22 / 39 (missing spp. added in phylogeny)	56.4%	11.18	stem			0	9.25 (corrected: 0.84, pers. com., von Hagen, 23 May 08)					von Hagen & Kadereit
2003	African Restionaceae	20 / 350	5.7%	65							+2.6* +MCCR			Linder et al.
				65-20							+2*			
	SW Australian Restionaceae	24 / 146	16.4%	60 65-20							-0.64 +MCCR -0.83			
2004	*Primula* sect. *Auricula* (Primulaceae)	25 / 25	100%	2.4	crown									Kadereit et al.
	Western clade	15 / 15	100%	1.8							-4*	B/C dec.		
	Eastern clade	10 / 10	100%	2							-3.3*	B/C dec.		
	Globularia (Globulariaceae)	22 / 23	95.6%	7.6							-0.31	A		
	Gentiana sect. *Ciminalis* (Gentianaceae)	7 / 7	100%	0.76							-1.94*	B dec.		
2005	*Bursera* (Burseraceae)	65 / 84	77.4%	70	*D* for over-lapping 10 Myr intervals; γ for 60-34, 34-15, 15-1.5 Myr	+		0	0.02 - 0.15		for 3 time intervals: 0.27 / +1.91*/ -2.92*		5 M	Becerra
2006	*Agave* s.l. (Agavaceae)	26 / 208	12.5%	10	crown, but not N/2	+ Kend. var.[1]		0	0.32 ± 0.08		-4.40* +MCCR			Good-Avila et al.
		- / 208					+	0	0.51 ± 0.06					
	Yucca (Agavaceae)	- / 50		14.1			+		0.27 ± 0.03					
		- / 50		18.3			+		0.21 ± 0.02					
2006	*Lupinus* (Fabaceae)	81 / 81	100%	1.42 ± 0.29 /1.93 ± 0.35	crown		+		2.5 - 3.72/ 1.93 - 2.78					Hughes & Eastwood
2006	*Pinus* (Pinaceae)	83 / 111	74.8%	128 ± 4	crown						-1.52 +MCCR			Eckert & Hall
2008	Angiosperms						mom[4]	0.999		2.842 ± 0.73				Linder
2008	Burmanniaceae	41 / 92	44.6%	96 ± 3.3							-6.51* +MCCR	B dec.		Merckx et al.
2008	*Yucca* (Agavaceae)	34 / 34	100%	6.41	crown	+		0	0.33 ± 0.06		-3.23* +MCCR			Smith et al.
	Agave sensu latissimus	33 / 240	13.7%	9.52		+		0	0.21 ± 0.001		-4.56* +MCCR			
2008	*Acer* (Sapindaceae)	66 / 124(156)	53.2(42.3)%	60								A, B dec.		Renner et al.
2008	Psoraleeae (Fabaceae)	47 / 51	94%	5.8	crown								6 M	Egan & Crandall
2009	*Nigella* (Ranunculaceae)	19 / 20	95%	rel. age					0.03/rel. time unit			B/C inc.		Bittkau & Comes
2009	Proteaceae	81 / 1757 (all genera)	4.6%	91.4	stem		+	0	0.066					Saquet et al. (Table S2)
					crown		+	0	0.074					
					stem		mom[4]	0.9	0.046					
					crown		mom[4]	0.9	0.056					
2009	*Impatiens*	113/1000	11.3%	22.5 ± 5.6	crown						-6.27*			Janssens et al.
2009	Mirabelieae/ Bossiaeeae (Fabaceae)	~350/700	50%	52 - 55	crown						-2.04*+MCCR?		5 M	Crisp & Cook
	Arum clade (Araceae)	31 / 64	48.4%	40 ± 8	crown	+		0	0.124		-2.08 +MCCR	B/C dec.	5 M	Cusimano & Renner
							+	0	0.085					
							mom[4]	0.9	0.04	0.35				
							mom[4]	0.999	0.001	0.75				
	Typhonium (Araceae)	31 / 54	57.4%	48 ± 8	crown	+		0	0.07		-3.28* +MCCR	B/C dec.	5 M	
							+	0	0.07					
							mom[4]	0.9	0.03	0.26				
							mom[4]	0.999	$5x10^{-5}$	0.53				

species. Twenty studies sampled <60% of the extant species in their focal clade, nine sampled less than 20%, and one sampled 3.2% (Table 4.3). Fifteen analyses employed the γ statistic, SA, or BLD. Inconsistencies in the application of diversification estimators (Malcomber, 2002; von Hagen and Kadereit, 2003; Good-Avila et al., 2006) and in the handling of the missing species problem (Shaw et al., 2003; Merckx et al., 2008; Renner et al., 2008; Smith et al., 2008) are apparent from Table 4.3 and are taken up in the DISCUSSION.

Twelve studies (including the present) estimated diversification rates only with estimators that assume no extinction, five studies (additionally) considered extinction. For the *Arum* clade, the Moran/Kendall estimators yield 0.124 species/Myr, the ML estimator 0.085 species/Myr; for *Typhonium*, both estimators yield 0.07 species/Myr (Table 4.3, bottom). To estimate their diversification rates under the assumption of extinction, we explored extinction-speciation ratios κ of 0.9 or 0.999 (MATERIALS AND METHODS: *Diversification Rates*). With $\kappa = 0.9$, the speciation rates were 0.31 species/Myr for the *Typhonium* clade and 0.46 species/Myr for the *Arum* clade (Table 4.3, bottom). With $\kappa = 0.999$, speciation rates were about twice as high (0.53 and 0.75 species/Myr). Corresponding diversification rates were extremely low (Table 4.3, bottom). The net diversification rates we obtained for the *Typhonium* clade fit with those estimated by Magallón and Sanderson (2001) for the Arales (including 2480 species of Araceae and Lemnaceae in the classification then accepted), namely 0.07 or 0.05 (with $\kappa = 0$ or $\kappa = 0.9$). The diversification rate of the *Arum* clade is nearly twice as high and thus lies above the average angiosperm rate inferred by Magallón and Sanderson (2001).

Trees and Divergence Times for the Study Systems

The combined matrix (4341 characters from nuclear and chloroplast sequences, 76 ingroup and two outgroup taxa) yielded a well-supported Areae clade with for the most part monophyletic genera. *Typhonium* in the traditional wide sense, however, is polyphyletic (Online Supporting Material (OSM) Fig. S1). The taxonomic changes necessary to render *Typhonium* monophyletic are made in Cusimano et al. (in review), and in the present study we always refer to the *Typhonium* clade s.str., which contains 31 species of *Typhonium*, including the type species (OSM Fig. S1). The Mediterranean *Arum* clade also includes 31 species, which belong

to the genera *Arum, Biarum, Dracunculus, Eminium* and *Helicodiceros*. Both our focal clades, *Arum* and *Typhonium*, have high bootstrap support (OSM Fig. S1).

Figure 4.1 shows the chronogram obtained under a relaxed-clock model with correlated rates applied to the concatenated chloroplast data (with extended outgroup sampling). Age estimates were robust to different priors for the time units between root and tips and for the Brownian motion parameter, as well as to different root node constraints (Material and Methods, *Divergence time estimation*). The *Typhonium* crown group dates to the Early/Middle Eocene boundary (48 ± 8 Ma), the stem group age is 59 ± 7 Myr. The *Arum* clade crown group dates to 40 ± 8 Ma, the stem group to 43 ± 8 Myr (Fig. 4.1). Individual genera within the *Arum* clade mostly arose in the Oligocene and Miocene (20 – 25 Ma). Denser species sampling of these genera (21 spp. versus our 9 spp.) also yielded crown group ages of 20 Myr (Mansion et al., 2008). A relaxed clock model with uncorrelated rates gave similar results (data not shown).

Diversification Analysis with Different Methods for Handling Missing Species

Simulations of missing branching times with CorExS

The CorExS simulations require an appropriate value for m, the net diversification rate per unit of time. The obvious initial value to use for m is a clade's inferred net diversification rate, here from the Moran/Kendall estimator, 0.07 and 0.124. When many of the resulting branching times exceeded the focal clades' stem ages, we tried slightly higher values for m, namely 0.1, 0.15, 0.2 and 0.25 for the *Typhonium* clade and 0.25, 0.3, 0.35, 0.4, 0.45 and 0.5 for the *Arum* clade. For each of these values of m, we ran 1000 simulations. Based on the results, we chose a diversification rate of $m = 0.15$ to simulate the 23 missing splits of *Typhonium*, which adds them over the entire age range of the clade (Fig. 4.2B). A few simulations still added branching times that were older than the crown group of *Typhonium*, but the possibility that some missing species indeed diverged early on cannot be excluded, especially given the large difference between stem and crown group ages (Fig. 4.1). For the *Arum* clade, we have better information about the ages of the missing splits: Genera are smaller and all were sampled. Branching times before 25 Ma should therefore all be represented (Fig. 4.1). A diversification rate of $m = 0.45$ added the missing

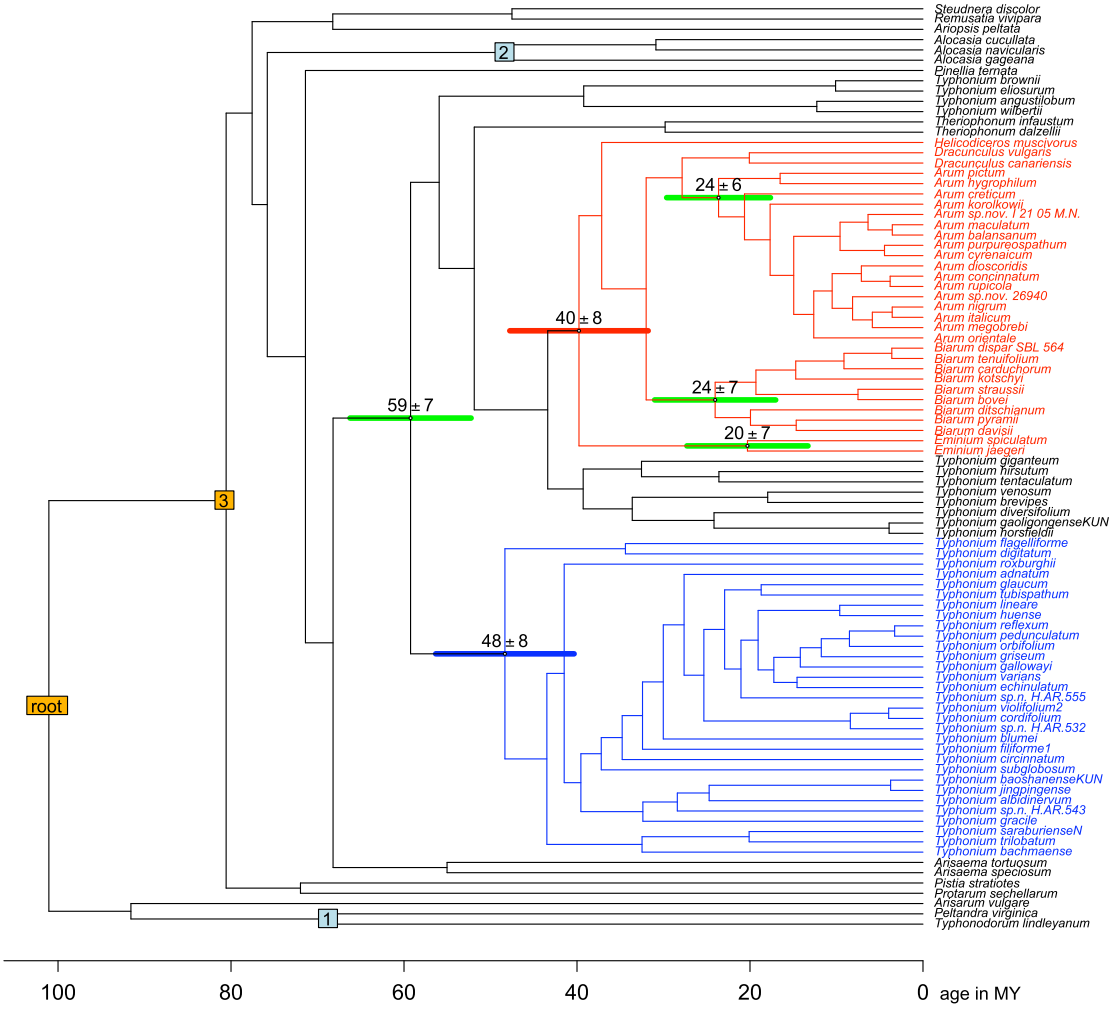


Figure 4.1: Chronogram for the Areae resulting from a relaxed molecular clock applied to the concatenated chloroplast data (3333 nt). Squares indicate constrained nodes (light blue: minimum age constraint; orange: maximum age constraint; MATERIAL AND METHODS). Ages (with standard deviations as bars) are shown for nodes important to the discussion.

branching times in a plausible way (Fig. 4.2E), i.e., most of them after 20 Ma and only occasionally one or two species before then. The LTT plots of the CorExS-corrected data for *Typhonium* and the *Arum* clade are shown in Figs. 4.2B and 4.2E.

Table 4.4: Results of BDL analysis of the CorExS-corrected data sets of the two focal clades. A summary of the AIC, ΔAIC_{rc} values and the model parameters are given as mean and standard deviation (SD). For *Typhonium*, only the 940 of the 1000 CorExS replicates that yielded a breakpoint time $T_c > 30$ Myr are considered for the summary, and all 1000 replicates for the *Arum* clade yielding all breakpoint times $T_c < 10$ Myr. The five models being compared were: YULE: constant-rate pure birth model; BD: constant-rates birth-death model; DDL: logistic density dependent (dd) model; DDX: exponential dd model, and Yule2rates (Y2r): pure birth model with a shift in diversification rate a certain time point, T_c. Model parameters: r = net diversification rate ($r1$ before, $r2$ after a breakpoint time T_c); a = extinction fraction; T_c = time of rate shift in Myr; k = carrying capacity parameter; x = rate change parameter.

		YULE	*r*	BD	*r*	*a*	DDX	*r*	*x*	DDL	*r*	*k*	Y2r	*r1*	T_c	*r2*	ΔAIC_{rc}
***Typhonium* clade** n = 940																	
m = 0.15	**Mean**	83.15	0.057	85.13	0.056	0.02	82.87	0.144	0.29	84.18	0.070	33,530	81.57	0.132	32.37	0.051	1.61
$T_c > 30$	**SD**	3.41	0.002	3.44	0.002	0.06	2.42	0.026	0.07	2.67	0.006	193,086	2.95	0.005	0.11	0.002	0.85
***Arum* clade** n = 1000																	
m = 0.45	**Mean**	-0.66	0.11	-6.64	0.052	0.73	-3.03	0.036	-0.36	1.35	0.109	1,085,238	-7.80	0.08	3.34	0.21	1.17
$T_c < 10$	**SD**	2.70	0.002	5.36	0.007	0.073	4.26	0.008	0.08	2.70	0.002	38,463	6.27	0.01	1.65	0.09	1.80

γ statistic

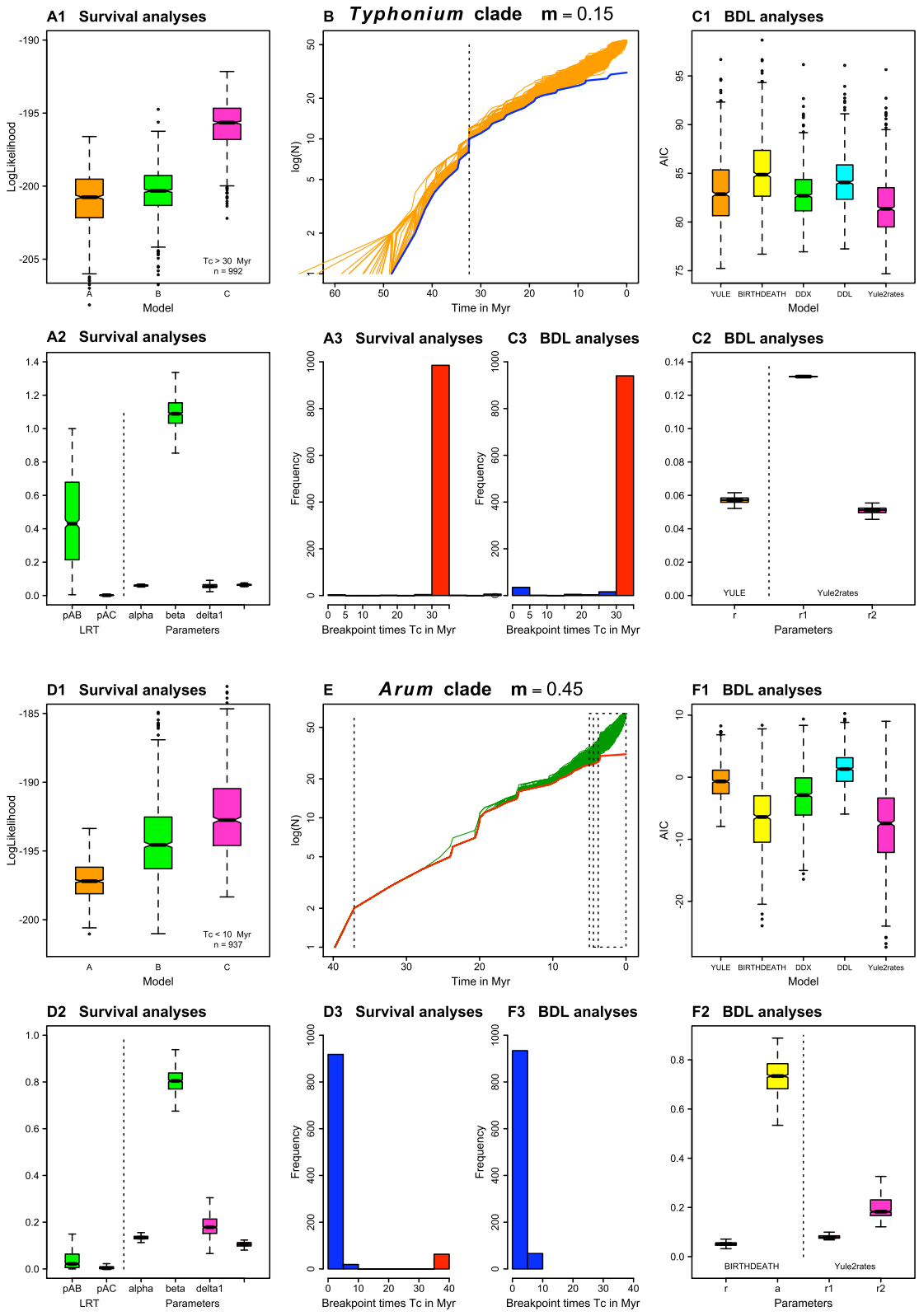
The γ values for the empirical data were -3.28 for *Typhonium* and -2.08 for the *Arum* clade. For the MCCR test, tree simulations in Phyl-O-Gen and Cass were carried out with a diversification rate of 0.0383 species/Myr for *Typhonium*, and 0.058 species/Myr for the *Arum* clade (Table 4.2). Judged against these null distributions, the γ value of *Typhonium* was significantly negative (critical values: -2.13 to -2.78), rejecting a constant diversification rate. The confidence level (CL) was 95%, when using the tree simulation of the `mccrTest` function and rose to 99% when using the simulations obtained from Phyl-O-Gen and Cass. For the *Arum* clade, the γ value of -2.08 was not significantly negative (critical values: -2.26 to -2.62), a constant diversification rate could therefore not be rejected. The CorExS-corrected data sets instead yielded a mean γ of -0.62 ± 0.53 for the *Typhonium* clade, implying that splits were more or less evenly distributed across

the tree. For the *Arum* clade, they yielded a mean γ of 2.67 $\pm$ 0.45, which is above the critical value of completely sampled phylogenies ($\gamma = 1.645$), implying that splits are concentrated towards the present, thus an increase in diversification rate.

BDL analyses

The results of the birth-death likelihood analyses of the empirical data are summarized in Table 4.2. A model of logistic density dependence (DDL) fit the *Typhonium* branching times best; the difference between this model and the best constant-rate model was $\Delta\text{AIC}_{rc} = 11.606$ (Table 4.2). In the ΔAIC_{rc} test, the confidence level for the DDL model being better than the null hypothesis (Yule model) was 99% when compared to the null distributions from Laser (`birthdeathSim`) and Cass, and 95% CL when compared to a null distribution from Phyl-O-Gen (Table 4.2). Also for the *Arum* clade, the DDL model fit the data best (Table 4.2), but the ΔAIC_{rc} of 3.879 was not statistically different from any of the three null distributions (Table 4.2). Cass simulations again gave better p values. For the *Arum* clade, this meant that results were nearly significant at a 90% CL. In BDL analyses of the CorExS-corrected *Typhonium* data, the mean values from the 1000 replicates indicated that the Yule model, with a mean diversification rate of

Figure 4.2 ***(facing page)*:** Results of CorExS analyses of the *Typhonium* clade (**A-C**) and the *Arum* clade (**D-E**). **B, E**: LTT plots of the simulated branching times added to the empirical data (orange, green) and LTT plots of only the empirical data (blue, red), dotted lines and boxes indicate inferred breakpoint times or time ranges; **A1-A3, D1-D3**: Results of survival analyses; **C1-C3, F1-F3**: Results of BDL analyses. **A1, D1**: Boxplots summarizing the likelihood values under models A, B, and C fitted to the data. **A2 and D2** depict the p values of the likelihood ratio tests (LRT) between models A and B (pAB), and A and C (pAC), and the parameters for models B (alpha, beta) and C (delta1, delta2); **C1, F1**: Boxplots summarizing the AIC values of the 5 models fitted to the data; **C2 and F2** depict the parameters for the best-fitting constant-rates and variable-rates models (r = diversification rate ($r1$ before, $r2$ after a breakpoint time T_c); a = extinction fraction); **A3, C3, D3, F3**: Histograms depicting the frequency of inferred breakpoint times under model C (survival analysis, **A3, D3**) and under the Yule2rate model (BDL analysis, **C3, F3**). Boxplots: black line: median; notches: 95% confidence interval of the median; boxes: upper and lower quartile, including 50% of the data; whiskers: minimum and maximum of the data, provided that their length does not exceed 1.5x the interquartile range; open dots: outliers.


A1 Survival analyses
LogLikelihood
Tc > 30 Myr
n = 992
Model
B Typhonium clade m = 0.15
log(N)
Time in Myr
C1 BDL analyses
AIC
YULE BIRTHDEATH DDX DDL Yule2rates
Model
A2 Survival analyses
pAB pAC alpha beta delta1
LRT
Parameters
A3 Survival analyses
Frequency
Breakpoint times Tc in Myr
C3 BDL analyses
Frequency
Breakpoint times Tc in Myr
C2 BDL analyses
YULE
Yule2rates
r r1 r2
Parameters
D1 Survival analyses
LogLikelihood
Tc < 10 Myr
n = 937
A B C
Model
E Arum clade m = 0.45
log(N)
Time in Myr
F1 BDL analyses
AIC
YULE BIRTHDEATH DDX DDL Yule2rates
Model
D2 Survival analyses
pAB pAC alpha beta delta1
LRT
Parameters
D3 Survival analyses
Frequency
Breakpoint times Tc in Myr
F3 BDL analyses
Frequency
Breakpoint times Tc in Myr
F2 BDL analyses
BIRTHDEATH
Yule2rates
r a r1 r2
Parameters

0.057 ± 0.002 species per Myr, was the best constant-rate model (Table 4.4). The best variable-rates model and overall best-fitting model was the Yule2rates model (Fig. 4.2C1, Table 4.4), with an abrupt rate change at 32.3 Ma (Fig. 4.2C3) in 940 of the 1000 replicates. The rate change involved a downturn from a mean rate $r1 = 0.13 \pm 0.005$ species/Myr to a rate $r2 = 0.05 \pm 0.002$ species/Myr (Fig. 4.2C2, Table 4.4). A Wilcoxon signed-rank test (CL = 99%) showed that the AIC value of the Yule2rates model (81.57 ± 2.95) was significantly lower than that of the Yule model (83.15 ± 3.41).

In BDL analyses of the CorExS-corrected *Arum* clade data, the best constant-rates model was the birth-death model with a mean diversification rate of 0.05 ± 0.007 species/Myr and a mean extinction fraction of $a = 0.73 \pm 0.073$. The overall best-fitting model was again the Yule2rates model, but with an increase in rate at breakpoint times <10 Ma and mean rates of $r1 = 0.08$ and $r2 = 0.21$ (Fig. 4.2F2, F3, Table 4.4). A Wilcoxon signed-rank test (CL = 99%) showed that the AIC value of the Yule2rates model (-7.8 ± 6.27) was lower than that of the birth-death model (-6.64 ± 5.36). However, the ΔAIC_{rc} (Table 4.4: 1.17 ± 1.8) is below the significance cut-off found in simulations (Rabosky, 2006b; in our case with c. 60 species it is about 5), meaning that the birth-death model could not be rejected.

Survival analyses

SA analysis requires complete data sets and was therefore only carried out for the CorExS-corrected data. For *Typhonium*, the mean likelihoods of model A (constant-rates pure-birth; -200.9 ± 1.9) and model B (gradual change in diversification; -200.4 ± 1.6; Fig. 4.2A1, Table 4.5) were not significantly different. Model C (rate change at a breakpoint; -195.8 ± 1.6) was significantly better than model A, judged by a likelihood ratio test (Fig. 4.2A1, A2, Table 4.5), and also than model B, judged by a Wilcoxon signed-rank test. Nearly all replicates (992 of 1000) yielded a breakpoint time of 32.38 Ma (Fig. 4.2B, 2A3). The rate decreases from 0.067 ± 0.018 to 0.058 ± 0.037 (Fig. 4.2A2, Table 4.4; in SA, data are read from the present to the past; Paradis, 1997). In SA analyses of the CorExS-corrected *Arum* clade, model C again fit the data best (-192.5 ± 3.1; likelihood ratio test against model A (-197.2 ± 1.5), Wilcoxon signed-rank test against model B (-194.3 ± 2.9); 99% CL; Fig. 4.2D1, D2, Table 4.5), with the most likely break-

Table 4.5: Results of survival analyses of the CorExS-corrected data sets for the two focal clades. A summary of the likelihood values and the model parameters are given as mean and standard deviation (SD). For *Typhonium*, only the 992 of the 1000 CorExS replicates that yielded a breakpoint time $T_c < 30$ Myr are considered for the summary, and for the *Arum* clade, 937 of the 1000 CorExS replicates that yielded a breakpoint time $T_c < 10$ Myr. The three models being compared were a constant-rate model (A), a constant rate change model (B), and a model with a rate change at a breakpoint time (C). Model parameters: alpha and beta of model B; rates delta1 and delta2 of model C. *p*AB and *p*AC are the *p* values resulting from the likelihood ratio test between models A and B, and models A and C.

		LH values for model			Parameters (B)		Parameters (C)		LH ratio test	
		A	**B**	**C**	**alpha**	**beta**	**delta1**	**delta2**	***p*AB**	***p*AC**
***Typhonium* clade** n = 992										
m = 0.15	**Mean**	-200.9	-200.4	-195.8	0.060	1.096	0.058	0.067	0.454	0.002
T_c > 30 Myr	**SD**	1.9	1.6	1.6	0.004	0.095	0.037	0.018	0.280	0.001
***Arum* clade** n = 937										
m = 0.45	**Mean**	-197.2	-194.3	-192.5	0.135	0.805	0.196	0.105	0.051	0.007
T_c < 10 Myr	**SD**	1.5	2.9	3.1	0.008	0.053	0.158	0.009	0.107	0.012

point times <10 Ma for 937 of the 1000 replicates (Fig. 4.2E, D3). The rate increases from 0.105 ± 0.009 to 0.196 ± 0.158 (Fig. 4.2D2, Table 4.5).

4.5 Discussion

In the present study, we first assessed the missing species problem in plant diversification studies and how it has been handled, and then used two similarly incomplete phylogenies to compare how the handling of missing species impacts inferences about diversification. Because they are the most widely used approaches, we applied the γ statistic, birth-death likelihood (BDL) analysis, and survival analysis (SA) in combination with "their" correction methods (MCCR and ΔAIC_{rc} test). Finally, we explored our new approach for handling missing species, the CorExS method. Table 4.6 summarizes how the different methods of simulating trees and handling missing species impacted results.

Missing species in plant diversification studies

Incomplete phylogenies are commonly used to infer diversification rates (Table 4.3), and the γ statistic with the MCCR test is the most commonly used method to correct for missing species. Studies using SA usually had complete species sampling (as assumed by the method; Paradis, 1997). Two studies, however, applied SA to incomplete species sets (3.2/9.6% and 44.6%) without adding missing species (Shaw et al., 2003; Merckx et al., 2008). The same studies also used tree simulations and random pruning to obtain a null distribution under rate-constancy for visual comparison of the empirical LTT plots with simulated trees. Resulting inferences about diversification in these poorly sampled clades are doubtful. One study applied SA analysis with only 53% of the clade's species sampled, but used censoring events in batches to correct for missing species (Renner et al., 2008). In plants, BDL analysis (with the ΔAIC_{rc} test) has only been applied twice, prior to this study (Becerra, 2005; Egan and Crandall, 2008). Two studies (Linder et al., 2003; Becerra, 2005) tried to infer diversification at different time intervals in the past, not using any correction methods, based on the rationale that up to a certain time all nodes are sampled. But even when deep nodes are all included, clade ages within specific time intervals will not all be the same.

Strengths and Weaknesses of Different Methods for Handling Missing Species

The γ statistic and BDL analysis both try to correct for missing splits (nodes, divergence events, or loosely "species" although this is not strictly correct) by creating null distributions under a model of rate-constancy to which the empirical data are then compared. This means that diversification estimation/model-fitting is done on the incomplete phylogeny and that results are evaluated afterwards by comparison with the simulated null distributions. The new approach presented here is based on the assumption that analyzing complete data sets is the best way for inferring diversification rate changes, and the general agreement among statisticians that model-based data augmentation and multiple imputation is the best way of dealing with missing data (Nakagawa and Freckleton, 2008). If instead of augmenting the data based on a model, one adds missing species "by hand" (e.g., Purvis et al., 1995; Barraclough and Vogler, 2002) this has three undesirable

Table 4.6: Summary of the results of the three diversification estimation methods and the two methods for correcting for missing species.

	Typhonium		*Arum clade*	
	Tree simulation	**CorExS-corrected data**	**Tree simulation**	**CorExS-corrected data**
γ statistic	Decreasing diversification	Constant diversification (Yule)	Constant diversification (Yule)	Increasing diversification
BDL	Decreasing diversification (DDL)	Abrupt rate change at 32 Ma	Constant diversification (Yule)	Constant diversification with background extinction
SA	—	Abrupt rate change at 32 Ma	—	Abrupt rate change at <10Ma

effects: It is subjective; one risks adding bias to the data if species sampling is extremely low and many non-sequenced species have to be added; the approach only works with sufficient knowledge of species relationships. In one of our focal clades (*Typhonium*), this precluded adding missing species by hand or even randomly to subclades (following Day et al., 2008).

The CorExS approach overcomes these problems. Missing data are added beforehand to create completed data sets, but with the major difference that missing splits are added, not minimum ages as required in SA nor species to specific branches. The missing splits are simulated under an exponential model and this procedure is repeated a 1000 times (yielding objectivity). The completed batches of data sets (consisting of the empirical splits plus the added ones) can then be analyzed with any of the available methods for diversification estimation, which will yield mean values and a standard deviations. An important advantage is that the missing splits are drawn from an exponential random distribution, which does not assume that missing splits have an even chance of falling anywhere along the clade, but instead fits what we know about diversification. (Of course, missing splits could also be generated under some other model.) If information about the ages of the missing splits is available, the CorExS approach allows including it (age information will influence the *a priori* diversification rate m).

Each of the approaches for assessing diversification rate changes and of handling missing splits has disadvantages (Table 4.1): The γ statistic gives a direction of rate change, but does not consider extinction. It is biased towards more negative values as clade size increases or clade age decreases (Phillimore and Price, 2008). It

may also be biased towards negative values if molecular clocks underestimate deep branches (Revell et al., 2005). However, the major drawback of null distributions of the type used in the γ statistic (and in BDL analysis) is that they assume that species in molecular phylogenies are sampled at random, which is rarely if ever the case (Pybus and Harvey, 2000, Cusimano and Renner, in review). Furthermore, these null distributions do not incorporate knowledge about the likely ages and distribution of missing splits. In the end, only the diversification pattern of the available incomplete phylogeny is tested, which is inferior to model-based data augmentation and multiple imputation (Nakagawa and Freckleton, 2008).

Simulation of trees itself also requires more mathematical analysis than hitherto realized (Hartmann et al., 2008). Problems introduced by tree simulation have to do with, among other things, how pending edges (tip branch lengths) are treated and whether simulations stop at the $n^{th} + 1$ tree and are then pruned to n tips (general sampling approach, as in Cass, Hartmann et al., 2008) or if they stop exactly when trees have reached n tips (simple sampling approach). When we tested model significance with null distributions obtained from three tree simulation programs (Phylo-O-Gen, Laser, Cass) we indeed found clear differences (Table 4.2). Although Hartman et al. (in review) state that for simulating trees under the Yule model the simple sampling approach may be sufficient, our analyses showed that Cass simulations always provided higher confidence levels than those obtained with Phyl-O-Gen or Laser (Results, Table 4.2).

A weakness of the CorExS approach is the need for an *a priori* value for the diversification rate m. In practice, the empirical data, namely clade age and species number, will determine the initial values of m, and LTT plots resulting from different values for m can then be used to chose a value that is biologically plausible, which can be a strength of the approach.

Rate Inferences in the two Study Systems

That the handling of missing species strongly impacts diversification rate inference is clear from the results obtained for the focal clades, which had a species sampling of 57 and 48% (Table 4.6): Analyses of the empirical *Typhonium* data with the γ statistic and BDL analysis, in combination with null distributions from different tree simulation approaches (Table 4.6), yielded a gradual downturn in the

diversification rate. Analyses of the CorExS-corrected data with the γ statistic instead could not reject constant diversification, and SA and BDL analysis of the CorExS-corrected data yielded an abrupt rate decrease at around 32 Ma (Model C, Yule2rates). Results from the γ statistic and the other two methods do not contradict each other because the γ statistic cannot reveal abrupt rate changes. For the *Arum* clade, the γ statistic and the BDL analyses of the empirical data could not reject a constant-rate model, while analysis of the CorExS-corrected data with the methods not incorporating extinction suggested a rate upswing (constant: γ >1.645, abrupt: Model C). The BDL analysis of the CorExS-corrected data instead yielded the birth-death model as the best fit. These results are again not contradictory because the γ statistic does not consider extinction, and SA does not disentangle speciation and extinction rates; these methods are therefore unable to link an upswing of the LTT plot to constant background extinction. Analysis of the 1000 completed CorExS data sets always yielded a single best model, and the Wilcoxon test was always significant. In all cases, the likelihood or AIC values of one model were lower that those of all competing models, meaning that although we obtained a "cloud" of LTT plots results unambiguously pointed to a single model of diversification.

These results illustrate the sensitivity of diversification modeling towards incomplete data and the way they are handled. Even the dependence of the CorExS method on an *a priori* diversification rate m, indirectly highlights a drawback of the *a posteriori* methods of handling missing species: All diversification modeling depends on the parameters of the simulations that are done to obtain the null distributions (and even the tree simulation algorithm will have an effect on significance levels, as shown by our comparison of Phyl-O-Gen, Laser, and Cass simulations).

4.6 Conclusions

The growing field of evolutionary diversification studies requires robust methods and their consistent application. Here, we present a new approach of handling a major problem of such studies, the handling of missing species. The CorExS method reduces the type I error rate because it makes the test of the γ statistic more conservative. Where something is known about the ages of the missing

species, it moreover offers the possibility of adding this information to the available data in an objective way by repeated simulations under an exponential model. When applying the CorExS approach to two Araceae clades, results differed from those obtained with the other methods of handling missing species, suggesting that great caution is warranted in interpreting diversification patters inferred from incomplete data sets.

The R functions `batch.dt` and `CorExS`, with the corresponding plotting functions, are available from the first author on request.

Acknowledgments

We thank T. Gernhard for access to at the time unpublished manuscripts and for help with her Python program Cass, and R. E. Ricklefs for helpful comments and suggestions. Supported by DFG grant RE 603/7-1.

4.7 References

Baldwin, B. G. and M. J. Sanderson. 1998. Age and rate of diversification of the Hawaiian silversword alliance (Compositae). Proc. Natl. Acad. Sci. USA 95:9402–9406.

Barraclough, T. G. and A. P. Vogler. 2002. Recent diversification rates in North American tiger beetles estimated from a dated mtDNA phylogenetic tree. Mol. Biol. Evol. 19:1706–1716.

Becerra, J. X. 2005. Timing the origin and expansion of the Mexican tropical dry forest. Proc. Natl. Acad. Sci. USA 102:10919–10923.

Bittkau, C. and H. P. Comes. 2009. Molecular inference of a Late Pleistocene diversification shift in *Nigella* s. lat. (Ranunculaceae) resulting from increased speciation in the Aegean archipelago. J. Biogeogr. 36:1346–1360.

Bogner, J. and G. Petersen. 2007. The chromsome numbers of the aroid genera. Aroideana 30:82–90.

Cabrera, L. I., G. A. Salazar, M. W. Chase, S. J. Mayo, J. Bogner, and P. Davila. 2008. Phylogenetic relationships of aroids and duckweeds (Araceae) inferred from coding and noncoding plastid DNA. Am. J. Bot. 95:1153–1165.

Crisp, M. D. and L. G. Cook. 2009. Explosive radiation or cryptic mass extinction? Interpreting signatures in molecular phylogenies. Evolution 63:2257–2265.

Cusimano, N., L.-B. Zhang, and S. S. Renner. 2008. Reevaluation of the *cox1* group I intron in Araceae and angiosperms indicates a history dominated by loss rather than horizontal transfer. Mol. Biol. Evol. 25:265–276.

Day, J. J., J. A. Cotton, and T. G. Barraclough. 2008. Tempo and Mode of Diversification of Lake Tanganyika Cichlid Fishes. PLoS. One 3:e1730.

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. PLoS. Biol. 4:e88.

Drummond, A. J. and A. Rambaut. 2007. Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7:214.

Eckert, A. J. and B. D. Hall. 2006. Phylogeny, historical biogeography, and patterns of diversification for *Pinus* (Pinaceae): Phylogenetic tests of fossil-based hypotheses. Mol. Phylogenet. Evol. 40:166–182.

Egan, A. N. and K. A. Crandall. 2008. Divergence and diversification in North American Psoraleeae (Fabaceae) due to climate change. BMC Biol. 6:55–67.

Friis, E. M. 1985. Angiosperm fruits and seeds from the Middle Miocene of Jutland (Denmark). Kongl. Danske Vid. Selskb. Biol. Skrifter 24:1–165.

Friis, E. M., K. R. Pedersen, and P. R. Crane. 2004. Araceae from the Early Cretaceous of Portugal: Evidence on the emergence of monocotyledons. Proc. Natl. Acad. Sci. USA 101:16565–16570.

Gernhard, T. 2008. Cass: A Python tool for analyzing and manipulating phylogenetic trees. `http://www-m9.ma.tum.de/twiki/bin/view/Allgemeines/TanjaGernhard` .

Good-Avila, S. V., V. Souza, B. S. Gaut, and L. E. Eguiarte. 2006. Timing and rate of speciation in *Agave* (Agavaceae). Proc. Natl. Acad. Sci. USA 103:9124–9129.

Hartmann, K., T. Stadler, and D. Wong. 2008. Sampling trees from evolutionary models. Syst. Biol. in review.

Harvey, P. H., R. M. May, and S. Nee. 1994. Phylogenies without fossils. Evolution 48:523–529.

Heibl, C. and N. Cusimano. 2008. LAGOPUS - Divergence time estimation in R using the multidivtime approach. Version 1.4. `www.christophheibl.de/mdt/mdtinr.html` .

Hey, J. 1992. Using phylogenetic trees to study speciation and extinction. Evolution 46:627–640.

Hughes, C. and R. Eastwood. 2006. Island radiation on a continental scale: Exceptional rates of plant diversification after uplift of the Andes. Proc. Natl. Acad. Sci. USA 103:10334–10339.

Janssens, S. B., E. B. Knox, S. Huysmans, E. F. Smets, and V. S. F. T. Merckx. 2009. Rapid radiation of *Impatiens* (Balsaminaceae) during Pliocene and Pleistocene: Result of a global climate change. Mol. Phylogenet. Evol. 52:806–824.

Kadereit, J. W., E. M. Griebeler, and H. P. Comes. 2004. Quaternary diversification in European alpine plants: pattern and process. Philos. Trans. R. Soc. Lond. B Biol. Sci. 359:265–274.

Kendall, D. G. 1949. Stochastic processes and population growth. J. Roy. Stat. Soc. B 11:230–264.

Linder, H. P. 2008. Plant species radiation: where, when, why? Philos. Trans. R. Soc. Lond. B Biol. Sci. 363:3097–3105.

Linder, H. P., P. Eldenas, and B. G. Briggs. 2003. Contrasting patterns of radiation in African and Australian Restionaceae. Evolution 57:2688–2702.

Maddison, W. P. and D. R. Maddison. 1992. MacClade: analysis of phylogeny and character evolution. Version 3.0. Sinauer, Sunderland, Mass.

Magallón, S. and M. J. Sanderson. 2001. Absolute diversification rates in angiosperm clades. Evolution 55:1762–1780.

Malcomber, S. T. 2002. Phylogeny of *Gaertnera* Lam. (Rubiaceae) based on multiple DNA markers: Evidence of a rapid radiation in a widespread, morphologically diverse genus. Evolution 56:42–57.

Mansion, G., G. Rosenbaum, N. Schoenenberger, G. Bacchetta, J. A. Rossello, and E. Conti. 2008. Phylogenetic analysis informed by geological history supports multiple, sequential invasions of the Mediterranean basin by the angiosperm family Araceae. Syst. Biol. 57:269–285.

McPeek, M. A. 2008. The Ecological Dynamics of Clade Diversification and Community Assembly. Am. Nat.. 172:E270–E284.

Merckx, V., L. W. Chatrou, B. Lemaire, M. N. Sainge, S. Huysmans, and E. F. Smets. 2008. Diversification of myco-heterotrophic angiosperms: Evidence from Burmanniaceae. BMC Evol. Biol. 8:178–193.

Moran, P. A. P. 1951. Estimation methods for evolutive processes. J. Roy. Stat. Soc. B 13:141–146.

Nakagawa, S. and R. P. Freckleton. 2008. Missing inaction: the danger of ignoring missing data. Trends Ecol. Evol. 23:592–596.

Nee, S. 2006. Birth-death models in macroevolution. Annu. Rev. Ecol. Syst. 37:1–17.

Nee, S., A. O. Mooers, and P. H. Harvey. 1992. Tempo and mode of evolution revealed from molecular phylogenies. Proc. Natl. Acad. Sci. USA 89:8322–8326.

Paradis, E. 1997. Assessing temporal variations in diversification rates from phylogenies: estimation and hypothesis testing. Proc. R. Soc. Lond. Ser. B-Biol. Sci. 264:1141–1147.

Paradis, E. 1998. Detecting shifts in diversification rates without fossils. Am. Nat. 152:176–187.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290.

Phillimore, A. B. and T. D. Price. 2008. Density-dependent cladogenesis in birds. PLoS. Biol. 6:483–489.

Purvis, A., S. Nee, and P. H. Harvey. 1995. Macroevolutionary Inferences from Primate Phylogeny. Proc. R. Soc. Lond. B Biol. Sci. 260:329–333.

Pybus, O. G. and P. H. Harvey. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. Proc. R. Soc. Lond. Ser. B-Biol. Sci. 267:2267–2272.

R Developmental Core Team. 2006. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rabosky, D. L. 2006a. LASER: a maximum likelihood toolkit for inferring temporal shifts in diversification rates. Evol. Bioinform. Online 2:257–260.

Rabosky, D. L. 2006b. Likelihood methods for detecting temporal shifts in diversification rates. Evolution 60:1152–1164.

Rabosky, D. L., S. C. Donnellan, A. L. Talaba, and I. J. Lovette. 2007. Exceptional among-lineage variation in diversification rates during the radiation of Australia's most diverse vertebrate clade. Proc. R. Soc. Lond. B Biol. Sci. 274:2915–2923.

Rambaut, A. 2002. PhyloGen: Phylogenetic tree simulator package. Version 1.1. http://evolve.zoo.ox.ac.uk/software/PhyloGen/main.html .

Renner, S. S., G. W. Grimm, H. Schneeweiss, T. Stuessy, and R. E. Ricklefs. 2008. Rooting and dating maples (*Acer*) with an uncorrelated-rates molecular clock: Implications for North American/Asian disjunctions. Syst. Biol. 57:795–808.

Renner, S. S. and L.-B. Zhang. 2004. Biogeography of the *Pistia* clade (Araceae): Based on chloroplast and mitochondrial DNA sequences and Bayesian divergence time inference. Syst. Biol. 53:422–432.

Revell, L. J., L. J. Harmon, and R. E. Glor. 2005. Underparameterized model of sequence evolution leads to bias in the estimation of diversification rates from molecular phylogenies. Syst. Biol. 54:973–983.

Ricklefs, R. E. 2006. Global variation in the diversification rate of passerine birds. Ecology 87:2468–2478.

Rohatgi, V. K. 1976. An introduction to probability theory and mathematical statistics. Wiley, New York.

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

Sanderson, M. J. and G. Bharathan. 1993. Does cladistic information affect inferences about branching rates? Syst. Biol. 42:1–17.

Sanderson, M. J. and M. J. Donoghue. 1994. Shifts in diversification rate with the origin of angiosperms. Science 264:1590–1593.

Sauquet, H., P. H. Weston, C. L. Anderson, N. P. Barker, D. Cantrill, A. R. Mast, and V. Savolainen. 2009. Contrasted patterns of hyperdiversification in Mediterranean hotspots. PNAS 106:221–225.

Shaw, A. J., C. J. Cox, B. Goffinet, W. R. Buck, and S. B. Boles. 2003. Phylogenetic evidence of a rapid radiation of pleurocarpous mosses (Bryophyta). Evolution 57:2226–2241.

Smith, C. I., O. Pellmyr, D. M. Althoff, M. Balcazar-Lara, J. Leebens-Mack, and K. A. Segraves. 2008. Pattern and timing of diversification in *Yucca* (Agavaceae): specialized pollination does not escalate rates of diversification. Proc. R. Soc. Lond. B Biol. Sci. 275:249–258.

Stamatakis, A., P. Hoover, and J. Rougemont. 2008. A rapid bootstrap algorithm for the RAxML web-servers. Syst. Biol. 57:758–771.

Stanley, S. M. 1979. Macroevolution, pattern and process. W. H. Freeman, S. Francisco.

Thorne, J. L. and H. Kishino. 2002. Divergence time and evolutionary rate estimation with multilocus data. Syst. Biol. 51:689–702.

Thorne, J. L., H. Kishino, and I. S. Painter. 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. 15:1647–1657.

Verboom, G. A., H. P. Linder, and W. D. Stock. 2003. Phylogenetics of the grass genus *Ehrharta*: Evidence for radiation in the summer-arid zone of the South African Cape. Evolution 57:1008–1021.

von Hagen, K. B. and J. W. Kadereit. 2001. The phylogeny of *Gentianella* (Gentianaceae) and its colonization of the southern hemisphere as revealed by nuclear and chloroplast DNA sequence variation. Org. Divers. Evol. 1:61–79.

von Hagen, K. B. and J. W. Kadereit. 2003. The diversification of *Halenia* (Gentianaceae): Ecological opportunity versus key innovation. Evolution 57:2507–2518.

4.8 Online Supporting Material

Table S1: Sources of the fourteen outgroup species and GenBank accession numbers of newly generated sequences as well as sequences downloaded from GenBank.

Species	Voucher	*trnK*	*PhyC*	*rpl20-rps12*
Alocasia cucullata (Lour.) G. Don	MO acc. 751658	EU886579	-	AY248908
Alocasia gageana Engl. & K. Krause	MO acc. 78364	EU886580	-	AY248909
Alocasia navicularis (Blume) Hook.	T. Croat & V. D. Nguyen 78014 (MO)	EU886581	-	AY248925
Ariopsis peltata J. Graham	J. Murata s.n. 16 Oct 2001	EU886587	-	AY248910
Arisarum vulgare Targ.Toz.	BG Bonn 11472	EU886582	-	EU886630
Caladium bicolor (Aiton) Vent.	T. Croat 60868 (MO)	EU886501	-	AY248943
Peltandra virginica Raf.	J. Bogner 2119 (M)	EU886583	-	AY248942
Pinellia ternata (Thunb.) Breit.	J. McClements s.n., 30 Jul 2001	EU886503	-	AY248931
Pistia stratiotes L.	J. Bogner, 18 Jul 2001, BG Munich	EU886585	-	AY248932
Protarum sechellarum Engl.	J. Bogner 2545 (M)	EU886588	-	AY248933
Remusatia vivipara (Lodd.) Schott	MO acc. 69705b	EU886584	-	AY248934
Steudnera discolor Bull	J. Bogner 1582 (M)	EU886586	-	EF517221
Typhonodorum lindleyanum Schott	J. Bogner s.n. (M)	EU886578	-	EU886627
Xanthosoma sagittifolium (L.) Schott & Endl.	MO acc. 850652b	EU886500	-	AY248944

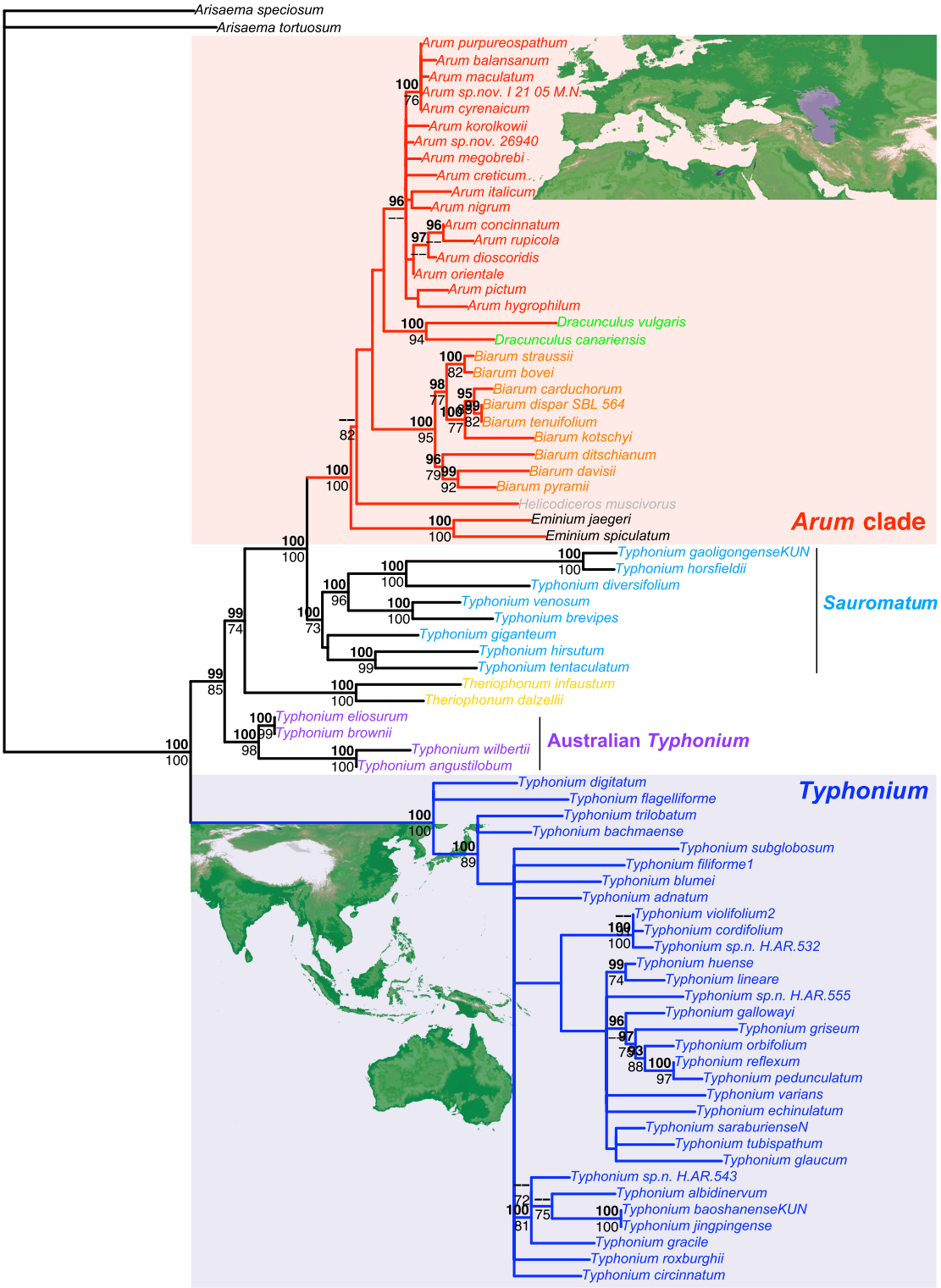


Figure S1:. Maximum likelihood phylogeny for the Areae obtained from combined chloroplast and nuclear data (4341 nt) analyzed under the GTR + I + Γ model of substitution. Values above branches refer to posterior probabilities from Bayesian inference (Material and Methods), those below branches to bootstrap support under maximum likelihood (1000 replicates).

Chapter 5

Slowdowns in diversification rates from real phylogenies may not be real

Cusimano, N. and S. S. Renner.
In review at *American Naturalist*.

5.1 Abstract

Studies of diversification patterns often find a slowing in lineage accumulation towards the present. This seemingly pervasive pattern of rate downturns has been taken as evidence for adaptive radiations, density-dependent regulation, and metacommunity species interactions. The significance of rate changes is evaluated with statistical tests that rely on null distributions, which assume that any non-sequenced ("missing") species are a random sample of the included species. However, sampling in phylogenies is generally not random but instead oversamples deep nodes. We studied the bias this introduces by experimentally pruning a complete empirical phylogeny and two simulated ones, and then applying γ statistic and fitting birth-death likelihood models with the standard tests for rate changes. Results show that oversampling deep nodes biases inferences towards downturns with high statistical confidence. The magnitude of the effect is such that it throws doubt on strong generalizations about rate downturns across major clades of animals and plants.

Key Words

Biodiversity patterns, density dependence, diversification rate, macroevolution, rates changes, species richness

5.2 Introduction

Numerous recent studies have used chronograms from molecular dating to infer diversification patterns, that is, the distribution of cladogenetic events across a tree (e.g., McPeek and Brown, 2007; Weir and Schluter, 2007; Linder, 2008; Phillimore and Price, 2008, 2009). In animal clades as diverse as birds, reptiles, beetles, and fishes, such studies have often revealed apparent slowing in lineage accumulation towards the present (Nee et al., 1992; Zink and Slowinski, 1995; Lovette and Bermingham, 1999; Price et al., 2000; Harmon et al., 2003; Rüber and Zardoya, 2005; Kozak et al., 2006; McKenna and Farrell, 2006; Weir, 2006; Phillimore and Price, 2008; Rabosky and Lovette, 2008a,b). A predominant pattern of rate slowdowns was also found in an analysis of 245 phylogenies (182

of them time-calibrated), including 39 angiosperm phylogenies (20 of them calibrated, McPeek, 2008). Other plant phylogenies (not included in McPeek, 2008) also exhibited slowing diversification rates (Kadereit et al., 2004; Good-Avila et al., 2006; Merckx et al., 2008; Smith et al., 2008, summarized in Cusimano and Renner, in review). This seemingly pervasive pattern of diversification downturns across major groups of animals and plants has been taken as evidence for adaptive radiations (Harmon et al., 2003; Weir, 2006), density- (diversity-) dependent regulation (Weir, 2006; Phillimore and Price, 2008, 2009; Rabosky and Lovette, 2008a,b), and metacommunity species interactions (McPeek, 2008).

The significance of rate downturns typically is evaluated with the γ statistic and the Monte Carlo constant-rates test (MCCR test; Pybus and Harvey, 2000) or with AIC scores in birth-death likelihood analyses (Rabosky, 2006a,b). To create the probability distribution of the null hypothesis, both statistics simulate phylogenies for the complete number of species in the focal clade. If the molecular phylogeny for the focal clade is incomplete, the simulated tree sets are then pruned to number of species actually sequenced. These null distributions are based on the assumption that the species in a tree represent a random sample of all species in a clade. Overdispersed sampling will raise the type I error of the MCCR test, whilst underdispersed sampling will raise the type II error (Pybus and Harvey, 2000). Random sampling is a critical assumption because incompletely sampled clades may be typical in molecular phylogenetics. For example, of the 245 phylogenies analyzed by McPeek (2008), 44% did not include all species of the studied clades. This non-random sampling probably arises because systematists try to strike a balance between the expense of time and funds on sequencing and the return in terms of insights into relationships. Such insights can often be gained from sparse sampling as long as each morphologically or geographically defined group is represented. The resulting overrepresentation of deep nodes relative to tip nodes, however, leads to overdispersion and hence an increased type I error when inferring diversification processes from trees. Here we quantify the magnitude of the bias introduced by phylogenetically informed sampling (i.e., overrepresentation of deep nodes). Our experiments involve a fully sampled real phylogeny that was pruned in different ways to determine how overrepresentation of deep nodes influences birth-death likelihood analyses and the γ statistic. To better quantify the magnitude of any bias, we conducted additional analyses using simulated data.

5.3 Material and Methods

We used a chronogram of the Cucurbitaceae genus *Momordica* that includes 58 of clade's 59 species and a crown group age of 31 ± 5 Myr (Schaefer and Renner, 2010). We treat this phylogeny as complete. To test the effect of different sampling strategies we reduced the complete data set in the following ways: We randomly pruned the chronogram to 80, 60, and 40% of the species, with each pruning performed 10 times. Additionally, we pruned the chronogram manually in two ways: First, we took out entire clades, yielding chronograms with 66 and 54% of the original species number. Next, we left one or two representatives of every major lineage, yielding trees with 44 and 30% of the species, but all deep nodes represented.

All together this yielded one complete and 34 pruned chronograms for which the corresponding branching times were analyzed with the γ statistic and the MCCR test (Pybus and Harvey, 2000) as well as birth-death likelihood analysis (BDL), both implemented in Laser 2.2 (Rabosky, 2006a). The γ statistic tests if splitting events across the phylogeny are evenly distributed or accumulated towards the root or the tips of the phylogeny. In the BDL analysis, five models were fitted to the 35 data sets, two constant-rate models of diversification (a pure-birth (Yule) model with constant speciation rate; a birth-death model with constant speciation and extinction rates) and three variable-rate models (logistic density dependence model (DDL); exponential density dependence model (DDX); a two-rates variant of the pure-birth model with a rate shift at a certain time point (Yule-2-rates)). The ΔAIC_{rc} is calculated for every data set as the difference in AIC scores of the best-fit rate-constant and the best-fit rate-variable model.

To obtain a null distribution of AIC scores, we simulated 1000 trees for each of the 35 data sets independently, with the number of tips corresponding to the complete number of species (58) under a pure-birth model with a diversification rate obtained by fitting the Yule model to the respective data set. The simulated trees were then pruned (if necessary) to the real number of species occurring in the variously pruned experimental data sets. Fitting each of the five diversification models to every simulated tree yielded a null distribution of AIC scores under the null model against which to compare the scores obtained with the respective "true" phylogeny. Tree simulations were carried out in Cass (Gernhard, 2008;

Hartmann et al., 2008). To obtain a null distribution of γ values against which to test the empirical (and experimental tree) γ values, we used the MCCR test (cf. Introduction).

Additionally, we simulated two trees under the Yule model with a diversification rate of 0.5, 60 and 150 tips, and slightly positive γ values ($\gamma = 0.076$ and $\gamma = 0.07$, respectively). These trees were also pruned 10 times to 80, 60, and 40% of their original species numbers. To test for the effect of oversampling deep nodes, we again conducted manual pruning, leaving 75, 65, 55 and 45% of the species. All resulting trees were analyzed with the γ statistic and the MCCR test (1000 replicates), as these are the most widely applied methods for diversification rate inference.

5.4 Results

Table 5.1 shows that the best model for the complete *Momordica* phylogeny is the Yule-2-rates model, with a rate decrease at 2 Ma. However, a simple Yule model could not be rejected, nor did the γ statistic reject a constant-rates model. The diversification rate is 0.08 species per million years (sp./Myr), and an LTT plot for *Momordica* is shown in Fig. 5.1, together with plots for the 34 pruned trees.

When 20% of the species were pruned at random (leaving 80% in the tree), the Yule model could still not be rejected in six cases (Table 5.1); the diversification rates became lower than they had been in the complete phylogeny (0.071-0.075 sp./Myr). In four other cases, the DDL, DDX, or Yule-2-rates models were preferred with a 90% confidence level (CL; Table 5.1).

Pruning 40% of the species from the tree at random had nearly the same

Table 5.1 ***(facing page)*:** Results from fitting five diversification models to the complete phylogeny and variously pruned subsets of it (Materials and Methods). The headers refer to the following models: Best RC = best rate-constant model; r1 = net diversification rate; Best RV = best rate-variable model; xp/k/r2 refer to the parameters associated with the DDX, DDL, and yule-2-rate models; xp = rate change parameter in the DDX model, k refers to the carrying capacity parameter in the DDL model, and r2 is the second net diversification rate after the breakpoint time st (in million years) in the Yule-2-rates model; st = break-point in the Yule-2-rates model; ΔAIC_{rc} is the difference between the best-fitting rate-variable and the best-fitting rate-constant model; cr. value = the critical value of the MCCR test.

Pruning Scheme	% of species remaining	Species	Branching times	Missing BT	Birth-death likelihood analysis: Best RC model	r1	Best RV model	Model parameters r1	xp/k/r2	st	dAICrc	*p* value	γ statistic: γ value	MCCR test: cr. value	*p* value
Complete Phylogeny	100%	58	57	-	**pure birth**	**0.08**	yule-2-rates	0.09	0.03	2.02	2.31	0.13	-1.25		0.11
							DDX	0.20	0.28						
							DDL	0.11	116.00						
Pruned randomly to	80%	46	45	12	**pure birth**	**0.07**	yule-2-rates				1.15	0.20	-0.99	-1.95	0.32
					pure birth	**0.07**	yule-2-rates				2.08	0.13	-1.49	-2.26	0.24
					pure birth	**0.07**	DDx				1.77	0.14	-1.51	-1.10	**0.02**
					pure birth	0.07	**DDL**	**0.12**	**66.14**		2.62	**0.10**	-1.88	-1.28	**0.01**
					pure birth	**0.07**	DDX				2.06	0.12	-1.57	-2.09	0.15
					pure birth	0.07	**DDX**	**0.27**	**0.45**		2.77	**0.09**	-1.87	-2.36	0.14
					pure birth	0.07	**DDX**	**0.28**	**0.46**		3.04	**0.07**	-1.88	-1.32	**0.01**
					pure birth	0.07	**yule-2-rates**	**0.09**	**0.03**	**3.27**	2.69	**0.09**	-1.53	-1.50	**0.05**
					pure birth	**0.08**	yule-2-rates				1.36	0.19	-1.16	-1.70	0.16
					pure birth	**0.07**	yule-2-rates				1.31	0.17	-1.40	-1.92	0.12
Pruned randomly to	60%	34	33	24	**pure birth**	**0.07**	DDX				1.03	0.30	-1.36	-2.10	0.25
					pure birth	0.06	**DDX**	**0.38**	**0.66**		4.64	**0.07**	-2.14	-2.88	0.21
					pure birth	**0.07**	DDX				0.09	0.44	-0.88	-2.06	0.43
					pure birth	0.06	**DDL**	**0.12**	**42.62**		3.71	**0.08**	-2.06	-2.61	0.16
					pure birth	0.06	**DDL**	**0.13**	**41.82**		4.24	**0.06**	-2.17	-2.39	**0.09**
					pure birth	**0.07**	DDL				1.10	0.31	-1.44	-1.93	0.16
					pure birth	**0.07**	yule-2-rates				0.76	0.31	-1.00	-2.27	0.51
					pure birth	**0.07**	DDX				0.77	0.33	-1.27	-1.76	0.17
					pure birth	**0.06**	yule-2-rates				3.01	0.12	-1.73	-1.97	**0.08**
					pure birth	0.06	**DDL**	**0.15**	**37.50**		9.44	**0.00**	-3.01	-2.74	**0.02**
Pruned randomly to	40%	22	21	36	pure birth	0.05	DDX				2.01	0.32	-1.52	-2.53	0.38
					pure birth	**0.05**	DDL				3.21	0.16	-1.86	-1.49	**0.01**
					pure birth	**0.05**	DDX				2.17	0.30	-1.33	-2.10	0.28
					pure birth	**0.06**	DDX				-0.91	0.82	-0.41	-2.82	0.93
					pure birth	**0.06**	DDX				-0.23	0.69	-0.93	-2.87	0.81
					pure birth	**0.05**	DDL				2.84	0.22	-1.78	-2.63	0.32
					pure birth	0.05	**DDX**	**0.51**	**0.95**		4.57	**0.10**	-1.78	-2.39	0.18
					pure birth	**0.05**	yule-2-rates				3.89	0.13	-1.73	-2.99	0.43
					pure birth	0.05	**DDL**	**0.13**	**24.45**		4.29	**0.10**	-2.08	-3.37	0.52
					pure birth	**0.05**	DDX				3.02	0.21	-1.44	-2.08	0.22
Cut off whole clades	66%	40	39	18	**pure birth**	**0.08**	DDX				-1.43	0.74	-0.42	-1.98	0.60
	54%	33	32	25	**pure birth**	**0.06**	DDL				2.21	0.20	-1.72	-2.15	0.13
Deep nodes sampled	44%	27	26	31	pure birth	0.05	**DDL**	**0.17**	**26.45**		14.39	**0.00**	-3.54	-3.20	**0.02**
	30%	19	18	39	pure birth	0.04	**DDL**	**0.24**	**17.36**		20.56	**0.00**	-3.91	-2.29	**0.00**

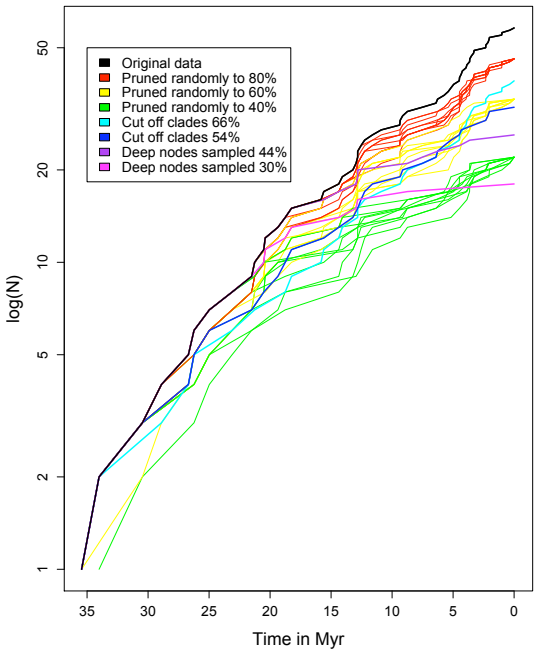


Figure 5.1: Lineage-through-time plots obtained from the complete phylogeny and the variously pruned subsets of it (Table 5.1)

effects (Table 5.1), but diversification rates under the Yule model became even lower (0.063-0.068 sp./Myr). Of the variable-rate models, DDL usually became the best fitting model, in one case even with 99.9% CL (Table 5.1).

When 60% of the species were pruned from the phylogeny at random, the Yule model became the preferred model (Table 5.1), with rates ranging from 0.051 to 0.064 sp./Myr. The DDL and the DDX model each were preferred once (90% CL). In 27% of the randomly pruned phylogenies, the γ statistic suggested a significant accumulation of nodes in the early part of the tree, hence a diversification rate downturn, with p values ranging from 0.01 to 0.09. In all others the constant-rate model could not be rejected.

Cutting-off entire clades from the phylogeny (with 60 or 54% of the species left) resulted in the Yule model becoming the preferred model in BDL analysis, with diversification rates of 0.079 and 0.064 sp./Myr. The constant-rate model was preferred by the γ statistic.

Cutting-off sister species near the tips and thus oversampling deep nodes, re-

sulted in DDL becoming the preferred model, irrespective of whether 44 or 30% of deep nodes where left in the tree (Table 5.1). Confidence levels in both cases were 99.9%. The extreme downturns are also seen in the LTT plots in Fig. 5.1. The obtained diversification rates (0.17 and 0.24 sp./Myr) and carrying capacity (k = 26.25 and 17.36) differ strongly from those obtained when fitting the DDL model to the complete phylogeny (0.11 sp./Myr and k = 116). Oversampling deep nodes thus gives starkly different results from randomly pruned phylogenies or phylogenies missing entire clades (for which the Yule model could be rejected with confidence only in one case). The γ statistic also inferred rate downturns with high confidence (Table 5.1).

For the simulated phylogenies of initially 60 species, random pruning resulted in inferred downturns in seven cases and in inferred constant diversification in the remaining 23 (Table 5.2). Downturns are inferred with higher confidence, the more severely pruned a phylogeny (20% pruned: CL = 90%, 40% pruned: CL = 95/99%; 60% pruned: CL = 95/99%). For the simulated phylogenies of initially 150 species, a downturn was inferred in only one case (CL = 95%; Table 5.2). For phylogenies oversampling deep nodes, irrespective of their initial species number, those with only 75% of the species never rejected the constant rate model, whereas those sampling 65, 55, and 45% of the species all inferred a downturn with high statistical support (Table 5.2). The γ values of all but two pruned phylogenies were negative, in contrast to the initial phylogeny ($\gamma = 0.076$, Table 5.2).

5.5 Discussion

These experiments reveal a strong and consistent effect of phylogenetically informed taxon sampling on the diversification patterns likely to be inferred. Trees in which deep nodes are oversampled will bias results towards rate slowdowns, that is, density-dependent diversification. Simulations suggest that this pattern may be independent of the initial γ value and the size of the original tree, and that it occurs in trees with a species sampling <70%. Even for trees with random species sampling, the true model (i.e., the one fitting the full 58 species of *Momordica*) could be inferred only in 63% (BDL) or 73% (γ statistic) of the cases for the empirical and simulated smaller phylogenies. For the remaining 37 or 23% of trees, a diversification slowdown was inferred. Random pruning of the 150-species phy-

		60 species			150 species		
		γ value	**cr. value**	***p* value**	**γ value**	**cr. value**	***p* value**
Complete phylogeny		**0.08**	-	0.53	**0.07**	-	0.53
Pruned randomly to	80%	-0.46	-1.63	0.39	-0.55	-2.33	0.61
		-0.37	-2.32	0.68	-0.93	-1.59	0.18
		0.64	-1.66	0.83	-0.20	-2.34	0.83
		-0.47	-1.73	0.36	-0.32	-2.08	0.65
		-0.55	-1.61	0.33	-1.17	-1.79	0.17
		-0.86	**-1.09**	**0.08**	-0.46	-1.70	0.41
		-0.40	-0.94	0.17	-0.81	-1.77	0.27
		0.19	-1.30	0.53	-0.19	-2.10	0.65
		-1.14	**-1.34**	**0.08**	-1.09	-1.52	0.14
		-0.05	-1.77	0.57	-0.51	-2.13	0.55
Pruned randomly to	60%	**-2.33**	**-1.62**	**0.01**	-1.20	-2.65	0.55
		-0.56	-2.37	0.74	**-1.92**	**-1.86**	**0.04**
		-1.71	**-1.49**	**0.03**	-1.08	-2.60	0.59
		-0.45	-2.57	0.73	-1.17	-3.18	0.78
		-0.92	-2.09	0.39	-0.92	-2.21	0.48
		-1.26	-2.33	0.35	-0.98	-2.29	0.46
		-0.02	-1.42	0.48	-1.64	-2.56	0.34
		-0.78	-1.64	0.29	-0.65	-2.57	0.72
		-1.91	-2.59	0.20	-1.05	-2.76	0.67
		-1.72	-2.67	0.33	-0.90	-2.75	0.72
Pruned randomly to	40%	**-2.17**	**-1.91**	**0.03**	-1.01	-4.03	0.98
		-2.92	**-1.66**	**0.00**	-0.96	-2.78	0.75
		-0.92	-1.91	0.40	-1.47	-3.22	0.72
		-2.07	**-1.59**	**0.01**	-1.37	-3.28	0.80
		-0.84	-2.26	0.57	-1.80	-3.26	0.61
		-1.96	-2.53	0.17	-2.20	-3.14	0.36
		-1.78	-2.57	0.26	-0.92	-3.30	0.94
		-1.24	-2.91	0.68	-1.21	-2.06	0.31
		-1.56	-2.17	0.21	-1.51	-3.71	0.89
		-1.43	-2.29	0.28	-0.79	-3.41	0.95
Deep nodes sampled	75%	-1.37	-1.91	0.14	-1.58	-2.59	0.31
	65%	**-2.60**	**-2.23**	**0.02**	**-2.51**	**-2.74**	**0.09**
	55%	**-4.00**	**-2.58**	**0.00**	**-2.94**	**-2.02**	**0.00**
	45%	**-3.73**	**-0.68**	**0.00**	**-3.68**	**-2.28**	**0.00**

Table 5.2: Results from fitting five diversification models to the two simulated phylogenies (under the Yule model with a diversification rate of 0.5 and 60 and 150 tips) and the variously pruned subsets. Oversampling of deep nodes was simulated by manual pruning, leaving 75, 65, 55 and 45% of the total species in the trees. Resulting trees were analyzed with the γ statistic and the MCCR test (1000 replicates); cr. value = the critical value of the MCCR test.

logeny did not result in a bias towards rate downturns (a downturn was inferred only once); however, a rate upswing was never inferred.

The magnitude of the downturn bias in smaller phylogenies (Tables 1, 2) is such that it throws doubt on strong generalizations about diversification slowdowns across major clades of animals and plants. A compounding effect may be the sensitivity of γ towards clade size and age demonstrated by Phillimore and Price (2008, 2009). Clades larger than 50 species and younger than 20 Myr tend to show more negative γ values (i.e., stronger slowdowns) than do smaller or older clades. For the *Momordica* data set, with its 58 species and c. 35 Myr age, this should not have been a major problem. However, γ is also biased towards negative values if molecular clocks underestimate deep branches (Revell et al., 2005), a problem of unknown magnitude. Another concern is that oversampling deep nodes also impacts phylogenetic tree imbalance (Heath et al., 2008). Using a range of tree imbalance measures, studies have found that published phylogenies reconstructed from empirical data are more imbalanced than predicted under the equal-rates Markov model (Heath et al., 2008). Randomized pruning of 50% of the species (in real trees), as well as removal of just the terminal branches, increases imbalance for nodes of a given size. This, in turn, may mislead analyses using tree shape. It will be important to be aware of all these biases as phylogenies are being increasingly used to study large-scale patterns of diversification.

Here we have focused not on effects of tree imbalance, but instead on the effects of oversampling deep nodes in incompletely sampled phylogenies. All methods that attempt to correct for missing species in diversification analyses assume that the distribution of missing species is random with respect to phylogenetic topology and clade age (Pybus and Harvey, 2000; Rabosky, 2006a). An additional problem may be that trees are simulated under assumed diversification rates obtained from the incompletely sampled phylogenies, which underestimate the true rate as shown here. Our results, which show that real trees, with the corrections for missing species by standard tree simulation and pruning (MCCR test), still lead to the wrong diversification models being inferred with confidence, stress the importance of correcting for missing species in ways that will reduce type I error. This may be possible if the missing splits (node ages) are instead simulated numerous times under an exponential distribution and then added to the available ones, permitting calculation of confidence levels (Cusimano and Renner, in review). This would lead

to more conservative inferences of diversification rates or other parameters from incomplete phylogenies.

Of course, this paper highlights only one problematic aspect of inferring diversification rates from empirical phylogenies, and the bias quantified here may only affect phylogenies that sample <80% of a clade's extant species. Nevertheless, incomplete data sets may be the rule rather than the exception, and being aware of the biases introduced by incomplete sampling is therefore important.

Acknowledgement

For discussion and comments on the manuscript we thank Robert E. Ricklefs and Albert B. Phillimore. Supported by DFG grant RE 603/7-1.

5.6 References

Gernhard, T. 2008. Cass: A Python tool for analyzing and manipulating phylogenetic trees. `http://www-m9.ma.tum.de/twiki/bin/view/Allgemeines/TanjaGernhard` .

Good-Avila, S. V., V. Souza, B. S. Gaut, and L. E. Eguiarte. 2006. Timing and rate of speciation in *Agave* (Agavaceae). Proc. Natl. Acad. Sci. USA 103:9124–9129.

Harmon, L. J., J. A. Schulte, A. Larson, and J. B. Losos. 2003. Tempo and mode of evolutionary radiation in Iguanian lizards. Science 301:961–964.

Hartmann, K., T. Stadler, and D. Wong. 2008. Sampling trees from evolutionary models. Syst. Biol. in review.

Heath, T. A., D. J. Zwickl, J. Kim, and D. M. Hillis. 2008. Taxon Sampling Affects Inferences of Macroevolutionary Processes from Phylogenetic Trees. Syst. Biol. 57:160–166.

Kadereit, J. W., E. M. Griebeler, and H. P. Comes. 2004. Quaternary diversification in European alpine plants: pattern and process. Philos. Trans. R. Soc. Lond. B Biol. Sci. 359:265–274.

Kozak, K. H., D. W. Weisrock, and A. Larson. 2006. Rapid lineage accumulation in a non-adaptive radiation: phylogenetic analysis of diversification rates in eastern North American woodland salamanders (Plethodontidae: *Plethodon*). Proc. R. Soc. Lond. B Biol. Sci. 273:539–546.

Linder, H. P. 2008. Plant species radiation: where, when, why? Philos. Trans. R. Soc. Lond. B Biol. Sci. 363:3097–3105.

Lovette, I. J. and E. Bermingham. 1999. Explosive speciation in the New World *Dendroica* warblers. Proc. R. Soc. Lond. B Biol. Sci. 266:1629–1636.

McKenna, D. D. and B. D. Farrell. 2006. Tropical forests are both evolutionary cradles and museums of leaf beetle diversity. Proc. Natl. Acad. Sci. USA 103:10947–10951.

McPeek, M. A. 2008. The Ecological Dynamics of Clade Diversification and Community Assembly. Am. Nat.. 172:E270–E284.

McPeek, M. A. and J. M. Brown. 2007. Clade age and not diversification rate explains species richness among animal taxa. Am. Nat. 169:E97–E106.

Merckx, V., L. W. Chatrou, B. Lemaire, M. N. Sainge, S. Huysmans, and E. F. Smets. 2008. Diversification of myco-heterotrophic angiosperms: Evidence from Burmanniaceae. BMC Evol. Biol. 8:178–193.

Nee, S., A. O. Mooers, and P. H. Harvey. 1992. Tempo and mode of evolution revealed from molecular phylogenies. Proc. Natl. Acad. Sci. USA 89:8322–8326.

Phillimore, A. B. and T. D. Price. 2008. Density-dependent cladogenesis in birds. PLoS. Biol. 6:483–489.

Phillimore, A. B. and T. D. Price. 2009. Ecological influences on the temporal pattern of speciation. Pages 140–156 *in* Speciation and Patterns of Diversity (R. Butlin, J. Bridle, and D. Schluter, eds.). Cambridge University Press, Cambridge.

Price, T. D., I. J. Lovette, E. Bermingham, H. Gibbs, and A. Richman. 2000. The imprint of history on communities of North American and Asian warblers. Am. Nat. 156:354–367.

Pybus, O. G. and P. H. Harvey. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. Proc. R. Soc. Lond. Ser. B-Biol. Sci. 267:2267–2272.

Rabosky, D. L. 2006a. LASER: a maximum likelihood toolkit for inferring temporal shifts in diversification rates. Evol. Bioinform. Online 2:257–260.

Rabosky, D. L. 2006b. Likelihood methods for detecting temporal shifts in diversification rates. Evolution 60:1152–1164.

Rabosky, D. L. and I. J. Lovette. 2008a. Density-dependent diversification in North American wood warblers. Proc. R. Soc. Lond. B Biol. Sci. published online.

Rabosky, D. L. and I. J. Lovette. 2008b. Explosive evolutionary radiations: Decreasing speciation or increasing extinction through time? Evolution 62:1866–1875.

Revell, L. J., L. J. Harmon, and R. E. Glor. 2005. Underparameterized model of sequence evolution leads to bias in the estimation of diversification rates from molecular phylogenies. Syst. Biol. 54:973–983.

Rüber, L. and R. Zardoya. 2005. Rapid cladogenesis in marine fishes revisited. Evolution 59:1119–1127.

Schaefer, H. and S. S. Renner. 2010. A three-genome phylogeny of *Momordica* (Cucurbitaceae) suggests seven returns from dioecy to monoecy and recent long-distance dispersal to Asia. Mol. Phylogenet. Evol. in press.

Smith, C. I., O. Pellmyr, D. M. Althoff, M. Balcazar-Lara, J. Leebens-Mack, and K. A. Segraves. 2008. Pattern and timing of diversification in *Yucca* (Agavaceae): specialized pollination does not escalate rates of diversification. Proc. R. Soc. Lond. B Biol. Sci. 275:249–258.

Weir, J. T. 2006. Divergent timing and patterns of species accumulation in lowland and highland neotropical birds. Evolution 60:842–855.

Weir, J. T. and D. Schluter. 2007. The latitudinal gradient in recent speciation and extinction rates of birds and mammals. Science 315:1574–1576.

Zink, R. M. and J. B. Slowinski. 1995. Evidence from molecular systematics for decreased avian diversification in the Pleistocene epoch. Proc. Natl. Acad. Sci. USA 92:5832–5835.

Chapter 6

Relationships within the Araceae: Comparison of Morphological Patterns with Molecular Phylogenies

Cusimano, N., Bogner, J., Mayo, S. J., Keating, R. C., Boyce, P. C., Wong, S. Y., Hetterscheid, W. L. A., Hesse, M. and J. C. French.
Draft.

6.1 Abstract

The first family-wide molecular phylogeny of the Araceae, a family of about 4000 species and 113 genera, became available in 1995 (89 genera; 488 characters), followed by a cladistic analysis of morphological and anatomical data by Mayo et al. (1997, 106 genera, 63 characters). A 2008 phylogeny included 102 genera (including the five genera of Lemnoideae) that were sequenced for 5188 aligned nucleotides of chloroplast DNA. Here we add species from 11 genera for a matrix of 4156 aligned nucleotides. We also analyze 81 morphological characters in the context of the molecular phylogeny, using an extended version of the morphological and anatomical data. The resulting Bayesian phylogeny is well resolved, and most of the 47 larger clades also have morphological synapomorphies and ecological or geographic cohesion. Of the 47 clades, 19 new ones are provisionally named here. Relationships at higher taxonomic level are well resolved, while those in the relatively derived Aroideae subfamily remain poorly supported, with the position of *Calla* presenting the most glaring problem. Its placement in Aroideae conflicts with the distribution of morphological, anatomical and palynological characters. These results provide a firm basis for a new classification of the family, which we present here.

6.2 Introduction

The classification of the Araceae has long been an active area of research and its recent history has been summarized in various publications (Nicolson, 1987; Grayum, 1990; Mayo et al., 1997; Keating, 2002; Croat, 2004; Bogner and Petersen, 2007; Cabrera et al., 2008). The monograph by Mayo, Bogner and Boyce (1997: The Genera of Araceae - GoA) covers the morphology of most of the currently recognized genera except the Lemnoideae, and is the most complete generic treatment since A. Engler's classification (1876), and his monographs for deCandolle's *Monographiae Phanerogamarum* (1879) and the series *Das Pflanzenreich* (Engler, 1905-1920). To provide the framework for the classification presented in the 1997 GoA book a maximum parsimony cladistic analysis using a matrix of 63 morphological and anatomical characters for 106 genera assembled from literature and the examination of living and herbarium material during the preparation of

generic descriptions was undertaken, the first maximum parsimony study based on morpho-anatomical data. It was motivated by 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, providing the basis for a classification, which he produced using informal cladistic methodology. This GoA study was carried out in parallel, but independently of that of French et al. (1995), who were the first authors to publish a computer-generated cladistic analysis of Araceae genera based on a molecular matrix. Both these cladistic analyses 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 the possibility of a combined analysis. This was realised in Chapter 21 of GoA (Mayo et al., 1997) presenting the classification adopted by combining results from the molecular (French et al., 1995) and the morpho-anatomical analyses. Although the latter analysis was also presented orally at a symposium of the Tokyo International Botanical Congress in 1993, it was never published in full.

Since the first study of French et al. in 1995, as in other families of Angiosperms, molecular data have been mainly used for phylogenetic studies and the basis for significant changes (Barabé et al., 2002; Renner et al., 2004; Renner and Zhang, 2004; Rothwell et al., 2004; Tam et al., 2004; Gonçalves et al., 2007; Cabrera et al., 2008; Gauthier et al., 2008). The most comprehensive analysis to date regarding the whole family has been provided by Cabrera et al. (2008), and effectively settled the long-standing question of the relationships of the duckweeds (the former Lemnaceae, now Araceae subfamily Lemnoideae). These data and that of French et al. (1995) got recently available. Additionally, over the years, the morpho-anatomical data set of the GoA has been expanded to 81 characters, sampled for nearly all of the 109 genera. This led us to redo a analysis based on a comparison and combination of these expanded and new morphological/anatomical and molecular data sets, to get, more than ten years after the publication of the GoA, a step towards a new formal classification. We therefore augmented the plastid DNA sequence data of Cabrera et al. (2008) with data from genera previously missing, resulting in a complete genus sampling of Araceae and (re-) analysed the three data sets separately. The main discussion concentrates on the well-supported molecular phylogeny in combination with the expanded morphological-anatomical

data set: With this combination of data we found 47 clades, which are strongly supported by molecular data and/or well characterised morphologically, what reinforces proposals made by Cabrera et al. (2008) regarding the family classification. 19 of these clades are newly presented here, yet considered as informal taxonomic groups. All early diverging subfamilies and the relationships among them are well-supported. Also the subclades within the Aroideae subfamily are all well-supported. However, relationships among the Aroideae subclades are still not resolved. The only case, in which molecular and morphological data contradict, is the position of *Calla* within the Aroideae.

6.3 Material and Methods

Character Matrix and Data Analyses

During the preparation of the genus descriptions for GoA, the morphology and anatomy of the stem, leaf, inflorescence, fruits, and seeds were re-examined using existing taxonomic literature supplemented by observations made from specimens in the herbaria 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 more recently published genera not included in GoA. The morphological and anatomical characters comprising the matrix presented here are described in Appendix 1. Where no references are given, GoA (Mayo et al., 1997) 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 2; for downloadable versions of both matrices see `http://scratchpad.cate-araceae.org/`).

Table 6.1: GenBank numbers and sources of the sequences of the newly added Araceae taxa.

Species	*matK*	*rbcL*	*tRNA-LEU*	*trnL-F* spacer, *tRNA-PHE*	Authors
Anaphyllum	new	-	-	-	this paper
Asterostigma cubense	EF173531	-	EF173566	-	Gonçalves et al., 2007
Bakoa lucens	GQ220894	-	-	GQ220962	Wong et al., unpublished
Croatiella inegrifolia	EF173538	-	EF173573	-	Gonçalves et al., 2007
Furtadoa	new	-	-	-	this paper
Incarum pavonii	EF173548	-	-	-	Gonçalves et al., 2007
Philonotion americanum	GQ220908	-	-	GQ220978	Wong et al., unpublished
Schottariella sarikeense	GQ220912	-	-	GQ221009	Wong et al., unpublished
Theriophonum dalzelii	EU886534	-	AY249011	AY248973	Cusimano&Renner, 2009; Renner&Zhang, 2004
Typhonium brownii	EU886538	-	-	-	Cusimano&Renner, 2009
Typhonium hirsutum			AY249014	AY248976	Renner&Zhang, 2004
Typhonium horsfieldii	EU193593	EU193202			Mansion et al., 2008
Zomicarpa steigeriana	EU542592	-	-	-	Batista et al., DS

Parsimony analyses of the morphological data set in PAUP version 4.0 (Swofford, 2002) used heuristic searches, TBR swapping without the steepest descent option. Ten random taxon addition replicates were used when the number of most parsimonious trees was limited, otherwise, when no limit was reached, searching was halted at 10,000 trees and these were branch-swapped to exhaustion. Only for this analysis, multiple states were chosen to be variable and gaps were treated as new states. Differing combinations of weightings and deletions of both taxa and characters were applied in the analyses, but in the final ones presented here all 81 characters were used without weighting. For bootstrap analysis we ran 1000 replicates with the same settings as used in the tree searching analysis with MaxTrees set to 100 and with only one random taxon addition replicate.

The chloroplast restriction site data matrix of French et al. (1995) included 89 genera and 488 characters with *Acorus* as outgroup (see Appendix 3; for downloadable version see `http://scratchpad.cate-araceae.org/`). Parsimony and bootstrap analyses were conducted in the same way described above.

The sequences of the five 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 Araceae genera and seven outgroup taxa, were obtained from GenBank (for GenBank Numbers see Cabrera et al., 2008: Appendix 1). We downloaded the matrix from TreeBase and completed it by adding sequences from the six accepted genera not then included (*Anaphyllum, Croatiella, Furtadoa, Theriophonum, Zomicarpa*, and *Asterostigma*; the included *Asterostigma* species being now classified as *Incarum pavonii*), two recently published

genera, *Bakoa* and *Schottariella* (Boyce and Wong, 2008, 2009), and three other genera that will be resurrected or published in the near future: *Philonotion* (Wong and Boyce, 2010), *Sauromatum*, and an Australian genus composed of species previously assigned to *Typhonium* (most probably *Lazarum*, Cusimano et al., 2010), resulting in a total of 113 genera (Table 6.1). We also added a second sample of *Calla* as a check, and used *Hedyosmum* (Chloranthaceae) as outgroup. Table 6.1 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 sequences were available in Genbank, and the new sequences were generated according to the methods described in Cabrera et al. (2008), and deposited in GenBank (Accessions No. xxx to xxx, Table 6.1).

Data were analysed under maximum likelihood (ML) with RAxML (Stamatakis et al., 2008) and with a Bayesian approach using BEAST (Drummond et al., 2006; Drummond and Rambaut, 2007). Gap coding was not used because methods based on ML models could not then have been applied. In ML analysis we chose the GTRCAT + I + Γ model implemented in RAxML for calculation of the best tree and bootstrap analysis; bootstrap values were obtained by running 1000 replicates. In the Bayesian analysis base frequencies were estimated, number of Gamma categories was set to four, proportion of invariable sites was estimated, the mean substitution rate was not fixed and the Yule model was chosen as model for speciation. Prior to the analysis we only constrained two monophyletic groups: the ingroup with all species except for *Hedyosmum*, and the nine species of the Areae. The posterior probabilities were obtained from 7700 trees by running 19,250,000 generations and sampling every 2500th (after running 1,925,000 generations as pre-burnin).

Evaluation

The names of suprageneric groups used in this paper refer to taxa recognized by Bogner and Petersen (2007), unless specifically referenced to other published classifications or synopses (e.g., Mayo et al., 1997; Keating, 2002, 2004). When discussing other groups, especially clades emerging from the present analyses which have posterior probability values greater than 0.97 (Fig. 6.1) and do not correspond

precisely to previously recognized taxa, we use the numbers and informal names given in Table 6.2. The three phylogenies resulting from the three data matrices (morphological, restriction site and sequence data) were described according to the classifications mentioned above and compared with each other, with the original phylogenies presented by French et al. (1995) and Cabrera et al. (2008). Finally, to compare the information of the morphological characters with the results from molecular data analysis we mapped morphological characters onto the sequence data phylogeny.

6.4 Results and Discussion

Phylogenetic Analyses

Morphological Data

The analysis of the morphological data matrix of 109 taxa and 81 characters resulted in a 90% majority rule consensus tree based on the first 10,000 most parsimonious trees (Table 6.2). The phylogeny placed the duckweeds (Lemnoideae) as the first diverging clade of the Araceae. Somewhat surprisingly, *Calla* is placed as sister to the Lemnoideae. The Orontioideae was recovered as a clade, but *Gymnostachys* did not cluster with it to form the Proto-Araceae clade detected by French et al. (1995), and recognized by GoA. The analysis failed to reveal a number of currently accepted groupings (following Bogner and Petersen, 2007) such as 1) Pothoideae (*Anthurium* and *Potheae*; 2) Monsteroideae – Spathiphylleae are separated from the other genera of Monsteroideae; 3) Areae and 4) Caladieae. The tree also expresses some old higher groupings, which were recognized in earlier, pre-molecular classifications, such as the association of the Thomsonieae and Nephthytideae (Engler, 1920), Cryptocoryneae with *Ambrosina, Arisarum, Arisaema, Pinellia* and the Areae (Bogner and Nicolson, 1991), and the clustering of the genera of the Caladieae and Colocasieae (Engler, 1920). *Stylochaeton* does not cluster with the Zamioculcadeae, reflecting their very different morphology.

Restriction Site Data

We reanalysed the chloroplast restriction site data of French et al. (1995) adding a bootstrap analysis that was not previously available. From the resulting 2047 most parsimonious trees a 90% Majority Rule consensus tree was computed. This phylogeny (OSM Fig. S1) does not differ substantially from the original published one. Bootstrap support is high for the True Araceae, and for many lower clades (unless otherwise indicated taxa follow Bogner and Petersen (2007): Monsteroideae, Lasioideae, Cryptocoryneae, Schismatoglottideae, Caladieae (sensu Keating, 2002), Thomsonieae, Arophyteae, the *Pistia* clade (Renner and Zhang, 2004), Areae, *Stylochaeton* + Zamioculcadoideae, *Philodendron* + Homalomeneae, Culcasieae, Spathicarpeae (sensu Gonçalves, 2002; Gonçalves et al., 2007), Nephthytideae, Aglaonemateae) whereas the remaining backbone has no support at all. The only difference is that *Pothos* and *Anthurium* are not sister to the Monsteroideae, but form one clade in a trichotomy with Monsteroideae and the rest of the Araceae. The next branching clade is the Lasioideae followed by *Calla*. As in the original tree (French et al., 1995), Zamioculcadoideae are embedded within the Aroideae, making the latter paraphyletic. *Lemna* is also embedded within the Aroideae, but in a position distant from *Pistia*. As in the molecular sequence data, (see below) *Alocasia* does not group with the other genera of the Colocasieae (sensu Bogner and Petersen, 2007) .

Sequence Data

The matrix of the four combined chloroplast markers consisted of 115 taxa and 4156 aligned characters. Major indels were excluded in all markers except for *rbcL* (1391 nt) for analysis, especially in the *trnL–trnF* genes and spacer, which were highly variable. Alignment length before and after gap exclusion were: 1122 / 537 nt (*tRNA-LEU*), 728 / 458 nt (*trnL-F* spacer, *tRNA-Phe*), 1965 / 1770 nt (*trnK*). Nevertheless, several indels were not excluded as they include a high number of informative characters. Analyses of the molecular data yielded similar topologies for the ML and the Bayesian approaches. However, whereas the ML analysis failed to resolve the backbone of the phylogeny (Fig. S2), in the Bayesian approach it is well resolved and well-supported. For this reason we chose the Bayesian tree (Fig. 6.1) for discussion and character mapping. We used only *Hedyosmum* as outgroup

Table 6.2: Taxonomical accepted and new (bold) names of the 47 clades with the numbering used for Fig. 6.1. For each clade is indicated if posterior probability obtained from the Bayesian analysis is < 0.97.

Clade No.	Group name	Genera included	PP > 0.97
1	Orontioideae (Bogner & Petersen, 2007)	*Lysichiton, Orontium, Symplocarpus*	yes
2	Lemnoideae (Keating, 2002)	*Landoltia, Lemna, Spirodela, Wolffia, Wolffiella*	yes
3	Potheae (Bogner & Petersen, 2007)	*Pedicellarum, Pothoidium, Pothos*	yes
4	*Heteropsis* clade (Tam et al., 2004; Cabrera et al. 2008)	*Alloschemone, Heteropsis, Rhodospatha, Stenospermation*	yes
5	Spathiphylleae (Bogner & Petersen, 2007)	*Holochlamys, Spathiphyllum*	yes
6	*Rhaphidophora* clade (Tam et al., 2004; Cabrera et al., 2008)	*Amydrium, Anadendrum, Epipremnum, Monstera, Rhaphidophora, Scindapsus*	yes
7	Lasioideae (Bogner & Petersen, 2007)	*Anaphyllopsis, Anaphyllum, Cyrtosperma, Dracontioides, Dracontium, Lasia, Lasimorpha, Podolasia, Pycnospatha, Urospatha,*	yes
8	Zamioculcadoideae (Bogner & Petersen, 2007)	*Gonatopus, Zamioculcas*	yes
9	Aglaonemateae (Bogner & Petersen, 2007)	*Aglaodorum, Aglaonema*	yes
10	Nephthytideae (Bogner & Petersen, 2007)	*Anchomanes, Nephthytis, Pseudohydrosme*	yes
11	Culcasieae (Bogner & Petersen, 2007)	*Cercestis, Culcasia*	yes
12	***Philodendron* clade**	*Furtadoa, Homalomena, Philodendron*	yes
13	Spathicarpeae (Gonçalves, 2002,; Gonçalves et al., 2007)	*Asterostigma, Bognera, Croatiella, Dieffenbachia, Gearum, Gorgonidium, Incarum, Mangonia, Spathantheum, Spathicarpa, Synandrospadix, Taccarum*	yes
14	Cryptocoryneae (Bogner & Petersen, 2007)	*Cryptocoryne, Lagenandra*	yes
15	Schismatoglottideae (Bogner & Petersen, 2007; Boyce & Wong, 2008)	*Aridarum, Bakoa, Bucephalandra, Phymatarum, Piptospatha, Schismatoglottis, Schottariella*	yes
16	Thomsonieae (Bogner & Petersen, 2007)	*Amorphophallus, Pseudodracontium*	yes
17	Caladieae (Keating, 2002)	*Caladium, Chlorospatha, Filarum, Hapaline, Jasarum, Scaphispatha, Syngonium, Ulearum, Xanthosoma, Zomicarpa, Zomicarpella*	yes
18	Arisareae (Keating, 2002)	*Ambrosina, Arisarum*	yes
19	Arophyteae (Bogner & Petersen, 2007)	*Arophyton, Carlephyton, Colletogyne*	yes
20	Peltandreae (Bogner & Petersen, 2007)	*Peltandra, Typhonodorum*	no
21	***Colocasia* clade**	*Ariopsis, Colocasia, Remusatia, Steudnera*	yes
22	Areae (Bogner & Petersen, 2007)	*Arum, Biarum, Dracunculus, Eminium, Helicodiceros, Lazarum, Sauromatum, Theriophonum, Typhonium*	yes
23	Proto-Araceae (Mayo, Bogner & Boyce, 1997)	*Gymnostachys, Lysichiton, Orontium, Symplocarpus*	yes
24	Pothoideae (Bogner & Petersen, 2007)	*Anthurium, Pedicellarum, Pothoidium, Pothos*	yes
25	Monsteroideae (Bogner & Petersen, 2007)	*Clades 4, 5 , 6*	yes
26	***Stylochaeton* clade**	*Stylochaeton, Gonatopus, Zamioculcas*	yes
27	***Anchomanes* clade**	*Clades 9, 10*	yes
28	***Homalomena* clade**	*Clades 11, 12*	yes
29	**Rheophytes clade**	*Philonotion, clades 14, 15*	yes
30	***Typhonodorum* clade**	*Clades 19, 20*	yes
31	***Alocasia* clade**	*Alocasia, Arisaema, Pinellia, clade 22*	yes
32	**Bisexual climbers clade**	*Clades 24, 25*	yes
33	***Zantedeschia* clade**	*Zantedeschia, Clades 13, 27, 28*	yes
34	***Colletogyne* clade**	*Clades 18, 30*	yes
35	*Pistia* clade (Renner & Zhang, 2004)	*Pistia, Protarum, clades 21, 31*	yes
36	***Amorphophallus* clade**	*Clades 16, 17*	yes

37	***Ambrosina* clade**	*Clades 34, 35*	yes
38	**Spadix appendix clade**	*Clades 36, 37*	yes
39	***Calla* clade**	*Calla, clade 38*	no
40	***Philonotion* clade**	*Calla, clades 29, 39*	yes
41	***Montrichardia* clade**	*Anubias, Montrichardia, clades 33, 40*	no
42	*Aroideae* clade	*Callopsis, clade 41*	yes
43	Unisexual flowers clade (Mayo, Bogner & Boyce, 1997)	*Clades 26, 42*	no
44	***Podolasia* clade**	*Clades 7, 43*	yes
45	True Araceae clade	*Clades 32, 44*	yes
46	***Spirodela* clade**	*Clades 2, 45*	yes
47	Araceae	*Clades 23, 46*	yes
Monogeneric clades			
	Anthurium		
	Stylochaeton		
	Calla		
	Callopsis		
	Montrichardia		
	Anubias		
	Zantedeschia		
	Philonotion		
	Protarum		
	Pistia		
	Alocasia		
	Pinellia		
	Arisaema		

because inclusion of the other outgroup taxa made no difference. The second accession for *Calla* was used to confirm the veracity of the original sequence, and its position in the tree. The two *Calla* accessions form a well-supported clade.

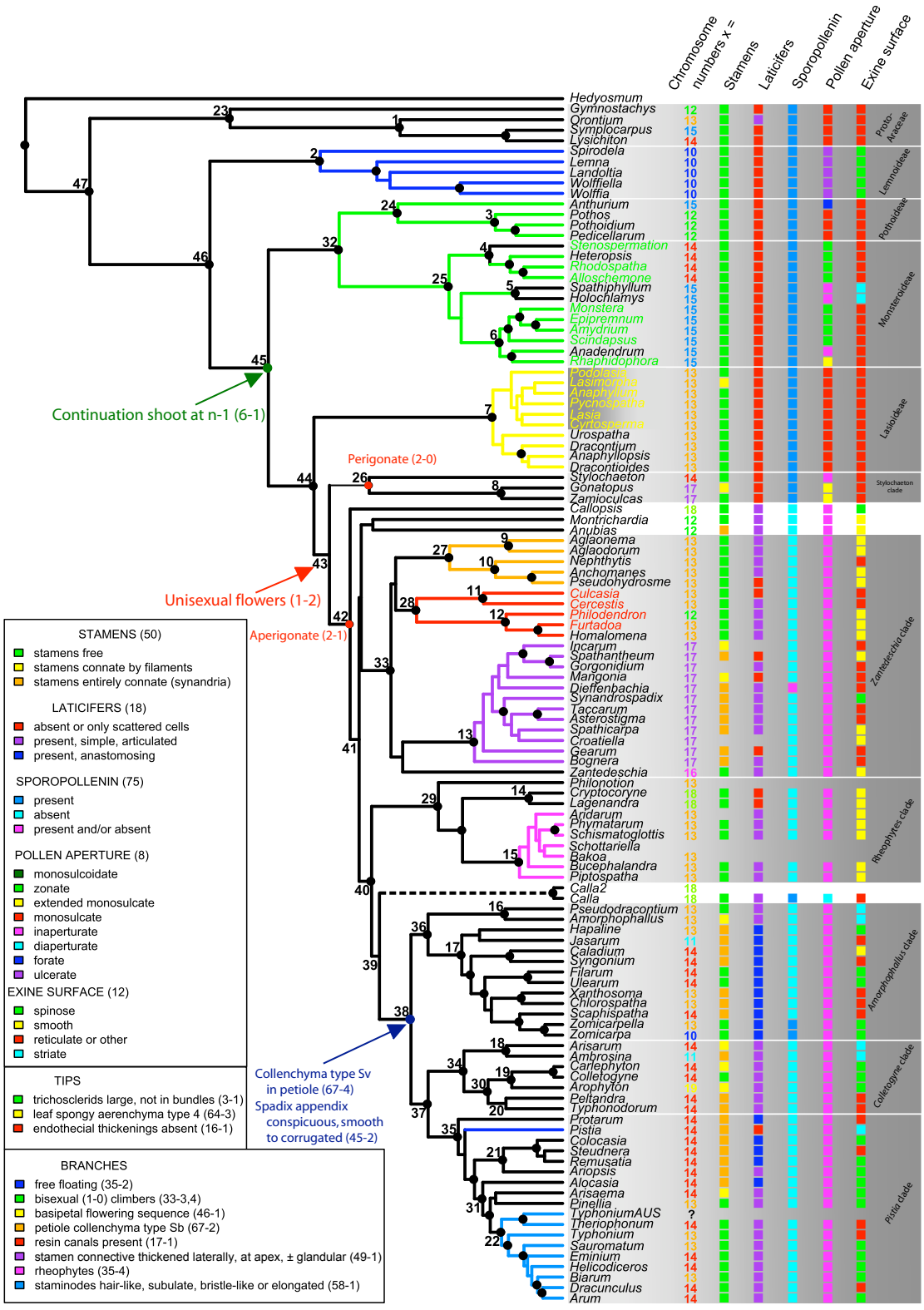
Our Bayesian summary tree, contrary to that of Cabrera et al. (2008) has a strong support for the backbone, and resolves the polytomy of the Lasioideae, *Stylochaeton*, Zamioculcadoideae, and Aroideae could be resolved to some extent: the first diverging clade is the Lasioideae (clade 7) followed by the *Stylochaeton* clade (clade 26), which is sister to the Aroideae (clade 42), although with low support (PP = 0.87), agreeing with the results of French et al. (1995) and supported by morphology (see below: *The molceular phylogeny in the context of morphology, anatomy and ecology*). In most of the other major features it is similar to that of Cabrera et al. (2008). Relationships between the subfamilies are well resolved. Within the Aroideae, the relationships between the major clades (*Callopsis, Anubias*, the Zantedeschia clade (clade 33), *Montrichardia*, and the *Philonotion* clade

(40); see Table 6.2 and Fig. 6.1 for clade/taxon circumscriptions) is still poor, clearly visible as very short branch length in the ML tree (Fig. S2). Within the *Zantedeschia* clade only the lower level clades are well-supported whereas the backbone is not. However, resolution within the Spadix Appendix clade (clade 38) is high. All the formerly missing, new and recently described taxa group with their hypothesized relatives based on morphology. As in the result of Cabrera et al. (2008), *Calla* falls well within the Aroideae clade, sister to the well-supported Spadix Appendix clade (clade 38, see further discussion below). In the Bayesian analysis with BEAST this grouping (*Calla* clade 39) has no significant support, but an additional analysis conducted with MrBayes supports this clade with a posterior probability of 1 (data not shown). Unlike the parsimony tree of Cabrera et al. (2008), our maximum parsimony analysis of the matrix (strict consensus of 1718 MP trees, data not shown) places *Calla* in the same position as the Bayesian analyses.

The Molecular Phylogeny in the Context of Morphology, Anatomy and Ecology

The most robust phylogeny was obtained from the Bayesian analysis of the molecular data set, based on the large number of characters (4156 nt) and a model-based method of analysis. We therefore favour this over the phylogenies from the other two data sets because the latter are much smaller (81 and 488 characters, respectively), could be analysed only with maximum parsimony, and have low statistical support. It is a common phenomenon that many characters that are of high importance for describing clades at lower taxonomic levels, become homoplasious when working at a higher taxonomic level. The number of morpho-anatomical charac-

Figure 6.1 ***(facing page)*:** Phylogeny obtained from Bayesian analysis of a molecular data set of 115 species and 4156 nucleotides from four chloroplast markers based on Cabrera et al. (2008); posterior probabilities are mean heights of 7700 trees from 19,250,000 generations; 19 morphological/anatomical characters are mapped by coloring branches or tips, or by plotting squares behind every taxon label with color coded-states (see legend; numbers in brackets are character numbers with the respective character state as used in Appendix 1); the dashed line highlights the position of Calla; numbers at nodes refer to the 47 clades defined below; grey boxes on the right: major clades of different taxonomic levels posterior probability > 0.97.


Chromosome numbers x =
Stamens
Laticifers
Sporopollenin
Pollen aperture
Exine surface
Hedyosmum
Gymnostachys
Qrontium
Symplocarpus
Lysichiton
Spirodela
Lemna
Landoltia
Wolffiella
Wolffia
Anthurium
Pothos
Pothoidium
Pedicellarum
Stenospermation
Heteropsis
Rhodospatha
Alloschemone
Spathiphyllum
Holochlamys
Monstera
Epipremnum
Amydrium
Scindapsus
Anadendrum
Rhaphidophora
Podolasia
Lasimorpha
Anaphyllum
Pycnospatha
Lasia
Cyrtosperma
Urospatha
Dracontium
Anaphyllopsis
Dracontioides
Stylochaeton
Gonatopus
Zamioculcas
Calla
Montrichardia
Anubias
Aglaonema
Aglaodorum
Nephthytis
Anchomanes
Pseudohydrosme
Culcasia
Cercestis
Philodendron
Furtadoa
Homalomena
Incarum
Spathantheum
Gorgonidium
Mangonia
Dieffenbachia
Synandrospadix
Taccarum
Asterostigma
Spathicarpa
Croatiella
Gearum
Bognera
Zantedeschia
Philonotion
Cryptocoryne
Lagenandra
Ariopsis
Phymatarum
Schismatoglottis
Schottariella
Bakoa
Bucephalandra
Piptospatha
Calla2
Calla
Pseudodracontium
Amorphophallus
Hapaline
Jasarum
Caladium
Syngonium
Filarum
Ulearum
Xanthosoma
Chlorospatha
Scaphispatha
Zomicarpella
Zomicarpa
Arisarum
Ambrosina
Carlephyton
Colletogyne
Arophyton
Peltandra
Typhonodorum
Protarum
Pistia
Colocasia
Steudnera
Remusatia
Ariopsis
Alocasia
Arisaema
Pinellia
TyphoniumAUS
Theriophonum
Typhonium
Sauromatum
Eminium
Helicodiceros
Biarum
Dracunculus
Arum
Proto-Araceae
Lemnoideae
Pothoideae
Monsteroideae
Lasioideae
Stylochaeton clade
Zantedeschia clade
Rheophytes clade
Amorphophallus clade
Colletogyne clade
Pistia clade
Continuation shoot at n-1 (6-1)
Perigonate (2-0)
Unisexual flowers (1-2)
Aperigonate (2-1)
Collenchyma type Sv in petiole (67-4)
Spadix appendix conspicuous, smooth to corrugated (45-2)
STAMENS (50)
stamens free
stamens connate by filaments
stamens entirely connate (synandria)
LATICIFERS (18)
absent or only scattered cells
present, simple, articulated
present, anastomosing
SPOROPOLLENIN (75)
present
absent
present and/or absent
POLLEN APERTURE (8)
monosulcoidate
zonate
extended monosulcate
monosulcate
inaperturate
diaperturate
forate
ulcerate
EXINE SURFACE (12)
spinose
smooth
reticulate or other
striate
TIPS
trichosclerids large, not in bundles (3-1)
leaf spongy aerenchyma type 4 (64-3)
endothecial thickenings absent (16-1)
BRANCHES
free floating (35-2)
bisexual (1-0) climbers (33-3,4)
basipetal flowering sequence (46-1)
petiole collenchyma type Sb (67-2)
resin canals present (17-1)
stamen connective thickened laterally, at apex, ± glandular (49-1)
rheophytes (35-4)
staminodes hair-like, subulate, bristle-like or elongated (58-1)

ters is rather low from a statistical point of view, and together with the relatively high level of homoplasy, accounts for the appearance of implausible groupings and low statistical support. The Bayesian phylogeny revealed 47 clades of different taxonomical levels for the Araceae (Fig. 6.1, Table 6.2). Of these, 19 clades are reported here for the first time and named informally. In the following we revealed 19 (out of the 81) characters describing morphology, anatomy or ecological preferences of the species that support the majority of the 47 clades, even those without support. These taxonomically relevant characters, discussed below and shown in Fig. 6.1, are: Occurrence and/or type of synandria (50), laticifers (18), collenchyma (67), trichosclerids (3), leaf spongy aerenchyma (64), stamen connective (49), staminodes (58), presence/absence of sporopollenin (75), endothecial thickenings (16), perigone (2) and resin canals (17), pollen aperture (8), exine surface (12), habit (33, 35), sexuality (1), flowering sequence (46), shoot architecture (6), and spadix appendix (45). Several others, non mapped characters, also support the phylogeny and are discussed below, whereas a high number do not and seem to result from independent, convergent evolution. In the following, numbers in brackets correspond to the clade numbers as described in Table 6.2.

Clades of Higher Taxonomic Level

The well-supported backbone of the molecular phylogeny defines four major clades (43-46) of high taxonomic level. Two are newly reported here, the *Spirodela* clade (46) and the *Podolasia* clade (44), two have been circumscribed before, the True Araceae (clade 45) and the Unisexual Flowers clade (43). These latter two clades are are also supported by morphological characters (see below). The *Spirodela* clade (46) includes all Araceae except for the Proto Araceae (clade 23). The *Spirodela* clade (46) includes the Lemnoideae as sister to the True Araceae (clade 45). That there does not exist any morphological character defining this group might be due to the extreme differentiation of the Lemnoideae because of their adaptation to an aquatic life form (see below). The True Araceae (clade 45) is characterized morphologically by shoot architecture (char. 6-1), namely the reiteration (continuation shoot) of the sympodial unit in mature stems arising in the axil of the penultimate foliage leaf (euphyll; Engler, 1877; translated by Ray and Renner, 1990). Only in the specialized climbing genera of Potheae, in *Heteropsis*,

in the *Schismatoglottis* Calyptrata Group and the *Homalomena bellula* complex is a different pattern observed within the True Araceae (char. 6-3). It remains to be investigated whether the predominant architectural model is adaptive and mediated some key evolutionary advantage. The *Podolasia* clade includes the Lasioideae as sister to the Unisexual flowers clade. This sister relationship between the two clades is not supported well by molecular data (see above), but from a morphological point of view, because the Unisexual Flowers clade (43) includes all species with unisexual flowers (char. 1-1). It corresponds to the Aroideae of GoA, which includes *Stylochaeton, Zamioculcas* and *Gonatopus*. *Calla* is the only member of the clade without this character.

Early Diverging Clades

The Proto-Araceae (clade 23) has no morphological support but is consistently found in all the molecular analyses (French et al., 1995; Tam et al., 2004; Cabrera et al., 2008). The morphology of *Gymnostachys* is quite unlike that of any other genus. This might be due to the long time the lineages diverged from each other, leading to morphologically highly divergent taxa. The duckweeds (clade 2, Lemnoideae) are also morphologically very distinct from all other Araceae, apart from the reduction in habit shared with *Pistia* associated with the free-floating life form (char. 35-2), which is highlighted further by the fossil genus *Limnobiophyllum* (Stockey et al., 1997; Rothwell et al., 2004; Bogner, 2009). Apart from the highly reduced structure, the Lemnoideae are supported by a chromosome base number of $x = 10$ (char. 57-5) and ulcerate pollen (char. 8-7). The evolution of the genera of the Lemnoideae has been thought to follow a logical sequence according to the following morphological reduction series: in *Spirodela* and *Landoltia* the fronds have veins and many roots, a prophyll is present, the inflorescence has a spathe and is situated at the side of the leaf sheath arising from the plant's growth point, the anther has two thecae; in *Lemna* the fronds have veins but only a single root, there is no prophyll but the inflorescence is similar to those of *Spirodela* and *Landoltia*; in *Wolffiella* the fronds lack both veins and roots, there is no prophyll, the inflorescence is situated on the upper side of the flat frond in a cavity, there is no spathe and the anthers have only a single theca; *Wolffia* differs from *Wolffiella* only in having globular to ellipsoid fronds. Our phylogeny (Figure 6.1) places

Spirodela, the species with the most ancestral phenotype, as sister to the other Lemnoid genera and groups *Wolffiella* and *Wolffia* as a well-supported subclade (Wolffioideae of Les et al., 2002; Wolffieae of Bogner and Petersen, 2007). However, the exact branching pattern between *Lemna* and *Landoltia* is not clear in our results; in the Bayesian analysis they are sister taxa but without support, and in the ML analysis (also unsupported) *Lemna* branches first, followed by *Landoltia* and the Wolffieae. The molecular data presented by Les et al. (2002) based on a comprehensive sampling of Lemnoid species also did not reveal the position of *Landoltia*. Only after adding morphological data to the analysis did it came out well-supported as sister to *Lemna, Wolffia* and *Wolffiella*. So it seems clear that there was evolution from more complex to the more reduced forms, perhaps twice, once leading to Lemna and once to the forms of *Wolffia* and *Wolffiella*.

The first-diverging clades within the True Araceae clade are the Pothoideae (clade 24) and Monsteroideae (clade 25), forming a robust monophyletic group called here the Bisexual climbers clade (32). Climbing habit is not a constant character within the clade – virtually all species in Spathiphylleae (clade 5) are terrestrial or helophytes, *Stenospermation* includes mainly perching epiphytes and most *Anthurium* are not climbers. Nevertheless, the climbing habit is predominant in twelve of the sixteen genera. The only other climbing genera are unisexual-flowered and occur in the *Homalomena* clade (*Philodendron, Culcasia, Cercestis* in clade 28), Caladieae (*Syngonium* in clade 17), and some species of Schismatoglottidae (clade 15). Furthermore, many *Alocasia* and Asian *Homalomena* species are functionally climbers extending the "rhizome" through the leaf litter layer for some considerable distance.

Within the Pothoideae (clade 24), *Anthurium* is differentiated from the Potheae (clade 3) by chromosome number (basic number $x = 15$ vs. $x = 12$; char. 57) and pollen aperture (forate vs. monosulcate; char. 8). Within the Monsteroideae (clade 25), the *Heteropsis* clade (clade 4) has a basic chromosome number of $x = 14$ (char. 57-2), whereas its sister clade, consisting of the Spathiphylleae (clade 5) and the *Rhaphidophora* clade (clade 6) and not supported by molecular data, has a basic number of $x = 15$. Although embedded, the Spathiphylleae differs from all other Monsteroideae by their combination of clustered trichosclereid structure (Keating, 2002), perigoniate flowers, and especially the polyplicate-multiaperturate rather than zona-aperturate pollen (Tarasevich, 1989; Hesse et al., 2000), perhaps as

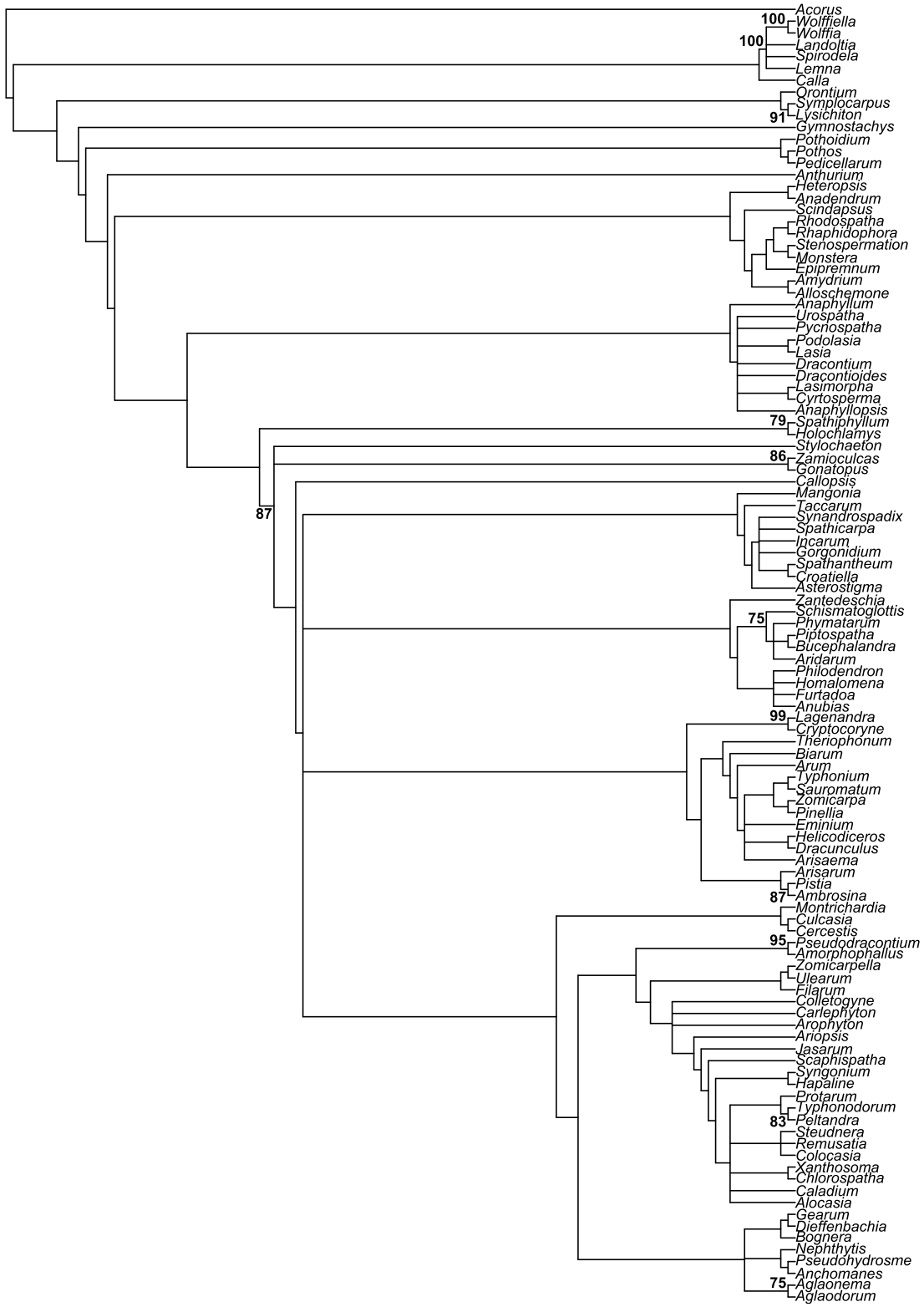
adaptation to their terrestrial (except *S. solomonense*), usually helophytic habit. Tam et al. (2004) that studied a much larger sample of species from the Monsteroideae with molecular methods, also found the Spathiphylleae embedded in the rest of the Monsteroideae and *Rhaphidophora* to be paraphyletic. However, support in their phylogeny is generally low.

The Lasioideae clade (7) is composed of tropical bisexual-flowered plants, which are mostly helophytes (char. 35-1); dryland species are found only in the genera *Dracontium* and *Pycnospatha*. Morpho-anatomical synapomorphies of this group are the basipetal flowering succession (char.46-1), a unique pollen aperture structure (Hesse, 2002), and a base chromosome number of $x = 13$ (char. 57-1). Lasioideae are also notable for the common occurrence of prickles on the petiole, peduncle, and main leaf veins (underside of leaf blade, char. 30-1) and deeply sagittate leaves in which the central vein of each basal lobe runs into the apex. This group has been extensively studied by Hay (1986, 1988); Hay and Mabberly (1991); Hay (1992).

Unisexual Flowers Clade

Clade 42 (Aroideae clade) is supported by the presence of both unisexual and aperigoniate flowers (char. 2-1), *Calla* again exccepted. Aperigoniate flowers also occur in few bisexual-flowered taxa, which must be the result of independent derivation: *Pycnospatha* is the only member of the Lasioideae (clade 7) with aperigoniate flowers; within the Monsteroideae (clade 25) the perigon may have been lost twice (in clades 4 and 6). Hesse (2006a,c) has discussed in some detail the significance of the switch to unisexual aperigoniate flowers in relation to important pollen characters. In clade 42 (again excepting *Calla*) pollen grains are always omniaperturate (= inaperturate), have a thick, spongy endexine and a highly reduced ektexine with either a very thin sporopollenin lamella or a non-sporopollenin outer exine layer (Hesse, 2006b). All genera diverging before clade 42 have aperturate pollen with a well-developed tectate-columellate sporopollenin

Figure 6.2 ***(facing page)*:** 90% majority rule consensus tree of 10000 most parsimonious trees obtained from maximum parsimony analysis of the morphological data matrix (81 char.) Values above nodes indicate bootstrap support values (1000 replicates).


Acorus
100
Wolffiella
Wolffia
100
Landoltia
Spirodela
Lemna
Calla
Orontium
Symplocarpus
91
Lysichiton
Gymnostachys
Pothoidium
Pothos
Pedicellarum
Anthurium
Heteropsis
Anadendrum
Scindapsus
Rhodospatha
Rhaphidophora
Stenospermation
Monstera
Epipremnum
Amydrium
Alloschemone
Anaphyllum
Urospatha
Pycnospatha
Podolasia
Lasia
Dracontium
Dracontioides
Lasimorpha
Cyrtosperma
Anaphyllopsis
79
Spathiphyllum
Holochlamys
Stylochaeton
86
Zamioculcas
Gonatopus
Callopsis
Mangonia
87
Taccarum
Synandrospadix
Spathicarpa
Incarum
Gorgonidium
Spathantheum
Croatiella
Asterostigma
Zantedeschia
Schismatoglottis
75
Phymatarum
Piptospatha
Bucephalandra
Aridarum
Philodendron
Homalomena
Furtadoa
Anubias
99
Lagenandra
Cryptocoryne
Theriophonum
Biarum
Arum
Typhonium
Sauromatum
Zomicarpa
Pinellia
Eminium
Helicodiceros
Dracunculus
Arisaema
Arisarum
Pistia
87
Ambrosina
Montrichardia
Culcasia
Cercestis
95
Pseudodracontium
Amorphophallus
Zomicarpella
Ulearum
Filarum
Colletogyne
Carlephyton
Arophyton
Ariopsis
Jasarum
Scaphispatha
Syngonium
Hapaline
Protarum
Typhonodorum
83
Peltandra
Steudnera
Remusatia
Colocasia
Xanthosoma
Chlorospatha
Caladium
Alocasia
Gearum
Dieffenbachia
Bognera
Nephthytis
Pseudohydrosme
Anchomanes
75
Aglaonema
Aglaodorum

ektexine and a thin endexine (Hesse, 2006b). Hesse also noted a correlation with anatomical characters possibly connected with chemical defence such as the presence of laticifers and biforines (char. 81-1). Clade 42 includes nearly all genera with laticifers, either simple articulated (char. 18-1) or anastomosing (char. 18-2). Exceptions are the presence of simple laticifers in *Orontium*, in the Proto-Araceae and their absence in *Cryptocoryne, Culcasia, Gearum, Lagenandra, Mangonia, Pistia, Pseudohydrosme*, and *Spathantheum*. *Calla* has simple laticifers. Biforines are found almost exclusively in clade (42) but patchily so, and not in *Calla*; outside clade 42 they are present only in *Stylochaeton*. Taken together these changes in morpho-anatomical character patterns seem to imply a major adaptive shift in the evolution of the family.

The sister group position of the *Stylochaeton* clade (clade 26) supports this with the intermediate condition of unisexual but perigoniate flowers. *Stylochaeton* combines unisexual with perigoniate flowers, but differs from its sister group the Zamioculcadeae (clade 8) in its omniaperturate pollen and much thinner, undifferentiated sporopollenin extexine (Hesse et al., 2001), the latter character also differentiating *Stylochaeton* from clade 42. *Callopsis* emerges as an isolated genus (as in previously published analyses, e.g., GoA) but with a robust sister-relation to the *Montrichardia* clade (clade 41) comprising all other Aroideae. Weber (2004) has observed a unique cuticle-like layer in the pollen wall of *Callopsis*.

The *Zantedeschia* clade (33), *Anubias, Montrichardia* and the Rheophytes clade (29) together correspond quite well to the distribution of smooth pollen (char. 12-1) suggesting that this pollen type evolved from the predominantly reticulate pattern at around the same time as the shift from bisexual to unisexual flowers. In the clades distal to the Rheophytes, spinose pollen becomes much more frequent (Hesse, 2006b). The *Zantedeschia* clade includes many of the genera assigned to subfamily Philodendroideae by Keating (2002, 2004). There are no obvious morpho-anatomical synapomorphies, but *Zantedeschia* is here brought into a more consistent relationship, although still essentially part of a polytomy because of lack of support in the spine of this clade. A notable subclade is the *Homalomena* clade (clade 28), which combines the Culcasieae (clade 11) and the *Philodendron* clade (12). Although supported only by the MrBayes analysis of our molecular sequence data (not shown), clade 28 is supported by anatomical characters observed by French (1985a, 1987a,b), including the occurrence of sclerotic

hypodermis and resin canals in the roots and absence of endothecial thickenings in the anthers. The recent molecular phylogeny of *Philodendron* by Gauthier et al. (2008) has confirmed the very close relationship between this genus and neotropical *Homalomena* species.

The Spathicarpeae (clade 13) have been thoroughly studied by Gonçalves (2002, Gonçalves et al., 2007). This provided strong molecular support for the inclusion of *Bognera* and *Dieffenbachia* (former Dieffenbachieae), which are vegetatively distinct from the other genera of the tribe. Synandria (char. 50-1 or -2) have evidently evolved in the Spathicarpeae independently of those in *Anubias* and in clades 17 (Caladieae) and 37 (*Ambrosina* clade). Chromosome base number $x = 17$ (char. 57-8) supports clade 13 uniformly, occurring also in the Zamioculcadoideae (clade 8), *Anubias, Montrichardia* and *Philodendron*.

The Rheophytes clade (29), corresponding closely to subfamily Schismatoglottidoideae of Keating (2002, 2004), includes the newly resurrected genus *Philonotion* following Boyce and Wong (2009), who are researching the phylogeny and systematics of this clade. Clade 29 has emerged consistently in molecular analyses (Cabrera et al., 2008; French et al., 1995) and includes a high concentration of rheophytes and aquatics (char. 35-3 or -4).

The Spadix Appendix clade (38) corresponds to subfamily Aroideae of Keating (2002, 2004) and is composed of the *Amorphophallus* clade (36) and the *Ambrosina* clade (37). Sterile terminal appendices (osmophores, Vogel, 1963; char. 45-1 or -2) are common in this clade, although many genera (14 out of 37) lack this character. Elsewhere sterile appendices are found in several genera of Schismatoglottideae (clade 15, *Aridarum, Bakoa, Bucephalandra, Schismatoglottis, Phymatarum*) and in a few *Homalomena* species. The Spadix Appendix clade also includes most genera with spinose pollen (char. 12-2; see also remarks by Hesse, 2006c), although there are other patterns present. The spinose pollen of the Lemnoideae must have arisen independently.

Four well-supported subclades make up the Spadix Appendix clade at a lower level: the Thomsonieae (clade 16), the expanded Caladieae sensu Keating (2002; 2004; clade 17), the *Colletogyne* clade (34), which includes the Arisareae (clade 18), Arophyteae (clade 19) and the Peltandreae (clade 20), and the *Pistia* clade (35) of Renner and Zhang (2004). Synandria, i.e., the androecium fused into a single structure, are common in these clades, predominating in clades 17, 19, 20,

Protarum, Pistia, clade 21 and *Alocasia*. Nearly all *Arisaema* species have at least partially fused stamens, whereas *Pinellia* and the Areae (clade 22) have free stamens; this result would be interesting to investigate with floral developmental studies. Similarly, anastomosing laticifers (char. 18-2) are present throughout the Caladieae (clade 17), the *Colocasia* clade (clade 21, except *Ariopsis*), and in *Protarum* and *Alocasia*, suggesting the possibility of a secondary derivation of simple laticifers in the Areae (clade 14), *Pinellia* and *Arisaema*.

The distinctive patterns of occurrence of petiole collenchyma tissue in the family, reported and discussed by Keating (2000, 2002, 2004) and Gonçalves et al. (2004), fit the Bayesian tree quite well, confirming their phylogenetic significance. The distribution of types B, Bi and Sb (Keating, 2000, his Table 3) suggests that the replacement of sclerenchymatous mechanical tissue by collenchyma, at least in the central portion of the petiole, occurred in conjunction with the appearance of the unisexual-flowered aroids, and later specialized into a single type (Sv, in the petiole, char. 67-4). In the independent report by Gonçalves et al. (2004) the collenchyma types are characterized as philodendroid (= types B, Bi, Sb) or colocasioid (= type Sv). Colocasioid collenchyma is characteristic of the Spadix Appendix clade (38), but also occurs in Cryptocoryneae (clade 14) while philodendroid collenchyma predominates in the Schismatoglottideae (clade 15).

Adaptation to Water-Associated Habitats

As previously noted by Cabrera et al. (2008), water-associated life forms of various kinds occur throughout the phylogeny of the Araceae in all major clades, even in the the Areae (clade 22, *Typhonium flagelliforme*). Individual aquatic or helophytic genera are often found embedded in otherwise non-aquatic clades, such as *Jasarum* (Caladieae, clade 17), *Aglaodorum* (clade 9), some species of *Dieffenbachia* (Spathicarpeae, clade 13), *Homalomena* in the *Philodendron* clade (12), *Peltandra* and *Typhonodorum* in the *Colletogyne* clade (34), and *Pistia* in the *Pistia* clade (35). Even in the Bisexual climbers clade (32) which is dominated by hemi-epiphytes and epiphytes, the Spathiphylleae (clade 5) stand out as a largely helophytic group. There are also several instances of genera that are both helophytic and rather isolated, failing to group consistently in most analyses hitherto: *Anubias, Montrichardia* and *Calla*. Clades, which are entirely or mostly

aquatic are the Orontioideae (clade 1), the Lemnoideae (clade 2), the Lasioideae (clade 7) and the Rheophytes (clade 29). It seems likely that a major theme in the phylogeny of the Araceae has been a repeated evolution to and from aquatic life forms, either to become more extreme aquatics or towards dry land geophytes and epiphytism of various kinds.

Towards a New Formal Classification

For the first time, analyses of a molecular sequence dataset with complete genus sampling of the Araceae is presented. The comprehensive morphological-anatomical data set and the re-analysis of the plastid restriction site data of French et al. (1995) have provided an opportunity to compare phylogenies arising from these data sets. The clades which are both well characterized morphologically, and strongly supported by molecular data, are highlighted by correlating the Bayesian phylogeny (Fig. 6.1) with "critical" morphological characters that have always been considered as taxonomically important, together with some new ones.

Using this approach we have defined 47 clades (Table 6.2) that could be considered as elements of a new formal classification. The majority has been described earlier, e.g., all subfamilies except for the Aroideae as circumscribed lastly by Bogner and Petersen (2007, *Calla* is included and *Stylochaeton* not) and lower taxonomic entities. 19 clades are circumcribed here for the first time (Table 6.2, Fig. 6.1) that are of higher taxonomic level in the bisexual taxa, and subdivide the Unisexual Flowers clade in major clades. Cabrera et al. (2008) gave a very detailed and complete discussion of the results of their keynote phylogenetic study, the implications for a revised classification, and probable evolutionary pathways of the Araceae, especially in relation to aquatic adaptation. Our re-analysis of an augmented version of their DNA sequence data set and the correlation of morphological characters supports most of their taxonomic proposals. Our results differs from theirs in the relative positions of the *Stylochaeton* clade (clade 26), and the Lasioideae (clade 7, Fig. 6.1): The Lasioideae are sister to the Unisexual Flowers clade (clade 43) including clade 26 as the sister of the Aroideae clade (clade 42). The Unisexual Flowers clade has no support but occurs in all analyses and is more parsimonious from an evolutionary point of view because then all taxa with unisexual aperigoniate flowers form a single clade (when excluding *Calla*).

Except for the morphologically well-supported Unisexual Flowers clade, only three other clades have no statistical support, but appear in all molecular analyses: the Peltandreae (clade 20) that have been revealed in further molecular studies as monophyletic (Renner and Zhang, 2004; Cusimano et al., 2008) and are also supported by morphology; the *Montrichardia* clade (clade 41) and *Calla* clade (clade 39). The latter two consequently have no support at all. Further molecular studies are needed to test if the *Montrichardia* clade is a natural entity in its present circumscription. That *Calla* is included and even nested high up in the Unisexual Flowers clade (clade 43) as sister to the Spadix Appendix clade (38) is in our opinion highly dubious due to several reasons:

Each of the three phylogenies places this genus in a different position. That based on morpho-anatomical data places *Calla* as the sister group of the duckweed genera; that from the restriction site data of French et al. (1995, Fig. S1) places it as sister to the unisexual-flowered clade (i.e., subfamily Aroideae sensu Mayo et al., 1997); and the phylogenies based on molecular sequence data (maximum parsimony, ML and Bayesian inference) place it within the Aroideae (Fig. 6.1). This lack of agreement between the three data sets regarding *Calla* is striking given that all the data sources used reflect, directly or indirectly, different sampling of the genomic diversity. Although the result from the Bayesian analysis of the molecular sequence data set is by far the most robust, its placing of *Calla* strongly jars with morphological character patterns that offer the possibility for ecological insights into the evolution of the unisexual-flowered aperigoniate Araceae (see Hesse, 2006b,a,c). Since this now appears to be the key event in aroid evolution, this is a problem that merits further study. The most important characters involved appear to be as follows. 1) *Calla* has aperturate (bicolpate) pollen, a massive tectate sporopollenin ektexine, thin endexine, and bisexual flowers. This character combination contradicts the otherwise almost uniform occurrence within the Aroideae clade (42) of omniaperturate (= inaperturate) pollen, highly reduced or absent sporopollenin ektexine structure, thickened endexine and unisexual flowers. As proposed by Hesse (2006a,c) it is likely that these characters are linked to a major evolutionary shift in floral ecology in which the transition from bisexual to unisexual flowers played a significant role, and led to the major crown radiation of the aroids – more than 65% of the family's genera belong to this clade. Accepting the inclusion of *Calla* within the Aroideae clade as sister to the Spadix

Appendix clade (38), requires bisexual flowers, aperturate pollen and a massive tectate sporopollenin ektexine to have re-evolved from unisexual-flowered, omniaperturate, sporopollenin-less ancestors. Although sporopollenin occurs in the exine of a few genera in the Aroideae clade, e.g., in the spines of *Remusatia* and *Zomicarpella*, it is in a form quite different from the tectate ectexine of bisexual-flowered genera. 2) The pattern of collenchyma distribution (Keating, 2000, 2002) argues against the placement of *Calla* in the Aroideae, since collenchyma is absent in the genus, whereas the unisexual clade is characterized by possession of type B, Bi, Sb or Sv collenchyma. The position of *Calla* next to the Lemnoideae in the morpho-anatomical tree (Fig. 3) is also anomalous given striking phenotypic differences such as the creeping (not free-floating) habit, smooth (not ribbed) testa, and dicolpate rather than ulcerate pollen. In our opinion, the morpho-anatomy of *Calla* suggests a position probably in the "transition zone" between the bisexual taxa and unisexual clades like the *Stylochaeton* clade (26). Resolving the position of *Calla* has now become an important issue for Araceae systematics.

The present study is the most comprehensive yet presented as regards different classes of data and is a step towards a new formal classification. A few important outstanding questions remain. Most of the major clades are well-supported, as well as their relationships among each other. It seems very likely that *Callopsis, Anubias, Montrichardia, Calla, Alocasia, Protarum* and *Pistia* are rather isolated and would best be treated as monogeneric higher taxa. It has so far proved impossible with molecular markers to clarify the branching pattern of *Protarum, Pistia*, and the rest of the *Pistia* clade. The relationship of *Arisaema, Pinellia*, and the Areae (clade 22) are another unresolved problem that emerged in earlier molecular analyses (Renner et al., 2004; Renner and Zhang, 2004) and is confirmed here. Contrary to recent classifications (Keating, 2004; Bogner and Petersen, 2007) *Arisaema* and *Pinellia* do not cluster into a unique clade by molecular data.

In addition to establishing a consistent position for *Calla*, the main focus of further analyses dealing with the classification of Araceae genera should be on the relationships of clades within the Aroideae. Analyses of phylogenetic relationships within clades at a finer taxonomic scale may reveal further genera or suppress others, but given the generally high level of agreement between the earliest molecular analysis (French et al., 1995) and the one presented here, based largely on Cabrera et al. (2008), it would be surprising if new work contradicted the general cladis-

tic patterns of aroid phylogeny as now understood; new work with whole genome analyses is nevertheless to be eagerly awaited.

A more likely source of phylogenetic novelty is the discovery of new aroid fossils. In recent years this has become an active and exciting field, yielding a number of remarkable finds with important implications not only for Araceae but for the evolution of Monocots as a whole (e.g., Smith and Stockey, 2003; Friis et al., 2004; Bogner et al., 2005; Stockey et al., 2007; Herrera et al., 2008). The importance of fossils further emphasizes the need for greater activity in the comparative study and classification of phenotypic character data of extant species, in order to be able to analyse the phylogenetic position of fossils with greater sophistication. Particularly important character fields are leaf venation, seed structure and pollen structure. While the latter two areas have received important studies in recent years (e.g., Grayum, 1992; Seubert, 1993; Hesse, 2002; Tillich, 2003; Hesse, 2006b,a,c), comparative leaf venation has generally been neglected since the monograph by Ertl (1932).

The morpho-anatomical matrix presented here (Appendices 1, 2) is a compilation from many different sources but expresses only very approximately the structural variability of the family. The spectrum of character variability and the character analyses employed are likely to change considerably with new research. New initiatives on the Internet have brought about the possibility to collectively build a more complete database, with entries fully documented to specimens and images and fully credited to every contributor. We hope that the compilation and electronic publication of such mega-matrix resources will increasingly become a major focus for collaborative taxonomic work and thus provide a more comprehensive foundation for understanding the phylogeny and evolution of the aroids.

Here, we have collected different kinds of these available up-to-date data, extended and analysed them with the newest methods. Although there are still unresolved questions, we got several new insights into Araceae phylogeny based on the most recent and complete data available and formed a basis for a new formal classification for the family.

6.5 References

Andrade, I. M. and S. J. Mayo. 1998. Dynamic Shoot Morphology in *Monstera adansonii* Schott var. *klotzschiana* (Schott) Madison (Araceae). Kew Bulletin 53(2): 399-417. 53:399–417.

Andrade, I. M. and S. J. Mayo. 2000. Dynamic shoot morphology in root-climbing Araceae: *Philodendron rudgeanum* Schott and *Ph. fragrantissimum* (Hook.) G.Don. Feddes Repertorium 111:295–314.

Barabé, D., A. Bruneau, F. Forest, and C. Lacroix. 2002. The correlation between development of atypical bisexual flowers and phylogeny in the Aroideae (Araceae). Plant Syst. Evol. 232:1–19.

Bogner, J. 2009. The free-floating Aroids (Araceae) – Living and fossil. Zitteliana in press.

Bogner, J. and E. G. Gonçalves. 2005. Two new species of *Xanthosoma* (Araceae) from South America and notes on the tribe Caladieae. Willdenowia 35:333–344.

Bogner, J. and M. Hesse. 2005. Zamioculcadoideae, a new subfamily of Araceae. Aroideana 28:3–20.

Bogner, J., G. L. Hoffman, and K. R. Aulenback. 2005. A fossilized aroid infructescence, *Albertarum pueri* gen. nov. et sp. nov. of Late Cretaceous (Late Campanian) age from the Horseshoe Canyon Formation of southern Alberta, Canada. Can. J. Bot. 83:591–598.

Bogner, J., K. R. Johnson, Z. Kvacek, and G. R. Upchurch Jr. 2007. New fossil leaves of Araceae from the Late Cretaceous and Paleogene of western North America. Zitteliana A47:133–147.

Bogner, J. and D. H. Nicolson. 1991. A revised classification of Araceae with dichotomous keys. Willdenowia 21:35–50.

Bogner, J. and G. Petersen. 2007. The chromsome numbers of the aroid genera. Aroideana 30:82–90.

Boyce, P. C. and S. Y. Wong. 2008. Studies on Schismatoglottidae (Araceae) of Borneo VII: *Schottarum* and *Bakoa*, two new genera from Sarawak, Malaysian Bornea. Botanical Studies 49:393–404.

Boyce, P. C. and S. Y. Wong. 2009. *Schottariella mirifica* P.C. Boyce & S.Y. Wong: a new name for *Schottarum sarikeense* (Araceae: Schismatoglottideae). Botanical Studies 50:269–271.

Cabrera, L. I., G. A. Salazar, M. W. Chase, S. J. Mayo, J. Bogner, and P. Davila. 2008. Phylogenetic relationships of aroids and duckweeds (Araceae) inferred from coding and noncoding plastid DNA. Am. J. Bot. 95:1153–1165.

Croat, T. 2004. History and current status of systematic research with Araceae. `http://www.aroid.org/literature/croat/croat_araceae_history04.pdf` .

Cusimano, N., M. Barrett, W. L. A. Hetterscheid, and S. S. Renner. 2010. A phylogeny of the Areae (Araceae) implies that *Typhonium, Sauromatum*, and *Lazarum* are distinct clades. Taxon 59:439–447.

Cusimano, N., L.-B. Zhang, and S. S. Renner. 2008. Reevaluation of the *cox1* group I intron in Araceae and angiosperms indicates a history dominated by loss rather than horizontal transfer. Mol. Biol. Evol. 25:265–276.

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. PLoS. Biol. 4:e88.

Drummond, A. J. and A. Rambaut. 2007. Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7:214.

Engler, A. 1876. Vergleichende Untersuchungen über die morphologischen Verhältnisse der Araceae. I. Theil. Natürliches System der Araceae. Acta Königl. Leopold.- Carol.- Deutschen Akad. Naturforscher 39:134–155.

Engler, A. 1877. Vergleichende Untersuchungen über die morphologischen Verhältnisse der Araceae. II. Theil. Ueber Blattstellung und Sprossverhältnisse der Araceae. Nov. Acta Königl. Leopold.- Carol.- Deutschen Akad. Naturforscher 39:158–232.

Engler, A. 1879. Araceae. Page 681 *in* Monographieae Phanerogamarum, vol. 2 (deCandolle, ed.). Masson, Paris.

Engler, A. 1920. Araceae – Pars generalis et index familiae generalis. *in* Das Pflanzenreich (A. Engler, ed.) vol. 74 (IV.23A) of *1-71*. Verlag Wilhelm Engelmann, Leipzig, Berlin.

Ertl, P. 1932. Vergleichende Untersuchungen über die Entwicklung der Blattnervatur der Araceen. Flora 126:115–248.

Eyde, R. H., D. H. Nicolson, and P. Sherwin. 1967. A survey of floral anatomy in Araceae. Am. J. Bot. 54:478–497.

French, J. C. 1985a. Patterns of endothecial wall thickenings in Araceae: subfamilies Calloideae, Lasioideae and Philodendroideae. Bot. Gaz. 146:521–533.

French, J. C. 1985b. Patterns of endothecial wall thickenings in Araceae: subfamilies Pothoideae and Monsteroideae. Am. J. Bot. 72:472–486.

French, J. C. 1986. Patterns of endothecial wall thickenings in Araceae: subfamilies Colocasioideae, Aroideae and Pistioideae. Bot. Gaz. 147:166–179.

French, J. C. 1987a. Systematic occurrence of a sclerotic hypodermis in roots of Araceae. Am. J. Bot. 74:891–903.

French, J. C. 1987b. Systematic survey of resin canals in roots of Araceae. Bot. Gaz. 148:360–371.

French, J. C. 1988. Systematic occurence of anastomosing laticifers in Araceae. Bot. Gaz. 149:71–81.

French, J. C., M. G. Chung, and Y. K. Hur. 1995. Chloroplast DNA phylogeny of the Ariflorae. Pages 255–275 *in* Monocotyledons: Systematics and Evolution (P. J. Rudall, P. Cribb, D. F. Cutler, and C. J. Humphries, eds.). Royal Botanic Gardens, Kew, England.

Friis, E. M., K. R. Pedersen, and P. R. Crane. 2004. Araceae from the Early Cretaceous of Portugal: Evidence on the emergence of monocotyledons. Proc. Natl. Acad. Sci. USA 101:16565–16570.

Gauthier, M.-P. L., D. Barabe, and A. Bruneau. 2008. Molecular phylogeny of the genus *Philodendron* (Araceae): delimitation and infrageneric classification. Bot. J. Linnean Soc. 156:13–27.

Gibernau, M., J. Albre, A. Dejean, and D. Barabé. 2002. Seed predation in *Philodendron solimoesense* (Araceae) by chalcid wasps. Int. J. Plant Sci. 163:1017–1023.

Gonçalves, E. G. 2002. Palynology and Phylogeny of the Araceae. Ph.D. thesis, Univ. Massachusetts, Amherst.

Gonçalves, E. G., S. J. Mayo, M.-A. Van Sluys, and A. Salatino. 2007. Combined genotypic-phenotypic phylogeny of the tribe Spathicarpeae (Araceae) with reference to independent events of invasion to Andean regions. Mol. Phylogenet. Evol. 43:1023–1039.

Gonçalves, E. G., E. A. S. Paiva, and M. A. Nadruz Coelho. 2004. A preliminary survey of petiolar collenchyma in the Araceae. Ann. Mo. Bot. Gard. 91:473–484.

Grayum, M. H. 1984. Palynology and Phylogeny of the Araceae. Ph.D. thesis Univ. Massachusetts (Amherst).

Grayum, M. H. 1990. Evolution and Phylogeny of the Araceae. Ann. Mo. Bot. Gard. 77:628–697.

Grayum, M. H. 1992. Comparative external pollen ultrastructure of the Araceae and putatively related taxa. Monogr. Syst. Bot. Missouri Bot. Gard 43:1–167.

Gunawardena, A., K. Sault, P. Donnelly, J. Greenwood, and N. Dengler. 2005. Programmed cell death and leaf morphogenesis in *Monstera obliqua* (Araceae). Planta 221:607–618.

Hay, A. 1986. *Cyrtosperma* Griff. and the origin of the Aroids. Doctoral thesis, Oxford University.

Hay, A. 1988. *Cyrtosperma* (Araceae) and its old world allies. Blumea 33:427–469.

Hay, A. 1992. A new Australian genus of Araceae, with notes on generic limits and biogeography of the Areae. Bot. J. Linnean Soc. 109:427–434.

Hay, A. and D. J. Mabberly. 1991. "Transference of Function" and the origin of aroids: their significance in early angiosperm evolution. Bot. Jahrb. Syst. 113:339–428.

Herrera, F. A., C. A. Jaramillo, D. L. Dilcher, S. L. Wing, and C. Gómez-N. 2008. Fossil Araceae from a Paleocene neotropical rainforest in Colombia. Am. J. Bot. 95:1569–1583.

Hesse, M. 2002. The uniquely designed pollen aperture in Lasioideae (Araceae). Aroideana 25:51–59.

Hesse, M. 2006a. Conventional and novel modes of exine patterning in members of the Araceae – the consequences of ecological paradigm shifts. Protoplasma 228:145–149.

Hesse, M. 2006b. Pollen wall ultrastructure of Araceae and Lemnaceae in relation to molecular classifications. Aliso 22:204–208.

Hesse, M. 2006c. Reasons and consequences of the lack of a sporopollenin ektexine in Aroideae (Araceae). Flora 201:421–428.

Hesse, M., J. Bogner, H. Halbritter, and M. Weber. 2001. Palynology of the perigoniate Aroideae: *Zamioculcas, Gonatopus* and *Stylochaeton* (Araceae). Grana 40:26–34.

Hesse, M., M. Weber, and H. Halbritter. 2000. A comparative study of the polyplicate pollen types in Arales, Laurales, Zingiberales and Gnetales. Pages 227–239 *in* Pollen and Spores: Morphology and Biology (M. M. Harley, C. M. Morton, and S. Blackmore, eds.). Royal Botanic Gardens, Kew.

Keating, R. C. 2000. Collenchyma in Araceae: trends and relation to classification. Bot. J. Linnean Soc. 134:203–214.

Keating, R. C. 2002. Acoraceae and Araceae. Pages 1–327 *in* Anatomy of the monocotyledons (M. Gregory and D. F. Cutler, eds.) vol. 9. Oxford University Press, Oxford, UK.

Keating, R. C. 2004. Vegetative anatomical data and its relationship to a revised classification of the genera of Araceae. Ann. Mo. Bot. Gard. 91:485–494.

Landolt, E. 1986. Biosystematic investigations in the family of duckweeds (Lemnaceae), vol. 2. Page 566 *in* The family of Lemnaceae – a monographic study, vol. 1 (E. Landolt, ed.) vol. Heft 71. Veröff. Geobot. Inst. ETH, Stiftung Rübel, Zürich.

Landolt, E. 1998. Lemnaceae. Pages 264–270 *in* The Families and Genera of Vascular Plants. IV. Flowering Plants. Monocotyledons. Alismatanae and Commelinanae (except Gramineae) (K. Kubitzki, ed.). Springer, Berlin.

Landolt, E. and R. Kandler. 1987. Biosystematic investigations in the family of duckweeds (Lemnaceae),vol. 4. Page 638 *in* The family of Lemnaceae – a monographic study. Vol. 2 (E. Landolt, ed.) vol. Heft 95. Veröff. Geobot. Inst. ETH, Stiftung Rübel, Zürich.

Les, D. H., D. J. Crawford, E. Landolt, J. D. Gabel, and R. T. Kimball. 2002. Phylogeny and systematics of Lemnaceae, the duckweed family. Syst. Bot. 27:221–240.

Madison, M. 1981. Notes on *Caladium* (Araceae) and its allies. Selbyana 5:342–377.

Mayo, S. J. 1989. Observations of gynoecial structure in *Philodendron* (Araceae). Bot. J. Linnean Soc. 100:139–172.

Mayo, S. J. 1991. A revision of *Philodendron* Subgenus *Meconostigma* (Araceae). Kew Bull. 46:601–681.

Mayo, S. J. and J. Bogner. 1988. A new species of *Caladium* (Araceae) with notes on generic delimitation in the Colocasioideae-Caladieae. Willdenowia 18:231–242.

Mayo, S. J., J. Bogner, and P. C. Boyce. 1997. The genera of Araceae. The Trustees, Royal Botanical Gardens, Kew, UK.

Nicolson, D. H. 1987. History of aroid systematics. Aroideana 10:23–30.

Ray, T. 1986. Growth correlations within the segment in the Araceae. Am. J. Bot. 73:993–1001.

Ray, T. 1987a. Cyclic heterophylly in *Synongium* (Araceae). Am. J. Bot. 74:16–26.

Ray, T. 1987b. Diversity of shoot organization in the Araceae. Am. J. Bot. 74:1373–1387.

Ray, T. 1987c. Leaf types in the Araceae. Am. J. Bot. 74:1359–1372.

Ray, T. 1988. Survey of shoot organization in the Araceae. Am. J. Bot. 75:56–84.

Ray, T. 1990. Metamorphosis in the Araceae. Am. J. Bot. 77:1599–1609.

Ray, T. S. and S. S. Renner. 1990. Comparative studies on the morphology of the Araceae. Transl. from A. Engler, 1877, with an introduction, updated nomenclature and a glossary. Englera 12:1–140.

Renner, S. S. and L.-B. Zhang. 2004. Biogeography of the *Pistia* clade (Araceae): Based on chloroplast and mitochondrial DNA sequences and Bayesian divergence time inference. Syst. Biol. 53:422–432.

Renner, S. S., L.-B. Zhang, and J. Murata. 2004. A chloroplast phylogeny of *Arisaema* (Araceae) illustrates tertiary floristic links between Asia, North America, and East Africa. Am. J. Bot. 91:881–888.

Rothwell, G. W., M. R. van Atta, H. W. Ballard Jr., and R. A. Stockey. 2004. Molecular phylogenetic relationships among Lemnaceae and Araceae using the chloroplast *trnL–trnF* spacer. Mol. Phylogenet. Evol. 30:378–385.

Rudall, P. J., P. J. Cribb, D. F. Cutler, and C. J. Humphries. 1995. Monocotyledons: Systematics and Evolution. 2 vols. Royal Botanic Gardens, Kew.

Seubert, E. 1993. Die Samen der Araceen: Die Samenmerkmale der Araceen und ihre Bedeutung für die Gliederung der Familie. Ph.D. thesis Koeltz, Königstein.

Smith, S. Y. and R. A. Stockey. 2003. Aroid seeds from the Middle Eocene Princeton Chert (*Keratosperma allenbyense*, Araceae): comparisons withm extant Lasioideae. Int. J. Plant Sci. 164:239–250.

Stamatakis, A., P. Hoover, and J. Rougemont. 2008. A rapid bootstrap algorithm for the RAxML web-servers. Syst. Biol. 57:758–771.

Stockey, R. A., G. L. Hoffman, and G. W. Rothwell. 1997. The fossil *Limnobiophyllum scutatum*: resolving the phylogeny of Lemnaceae. Am. J. Bot. 84:355–368.

Stockey, R. A., G. W. Rothwell, and K. R. Johnson. 2007. *Cobbania corrrugata* gen. et comb. nov. (Araceae): a floating aquatic monocot from the Upper Cretaceous of western North America. Am. J. Bot. 94:609–624.

Swofford, D. L. 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Version 4.0b. Sinauer Associates, Sunderland, Massachusetts.

Tam, S., P. Boyce, T. Upson, D. Barabe, A. Bruneau, F. Forest, and J. Parker. 2004. Intergeneric and infrafamilial phylogeny of subfamily Monsteroideae (Araceae) revealed by chloroplast *trnL-F* sequences. Am. J. Bot. 91:490–498.

Tarasevich, V. F. 1989. Pollen grain ultrastructure in the genus *Anthurium* (Araceae) in connection with its systematics. Bot. Zh. 74:314–424.

Tillich, H.-J. 2003. Seedling diversity in Araceae and its systematic implications. Feddes Repertorium 114:454–487.

Vogel, S. 1963. Duftdrüsen im Dienste der Bestäubung: über Bau und Funktion der Osmophoren. Abhandl. Math.-Naturwiss. Kl. Akad. Wiss. Mainz 1962:599–763.

Weber, M. 2004. Unique pollen features in Araceae. Pollen 14:18.

Wilde, V., Z. Kvacek, and J. Bogner. 2005. Fossil leaves of the Araceae from the European Eocene and notes on other aroid fossils. Int. J. Plant Sci. 166:157–183.

Wong, S. Y. and P. C. Boyce. 2010. Studies on Schismatoglottideae (Araceae) of Borneo IX: A new genus, *Hestia*, and resurrection of *Apoballis*. Botanical Studies 51:249–255.

Young, H. 1986. Beetle pollination in *Dieffenbachia longispatha* (Araceae). Am. J. Bot. 73:931–944.

6.6 Appendices

Appendix 1: Description of the Morphological Characters

1. **Flower sexuality**: bisexual < 0 >; unisexual < 1 >
   Flower sexuality is normally unambiguous. However in certain genera such as *Calla*, the uppermost flowers of the spadix are unisexual in behaviour although bisexual in structure.

2. **Perigon**: present < 0 >; absent < 1 >
   In some genera the perigon can be inconspicuous, e.g., *Anadendrum*.

3. **Trichosclereids**: absent < 0 >; present not in bundles, ± large < 1 >; present in bundles, ± small < 2 >
   Trichosclereids are found sporadically in *Pothos* (Keating, 2002). In some genera of the Monsteroideae (e.g. *Amydrium*), the trichosclereids are not necessarily present in all organs (check Keating).

4. **Spathe structure**: spathe bract-like, unspecialized, ± inconspicuous < 0 >; spathe modified for attraction and display or specialized in some other way < 1 >; spathe absent < 2 >
   The spathe is almost always well developed and functions as an integral part of the inflorescence. However in *Gymnostachys* and *Orontium* the structure homologous to a spathe is the simple bract that subtends the spadix, while in *Pothoidium* many spadices entirely lack a spathe or homologous structure.

5. **Inflorescence stipe and peduncle**: major internode of the inflorescence is the stipe situated between spadix and spathe < 0 >; major internode of the inflorescence is the peduncle situated between spathe and next leaf below < 1 >
   The long stipe (character state 0) is most clearly seen in *Orontium* and *Lysichiton*.

6. **Shoot architecture**: continuation shoot in last leaf axil before spathe < 0 >; continuation shoot in axil of penultimate leaf before spathe < 1 >; culm type inflorescence with cauline bracts apart from spathe (*Gymnostachys*) < 2 >; monopodial scandent structure < 3 >
   Engler (1877) made the first comparative survey of shoot architecture in Araceae, later further developed by (Ray, 1986, 1987c,b,a, 1988, 1990) and Andrade and Mayo (1998, 2000). The monopodial shoot structure (character state 3) is found only in climbing hemi-epiphytes and is probably a derived condition specialized for that ecological niche.

7. **Phyllotaxy**: distichous < 0 >; spiral < 1 >

8. **Pollen aperture** : monosulcate < 0 >; extended monosulcate < 1 >; zonate < 2 >; diaperturate < 3 >; forate < 4 >; inaperturate < 5 >; monosulcoidate < 6 >; ulcerate < 7 >
   The analysis of characters 8 to 12 is based primarily on the work of Grayum (1984, 1992), Hesse et al. (2001); Hesse (2006b) and Bogner and Hesse (2005).

9. **Pollen units**: monads < 0 >; tetrads < 1 >; dyads < 2 >
   Tetrad pollen was thought previously to be a good diagnostic character to separate *Xanthosoma* from *Caladium* (Mayo and Bogner, 1988), but more recent studies have shown that this is not the case (Gonçalves et al., 2007).

10. **Pollen shape**: ellipsoid < 0 >; hamburger-shaped < 1 >; globose < 2 >

11. **Mean pollen size**: small 10-24 μm. $< 0 >$; medium 25-49 μm. $< 1 >$; large 50-99 μm. $< 2 >$; very large 100-199 μm. $< 3 >$

12. **Pollen exine surface**: reticulate or other $< 0 >$; smooth $< 1 >$; spinose $< 2 >$; striate $< 3 >$
    We have included within the character state "reticulate or other" (0) a range of other types named by Grayum (1992).

13. **Petiole geniculum**: absent $< 0 >$; present $< 1 >$
    The geniculum is a pulvinus at the apex of the petiole which in some genera (e.g. *Bognera*, some species of *Philodendron*) is only clearly visible in living plants.

14. **Pollen starch**: absent $< 0 >$; present $< 1 >$
    This analysis is based on the survey by Grayum (1992) and additional observations by Hesse (pers. comm.).

15. **Sclerotic hypodermis in roots**: absent $< 0 >$; present $< 1 >$
    This analysis is based on the survey by (French, 1987a).

16. **Endothecial thickenings**: present $< 0 >$; absent $< 1 >$
    This analysis is based on the survey by French (1985a,b, 1986).

17. **Resin canals in roots**: absent $< 0 >$; present $< 1 >$
    This analysis is based on the survey by French (1987b).

18. **Laticifers**: absent or only scattered cells $< 0 >$; present, simple, articulated $< 1 >$; present, anastomosing $< 2 >$
    This analysis is based on the surveys by French (1988) and especially (Keating, 2002).

19. **Latex s bodies**: absent $< 0 >$; present $< 1 >$
    The analysis and data for Characters 19 to 22 are based on the survey made by Fox & French (1988), which covered only the genera with abundant milky latex of the family Colocasioideae (as recognized by Bogner and Nicolson, 1991).

20. **Latex rubber (r) particles**: absent $< 0 >$; present $< 1 >$

21. **Latex y bodies**: absent $< 0 >$; present $< 1 >$

22. **Latex r bodies**: absent $< 0 >$; present $< 1 >$

23. **Primary leaf venation – midrib**: midrib of primary veins $\pm$ absent with veins arcuate from the base $< 0 >$; midrib of primary veins well developed, i.e. pinnate veins in anterior division $< 1 >$
    The midrib is normally formed from the fusion of the proximal portions of the major (primary) veins of the anterior division of the leaf blade. In their distal portions these

veins diverge successively from the midrib in a pinnate pattern and are then referred to as primary lateral veins. The midrib, as defined here, refers only to the compound structure created by the proximal fusion of the primary veins. In some genera (e.g. *Filarum, Ulearum*) there is no basal connation of the primary veins and we therefore score this as the absence of a midrib, despite the presence of a single central primary vein nevertheless. Likewise, those genera with pedately divided leaves (e.g. *Dracunculus*) are scored as lacking a midrib (state 0) since the primary veins are the central veins of each leaflet. This analysis must be regarded as provisional, because apart from the study by Ertl (1932) there is little data on the ontogenesis of leaf venation in the Araceae. An improved classification of venation patterns in the future would, incidentally, facilitate the taxonomic interpretation of leaf fossils (Wilde et al., 2005; Bogner et al., 2007).

24. **Primary leaf venation – basal ribs**: basal ribs of primary veins absent < 0 >; basal ribs of primary veins very well developed, i.e., ± tripartite primary development < 1 >; basal ribs of primary veins distinct but short, i.e. as found in hastate, sagittate, pedate, trilobed and radiate primary vein patterns < 2 >
Basal ribs are defined here as exactly analogous to the midrib, but formed by the fusion of primary veins of the posterior divisions (see GoA page 8, fig. 6). Basal ribs occur only in leaves which have posterior divisions and are particularly well-developed in the Lasioideae, where in some species of *Cyrtosperma* they may be more strongly developed than the midrib (e.g. GoA page 139, plate 26).

25. **Primary leaf venation – blade**: distinct blade not differentiated, leaf ± linear < 0 >; distinct blade differentiated < 1 >; distinct blade not differentiated, leaf and shoot fused into a thallus-like body called a frond < 2 >
Clear differentiation of petiole and leaf blade is near-universal in Araceae, except *Gymnostachys*, and some species of *Biarum*. In the Lemnaceae we interpret the pouch region of the frond of the genera *Lemna, Landoltia* and *Spirodela* as homologous with a petiole and sheath, in contrast to *Wolffia* and *Wolffiella* in which the leaf and stem structures are conceived to be congenitally or rather, as a highly reduced neotenic form without such differentiation.

26. **Primary leaf venation – marginal venation of anterior division or ultimate leaf lobes**: ± no sympodial marginal or inframarginal vein, primary veins usually fusing only near apex < 0 >; sympodial marginal vein formed of ± all primary veins, no sympodial inframarginal vein < 1 >; sympodial inframarginal vein formed by majority of primary veins, lowermost primary veins forming non-sympodial marginal veins < 2 >
This analysis is based on a reading of Ertl (1932) from which a transformation series can be conceptualized beginning with a *Hydrocleys*-like pattern in which the primary veins diverge at the petiole insertion and curve around to join together at or very near the leaf apex. In the Araceae this pattern is found in *Anthurium* sect. Digitinervium and is approached in *Pistia* and *Ambrosina*. In most Araceae the primary lateral veins join together at the

margin into a sympodial marginal vein (e.g. *Philodendron, Schismatoglottis, Heteropsis*). Sometimes a sympodial submarginal collective vein is also formed (e.g. GoA, pages 104 – 107, plates 8i to 8iv) apparently by fusion of the distal portions of some or nearly all the primary lateral veins. In these cases the marginal vein is formed by the distal portion of more basal primary lateral veins. Some genera (e.g. *Peltandra, Typhonodorum*) characteristically have a series of marginal and submarginal veins running parallel to each other near the margin (GoA, page 248, plate 84A, B)

27. **Marginal form of leaf**: margin not lobed individually around primary veins < 0 >; margin lobed pinnately in the anterior division < 1 >; margin lobed pedately or radiately < 2 >; margin lobed both pinnately and pedately (dracontioid) < 3 >; margin lobed trifidly including deeply sagittate < 4 >; margin bi- to tri- to quadripinnate < 5 >
The patterns of leaf lobing are among the most striking characters of Araceae, but explanations of the relationships between these patterns are still ad hoc and require more thorough morphogenetic studies. There is a close link between the differentiation of major veins and the appearance of leaf lobes or segments, but the degree of marginal lobing may vary considerably among taxa with essentially similar primary vein patterns (e.g. *Philodendron* subgen. *Meconostigma*, Mayo, 1991). Dracontioid leaves (state 3) are the result of subdivision of leaves in which the posterior and anterior divisions are approximately equally well-developed. Pedate leaves represent the condition of subdivision of the posterior divisions with the anterior division remaining entire (e.g. *Philodendron goeldii*). In some pinnately lobed and pinnatisect leaves the lobing of posterior divisions is only weakly or not at all developed (e.g. *Anaphyllum, Zamioculcas, Gorgonidium*).

28. **Fine venation**: secondaries and tertiaries forming mostly cross veins to primaries < 0 >; secondaries and tertiaries parallel to primaries, joined by cross veins only < 1 >; secondaries and tertiaries ± parallel to primaries, often forming interprimary sympodial veins, cross connections often reticulated < 2 >; secondaries and tertiaries mostly reticulating freely, forming interprimary sympodial veins < 3 >; absent, only primary veins present < 4 >
The type of fine leaf venation has been used since Engler (1876) as a subfamily character within the Araceae. Subsequent work by (Ertl, 1932) suggested that these differences were less distinct than had been thought and recent molecular phylogenetic studies have suggested that this character is useful only at lower taxonomic categories. Engler's subfamilies Philodendroideae and Colocasioideae Engler (1920) were characterized respectively by parallel-pinnate venation (secondaries and tertiaries parallel to pinnately organized primary lateral veins; GoA, page 311, Venation types F) and colocasioid venation (GoA, page 311, Venation types E). However, subsequent phylogenetic studies have shown that parallel-pinnate and reticulate venation may both occur within a single tribal clade (e.g. Spathicarpeae sensu Gonçalves et al., 2007) and the two tribes characterized by colocasioid venation, the Caladieae and the Colocasieae, are found in widely separate subclades of the subfamily Aroideae (Cabrera et al., 2008). We have therefore avoided the older anal-

ysis terms into "reticulate", "philodendroid" and "colocasioid" and instead attempted a description of each state in terms of the mutual relationships of primary, secondary and tertiary lateral veins. This classification vein patterns remains provisional and needs further studies to place on a firmer footing.

29. **Leaf fenestration**: fenestration by necrosis between primary veins absent < 0 >; fenestration by necrosis between primary veins present < 1 >
The best-known examples of leaf fenestration occur in Monstera and some other genera of the tribe Monstereae. While normally understood as the occurrence of perforations in an otherwise simple leaf, we have followed previous authors (e.g., Grayum, 1984, 1990) in interpreting the pinnatisect and dracontioid leaves of some other genera (*Rhaphidophora, Epipremnum, Dracontium, Dracontioides, Anchomanes*) as the result of necrotic fenestration (programmed cell death) of an entire leaf blade rather than differential marginal growth. No comparative study of programmed cell death has yet been made in Araceae but an interesting recent report in *Monstera* is that of Gunawardena et al. (2005).

30. **Prickles on stem or petiole**: absent < 0 >; present < 1 >

31. **Stem producing erect shoots with bulbils**: absent < 0 >; present < 1 >
This character is only present in the genera *Remusatia* and *Gonatanthus*, the latter genus being now a synonym of the former.

32. **Stem type − thickening**: not condensed and strongly thickened < 0 >; condensed, strongly thickened but not depressed-globose < 1 >; condensed, strongly thickened into depressed-globose tuber (corm) < 2 >; stem reduced to a minute button, or indistinguishable from the thalloid structure < 3 >
Stem morphology, like leaf venation, is another character field in need of a more analytical understanding of homologies to substitute the present rather intuitive expression of character states. Anatomical and morphogenetic studies are needed to provide the basis for this desired improvement in understanding. The analysis presented here (characters 32 and 33) focuses on separating stem thickening for nutrient storage from the habit, suggesting these may not be completely independent characters.

33. **Stem type − habit**: subaerial, creeping to erect < 0 >; ± erect at least distally, aerial < 1 >; entirely subterranean < 2 >; aerial and climbing < *includinghemiepiphytes* >< 3 >; aerial and truly epiphytic < 4 >
Subterranean stems tend to be thickened for nutrient storage, i.e. correlation with character 33 but this is not always the case; for example in some *Stylochaeton* and *Gearum* the roots seem to have such a role, being often very thick and fleshy and the stem relatively slender despite the geophytic habit. Epiphytes and hemiepiphytes tend to have elongated and relatively slender green stems, but *Remusatia*, with a subglobose tuberous stem is frequently found as an epiphyte. Many other such examples exist which make the independence of stem storage thickening and habit at least a reasonable working hypothesis.

34. **Peltate leaves**; absent < 0 >; present < 1 >

35. **Aquatic habit**: not aquatic < 0 >; helophytes < 1 >; floating aquatics < 2 >; submerged aquatics < 3 >; rheophytes < 4 >
    We use "not aquatic" in preference to "terrestrial" because the latter state would epiphytic and hemi-epiphytic taxa, and this habit difference is dealt with in character 33.

36. **Petiole ligule**: petiole sheath not long-ligulate apically < 0 >; petiole sheath long-ligulate apically < 1 >
    This character is common to several genera of the Schismatoglottideae but also occurs in *Calla*.

37. **Infravaginal squamules**: absent < 0 >; present < 1 >
    Infravaginal squamules are similar to colleters, epidermal structures that appear to have a secretory function in the early development of the shoot and later may become stiff and even prickle-like in certain species of *Philodendron* subgen. *Meconostigma*. In the Araceae they occur only in the *Cryptocoryneae* and *Philodendron*.

38. **Spathe behaviour**: no differentiation in persistence, entire spathe persistent or withering slowly without distinctive abscission < 0 >; tube or lower half persistent, blade marcescent or caducous < 1 >; no differentiation, entire spathe soon deciduous or marcescent with distinct basal abscission < 2 >
    Some unisexual-flowered genera show a marked differentiation of persistence and colour between spathe tube and blade in which the paler (often white) blade, along with the associated portion of the spadix, speedily withers, decomposes or just falls off following a rapid absicission. Prior to the results of French et al. (1995) and GoA (1997) this was considered characteristic of the subfamily Colocasioideae, but it is also typical of the Schismatoglottideae. In *Piptospatha*, the spathe is sub-globose but exhibits rapid post-floral abscission in most species, thus demonstrating that this character is not necessarily correlated with the presence of the spathe constriction characteristic of many genera e.g., *Xanthosoma, Colocasia*. Spathe which fall soon after flowering are typical of the tribe Monstereae. Ecologically there must be a connection between spathe behaviour and the mode of protection of the developing seeds, but this has been little studied since Madison (1979). In Monstereae for example, the spathe does not protect the developing fruits within the spadix as in e.g., *Philodendron*). Instead the flowers are full of trichosclereidswhich protects them from herbivores.

39. **Spathe shape**: fully expanded, often reflexed < 0 >; boat- shaped, ± convolute basally or not < 1 >; convolute basally into distinct tube with distinctly different, ± expanded blade < 2 >
    The presence of strong shape differentiation between tube and blade is found only in unisexual-flowered genera, the canonical example being *Arum* itself. This character is often accompanied by the presence of a more (e.g. *Xanthosoma*) or less (many *Philodendron*

species) strongly marked constriction between the two portions of the spathe. In *Arisaema* the spathe tube tends to be cylindric without a constriction despite an often remarkable differentiation of the blade (e.g. *A. griffithii*). Within the unisexual genera there are many in which spathe shapes are much simpler, e.g. the boat-shaped spathes of *Anchomanes*. As with other features of spathe and spadix morphology, biological understanding requires further knowledge about floral ecology, pollination and dispersal.

40. **Spadix-spathe fusion – chambers**: spathe and spadix not forming 2 distinct chambers by fusion < 0 >; spathe and spadix forming 2 distinct chambers by partial fusion < 1 >
State 1 describes a more complete separation of tube and blade by the presence of partial septum with only a narrow passage allowing the movement of pollinators.

41. **Spadix-spathe fusion – internal flap**: spathe without internal flap covering and adnate to spadix apex < 0 >; spathe with internal flap covering and adnate to spadix apex < 1 >
This character (state 1) is unique to the Cryptocoryneae.

42. **Spathe margins**: margins free or connate only at extreme base < 0 >; connate for distinct distance < 1 >
Connate spathe tubes are probably only synapomorphic for the genera of the Cryptocoryneae.

43. **Spadix-spathe fusion – dorsal fusion**: spadix dorsally free of spathe < 0 >; spadix female zone dorsally adnate to spathe < 1 >; spadix entirely dorsally adnate to spathe < 2 >
Completely adnate spadices (2) occur in *Spathicarpa* and *Spathantheum*.

44. **Spadix zonation**: no zonation < 0 >; male and female zones only < 1 >; female, sterile, male zones < 2 >; female, male, sterile zones < 3 >; female, sterile, male, sterile zones < 4 >
The zonation is to be thought of as extending from base to apex in the above schema. All bisexual-flowered genera are counted as "no zonation" (0), and this character is thus partly dependent on flower sexuality (character 1).

45. **Spadix appendix**: appendix absent or inconspicuous < 0 >; appendix a conspicuous and well developed organ, staminodial < 1 >; appendix a conspicuous and well developed organ, smooth to corrugated < 2 >

46. **Basipetal flowering**: flowering sequence of spadix not basipetal < 0 >; flowering sequence of spadix basipetal < 1 >
This character is only known in subfamily Lasioideae.

47. **Female zone length**: composed of more than 1 flower < 0 >; composed of only 1 flower < 1 >

In *Pistia* and *Ambrosina* the female zone consists of only a solitary gynoecium. In *Arisarum*, there may be as few as two.

48. **Stamen thecae horns**: thecae not horned < 0 >; thecae horned < 1 >
This very interesting character occurs in the tribes Cryptocoryneae and Schismatoglottideae, but not in *Schismatoglottis* and *Piptospatha*.

49. **Stamen connective**: not strongly thickened < 0 >; strongly thickened laterally, at apex, and ± glandular < 1 >
Thick stamen connectives are found in many unisexual-flowered genera, a typical example being *Philodendron*. Based on the study of osmophores by Vogel (1963) we have assumed that in these genera the stamen connectives generate floral odours.

50. **Stamen connation**: stamens free < 0 >; connate by filaments < 1 >; entirely connate (including connectives) < 2 >
Connation of the stamens of the floral unit is mostly found in unisexual-flowered genera, although in Lasimorpha (subfamily Lasioideae) the filaments are more-or-less connate. In genera with thickened stamen connectives (character 49) connation of the stamens creates a large mass of osmophoric glandular tissue within each male flower (e.g. Anubias, Xanthosoma).

51. **Anther dehiscence**: dehiscing by longitudinal slits < 0 >; dehiscing by oblique pore-like slits < 1 >; dehiscing by apical pores < 2 >
The manner in which the anthers dehisce and present the pollen varies widely. In many unisexual-flowered genera the pollen is extruded in strings from pore-like anther stomia. Pollen presentation modes may be correlated with other inflorescence attributes such as secretion sticky substances such as resins (Grayum, 1990).

52. **Staminodes in female zone**: absent < 0 >; present < 1 >
The presence of staminodes in the female flower is characteristic of the tribe Spathicarpeae (including *Dieffenbachia* as reformulated by Gonçalves et al., 2007), the *Peltandreae, Protarum*, most *Homalomena* and *Furtadoa*. Little is known of the function of these organs, except in *Dieffenbachia* where they serve as food for pollinating beetles (Young, 1986).

53. **Ovary locules**: two to three < 0 >; unilocular < 1 >; more than three < 2 >
Unilocular ovaries in Araceae are regarded as pseudomonomerous (Eyde et al., 1967) and usually provide anatomical evidence of derivation from multilocular ancestors. Ovaries with high numbers of locules are found in some genera of Spathicarpeae and especially in Philodendron, although in the latter case there are species with as few as 2 or 3 locules (Mayo, 1989). High locule number may be a derived feature linked to parasitism of ovules and seeds by chalcid wasps (Gibernau et al., 2002).

54. **Ovule type**: anatropous or hemianatropous < 0 >; campylotropous < 1 >; hemiorthotropous < 2 >; strictly orthotropous < 3 >

55. **Endosperm**: copious, embryo relatively small < 0 >; sparse to absent, embryo relatively large < 1 >
    Important sources of information on endosperm characters in Araceae are Seubert (1993) and Tillich (2003).

56. **Placentation**: axile or on strongly intrusive placentae < 0 >; apical < 1 >; parietal, septa very reduced < 2 >; basal, basal-parietal or basal-axile < 3 >; basal and apical < 4 >
    Placental characters are normally observed using low-power microscopy or even hand lenses fom herbarium or living specimens but the analysis of this character would benefit from more detailed comparative anatomical studies.

57. **Chromosome base number**: $x = 12 < 0 >$; $x = 13 < 1 >$; $x = 7 < 2 >$; $x = 11 < 3 >$; $x = 21 < 4 >$; $x = 10 < 5 >$; $x = 15 < 6 >$; $x = 8 < 7 >$; $x = 17 < 8 >$; $x = 18 < 9 >$
    Bogner and Petersen (2007) have reviewed this character recently and added a number of important new counts. This paper includes the most recent synopsis of the family classification based on GoA but including the Lemnacae as subfamily Lemnoideae.

58. **Staminodes**: staminodes in interfertile zones absent, prismatic or fungiform < 0 >; staminodes in interfertile zones hair-like, subulate, bristle-like or clavate-elongated < 1 >
    Character state (1) occurs almost exclusively in the tribe Areae.

59. **Style**: styles not laterally thickened or extended < 0 >; styles laterally thickened or extended into "mantle" and contiguous < 1 >
    Laterally extended or thickened styles (state 1) are characteristic of the genera *Xanthosoma* and *Chlorospatha* in tribe Caladieae and the absence of this character state has been used to distinguish *Caladium* from these genera Madison (1981); Mayo and Bogner (1988); Bogner and Gonçalves (2005); Mayo and Bogner (1988)

60. **Ovule number**: 3 or more per locule < 0 >; 1-2 per locule < 1 >

61. **Perigon-tepal connation**: tepals free or partly free < 0 >; perigon a single unit < 1 >

62. **Perigon-tepal apex**: unthickened < 0 >; thickened < 1 >

63. **Male flowers**: no pistillode present < 0 >; pistillode or vestige, e.g., stigmatoids, present < 1 >
    Pistillodes in male flowers are characteristic of tribes Spathicarpeae, Stylochaetoneae and subfamily Zamioculcadoideae.

Characters 64 – 69 are from the comprehensive monograph of Araceae vegetative anatomy by Keating (2002).

64. **Leaf TS spongy aerenchyma**: compact tissue or cell-sized air spaces type 1, type 1a < 0 >; intercellular spaces 2-3 times cell diameters, type 2 < 1 >; intercellular spaces large with uniseriate partitions, type 3 < 2 >; intercellular spaces with multiseriate partitions, type 4 < 3 >

65. **Leaf TS collenchyma**: banded, type B < 0 >; banded, interrupted, type Bi < 1 >; strands irregular, often flattened, type Sb < 2 >; strands as phloem caps, type Svc < 3 >; strands discrete, usually circular, type Sv < 4 >; absent < 5 >

66. **Leaf TS vascular bundles**: broad, phloem strands semi-circular, type 1 < 0 >; narrow, phloem strands elliptical, type 2 < 1 >; small, xylem a wide cell or lacuna, type 3 < 2 >

67. **Petiole TS collenchyma**: banded, type B < 0 >; banded interrupted, type Bi < 1 > irregular flattened strands, not aligned with vascular bundles, type Sb < 2 >; bundle caps, type Svc < 3 >; strands aligned with vascular bundles, type Sv < 4 >; ensheathing vascular bundles < 5 >; absent < 6 >

68. **Petiole TS ground tissue**: cells compact in centre < 0 >; cells loosely packed (cell-sized cavities), type 1 and 2 < 1 >; uniseriate partitions separating cavities, type 3 < 2 >; multiseriate partitions separating cavities, type 4 < 3 >

69. **Petiole TS vascular bundles**: large, type 1 < 0 >; medium-sized, type 2 < 1 >; small, with dominant single xylem element, type 3 < 2 >

Characters 70 – 74 are from Tillich (2003):

70. **Cotyledon type**: haustorial < 0 >; storage < 1 >

71. **Cotyledon shape**: compact < 0 >; hypophyll widened, blade-like < 1 >

72. **First leaf**: cataphyll < 0 >; eophyll < 1 >

73. **hypocotyl presence**: present < 0 >; absent < 1 >

74. **Primary root presence**: present < 0 >; absent < 1 >

75. **Sporopollenin**: present < 0 >; absent < 1 >
This character is based on the studies of Hesse

Characters 76 – 79 focus on the differentiation of the Lemnoid genera from the rest of the family, based on (Landolt, 1998):

76. **Roots**: present < 0 >; absent < 1 >

77. **Root** branching: present < 0 >; absent < 1 >

78. **Leaf** veins: present < 0 >; absent < 1 >

79. **Anther thecae**: tetrasporangiate < 0 >; disporangiate < 1 >

80. **Raphides**: absent < 0 >; present < 1 >
    All Araceae (including the Lemnoideae) have raphides, but not *Acorus*.

81. **Biforines**: absent < 0 >; present < 1 >
    This character is based on information from Keating (2002), who found that biforines are only observed in unisexual-flowered genera.

Appendix 2: Matrix of morphological data with codes as described above. Codes for polymorphisms : a = (01); b = (12); c = (89); d = (012); e = (02); f = (04); g = (24); h = (26); i = (0123).

	Flower sexuality 1	Perigon 2	Trichosclereids 3	Spathe structure 4	Inflorescence stipe/peduncle 5	Shoot architecture 6	Phyllotaxy 7	Pollen aperture 8	Pollen units 9	Pollen shape 10	Mean pollen size 11	Pollen exine surface 12	Petiole geniculum 13	Pollen starch 14	Sclerotic hypodermis in roots 15	Endothecial thickenings 16	Resin canals in roots 17	Laticifers 18	Latex S bodies 19	Latex rubber (r) particles 20	Latex Y bodies 21	Latex R bodies 22	Primary leaf venation midrib 23	Primary leaf venation basal ribs 24	Leaf/petiole differentiation 25	Primary leaf venation marginal venation of anterior division or ultimate leaf lobes 26	Marginal form of leaf 27	Fine venation 28	Leaf fenestration 29	Prickles on stem or petiole 30	Stem producing erect shoots with bulbils 31	Stem type - thickening 32	Stem type - habit 33	Peltate leaves 34	Aquatic habit 35	Petiole ligule 36	Infravaginal squamules 37	Spathe behaviour 38	Spathe shape 39	Spadix-spathe fusion - chambers 40	Spadix-spathe fusion - internal flap 41	Spathe margins 42	Spadix-spathe fusion - dorsal fusion 43
Acorus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0
Aglaodorum	1	1	0	1	1	1	1	5	0	0	2	1	0	1	0	0	0	1	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	1	0	0	2	1	0	0	0	0
Aglaonema	1	1	0	1	1	1	1	5	0	0	2	1	0	1	0	0	0	1	0	0	0	0	1	0	1	0	0	1	0	0	0	0	1	0	0	0	0	2	1	0	0	0	0
Alloschemone	0	1	1	1	1	1	0	2	0	0	1	0	1	?	?	?	?	0	0	0	0	0	1	0	1	1	0	1	1	0	0	0	3	0	0	0	0	2	1	0	0	0	0
Alocasia	1	1	0	1	1	1	1	5	0	2	1	2	0	1	0	0	0	2	0	0	1	0	1	2	1	2	0	2	0	0	0	1	b	0	0	0	0	1	2	0	0	0	0
Ambrosina	1	1	0	1	1	1	1	5	0	0	1	3	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	3	0	0	0	2	2	0	0	0	0	0	2	1	0	1	1
Amorphophallus	1	1	0	1	1	1	1	5	0	0	b	i	0	1	0	0	0	1	0	0	0	0	1	1	1	2	3	3	0	0	0	2	2	0	0	0	0	0	1	0	0	0	0
Amydrium	0	1	0	1	1	1	?	2	0	1	1	0	1	a	0	0	0	0	0	0	0	0	1	0	1	1	0	3	1	0	0	0	3	0	0	0	0	2	1	0	0	0	0
Anadendrum	0	0	0	1	1	1	0	5	0	2	0	0	1	?	0	0	0	0	0	0	0	0	1	0	1	1	0	3	0	0	0	0	3	0	0	0	0	2	1	0	0	0	0
Anaphyllopsis	0	0	0	1	1	1	1	0	0	0	1	0	0	?	0	0	0	0	0	0	0	0	1	1	1	2	3	3	1	0	0	2	2	0	0	0	0	0	1	0	0	0	0
Anaphyllum	0	0	0	1	1	1	1	0	0	0	1	0	1	?	0	0	0	0	0	0	0	0	1	1	1	b	3	3	0	0	0	0	0	0	a	0	0	0	0	0	0	0	0
Anchomanes	1	1	0	1	1	1	1	5	0	0	2	1	1	1	0	0	0	1	0	0	0	0	1	1	1	2	0	3	1	1	0	1	2	0	0	0	0	0	1	0	0	0	0
Anthurium	0	0	0	1	1	1	1	4	0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
Anubias	1	1	0	1	1	1	1	5	0	2	0	1	1	0	1	0	0	1	0	0	0	0	1	0	1	1	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0
Aridarum	1	1	0	1	1	1	0	5	0	0	0	1	0	?	0	0	0	1	0	0	0	0	1	0	1	1	0	1	0	0	0	0	1	0	4	1	0	1	b	0	0	0	0
Ariopsis	1	1	0	1	1	1	1	5	0	2	0	2	0	0	0	0	0	1	0	0	0	0	0	0	1	2	0	2	0	0	0	2	0	1	0	0	0	0	1	0	0	0	1
Arisaema	1	1	0	1	1	1	1	5	0	2	0	2	0	1	0	0	0	1	0	0	0	0	0	2	1	2	2	3	0	0	0	2	2	0	0	0	0	0	2	0	0	0	0
Arisarum	1	1	0	1	1	1	1	5	0	0	1	3	0	1	0	0	0	1	0	0	0	0	1	0	1	2	0	3	0	0	0	1	2	0	0	0	0	0	2	0	0	1	1
Arophyton	1	1	0	1	1	1	1	5	0	2	1	2	0	1	0	0	0	1	0	0	0	0	0	0	1	2	0	2	0	0	0	0	g	0	0	0	0	a	2	0	0	0	0
Arum	1	1	0	1	1	1	1	5	0	2	1	2	0	0	0	0	0	1	0	0	0	0	1	2	1	2	0	3	0	0	0	2	2	0	0	0	0	0	2	0	0	0	0
Asterostigma	1	1	0	1	1	1	1	5	0	0	1	a	0	1	0	0	0	1	0	0	0	0	1	0	1	1	1	3	0	0	0	2	2	0	0	0	0	0	2	0	0	0	0
Biarum	1	1	0	1	1	1	1	5	0	2	1	2	0	1	0	0	0	1	0	0	0	0	1	0	1	2	0	3	0	0	0	2	2	0	0	0	0	0	2	0	0	a	0
Bognera	1	1	0	1	1	1	1	5	0	0	2	0	1	?	?	0	?	1	0	0	0	0	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
Bucephalandra	1	1	0	1	1	1	1	5	0	0	1	1	0	?	0	0	0	1	0	0	0	0	1	0	1	1	0	1	0	0	0	0	1	0	4	1	0	1	1	0	0	0	0
Caladium	1	1	0	1	1	1	1	5	0	2	1	1	0	1	0	0	0	2	1	1	0	0	1	2	1	2	0	2	0	0	0	2	2	1	0	0	0	1	2	0	0	0	0
Calla	0	1	0	1	1	1	0	3	0	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
Callopsis	1	1	0	1	1	1	1	5	0	2	1	0	0	1	0	0	0	1	0	0	0	0	1	0	1	0	0	3	0	0	0	0	2	0	0	0	0	0	0	0	0	0	1
Carlephyton	1	1	0	1	1	1	1	5	0	2	1	2	0	1	0	0	0	1	0	0	0	0	0	0	1	2	0	2	0	0	0	2	2	0	0	0	0	1	0	0	0	0	1
Cercestis	1	1	0	1	1	1	1	5	0	0	1	0	1	1	1	1	1	1	0	0	0	0	1	0	1	1	0	3	0	0	0	0	3	0	0	0	0	0	2	0	0	0	0
Chlorospatha	1	1	0	1	1	1	1	5	1	2	1	a	0	0	0	0	0	2	1	1	0	0	1	2	1	2	0	2	0	0	0	1	1	0	0	0	0	1	2	0	0	0	0
Colletogyne	1	1	0	1	1	1	1	5	0	2	1	2	0	1	0	0	0	1	0	0	0	0	0	0	1	2	0	2	0	0	0	2	2	0	0	0	0	0	0	0	0	0	1
Colocasia	1	1	0	1	1	1	1	5	0	0	1	e	0	1	0	0	0	2	0	0	1	1	1	2	1	2	0	2	0	0	0	1	b	1	0	0	0	1	2	0	0	0	0
Croatiella	1	1	0	1	1	1	1	5	0	0	1	1	0	?	?	?	?	?	?	?	?	?	1	2	1	2	0	3	0	0	0	2	2	0	0	0	0	0	1	0	0	0	1
Cryptocoryne	1	1	0	1	1	1	1	5	0	0	1	1	0	1	0	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	0	0	0	0	1	0	1	0	2	1	1	1	0
Culcasia	1	1	0	1	1	1	0	5	0	2	1	0	1	1	1	1	1	0	0	0	0	0	1	0	1	1	0	3	0	0	0	0	3	0	0	0	0	2	1	0	0	0	0
Cyrtosperma	0	0	0	1	1	1	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	1	1	1	4	3	0	1	0	1	2	0	1	0	0	0	1	0	0	0	0
Dieffenbachia	1	1	0	1	1	1	1	5	0	0	2	a	0	1	0	0	0	1	0	0	0	0	1	0	1	0	0	1	0	0	0	0	1	0	a	0	0	0	2	0	0	0	1
Dracontioides	0	0	0	1	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	4	3	1	0	0	0	2	0	1	0	0	0	2	0	0	0	0
Dracontium	0	0	0	1	1	1	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	1	1	2	3	3	1	0	0	2	2	0	0	0	0	0	1	0	0	0	0
Dracunculus	1	1	0	1	1	1	1	5	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	2	1	2	2	3	0	0	0	2	2	0	0	0	0	0	2	0	0	0	0
Eminium	1	1	0	1	1	1	1	5	0	2	1	2	0	1	?	0	?	1	0	0	0	0	0	2	1	2	2	3	0	0	0	2	2	0	0	0	0	0	2	0	0	a	0
Epipremnum	0	1	1	1	1	1	0	2	0	1	1	0	1	1	0	0	0	0	0	0	0	0	1	0	1	1	0	1	1	0	0	0	3	0	0	0	0	2	1	0	0	0	0
Filarum	1	1	0	1	1	1	1	5	0	2	0	2	0	?	?	0	?	2	0	0	0	0	0	0	1	2	0	3	0	0	0	2	2	0	0	0	0	0	1	0	0	0	1
Furtadoa	1	1	0	1	1	1	1	5	0	0	0	1	0	?	1	1	1	1	0	0	0	0	1	0	1	1	0	1	0	0	0	0	1	0	4	0	0	0	1	0	0	0	0
Gearum	1	1	0	1	1	1	1	5	0	0	2	a	0	?	?	0	?	0	0	0	0	0	1	2	1	2	1	3	0	0	0	1	2	0	0	0	0	0	2	0	0	0	0
Gonatopus	1	0	0	1	1	1	1	1	0	0	2	0	1	1	0	0	0	0	0	0	0	0	1	0	1	1	5	3	0	0	0	1	2	0	0	0	0	0	2	0	0	0	0
Gorgonidium	1	1	0	1	1	1	1	5	0	0	1	0	0	?	0	0	0	1	0	0	0	0	1	0	1	1	1	3	0	0	0	2	2	0	0	0	0	1	1	0	0	0	0
Gymnostachys	0	0	0	0	0	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0
Hapaline	1	1	0	1	1	1	1	5	0	2	1	2	0	1	0	0	0	2	1	1	0	0	1	2	1	2	0	3	0	0	0	2	2	0	0	0	0	1	2	0	0	0	1
Helicodiceros	1	1	0	1	1	1	1	5	0	2	1	2	0	?	0	0	0	1	0	0	0	0	0	2	1	2	2	3	0	0	0	2	2	0	0	0	0	0	2	0	0	0	0
Heteropsis	0	1	0	1	1	3	0	2	0	0	1	0	1	1	0	0	0	0	0	0	0	0	1	0	1	2	0	3	0	0	0	0	3	0	0	0	0	2	1	0	0	0	0
Holochlamys	0	0	2	1	1	1	0	5	0	0	1	3	1	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Homalomena	1	1	0	1	1	1	1	5	0	0	0	1	0	1	1	0	1	1	0	0	0	0	1	0	1	1	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0
Incarum	1	1	0	1	1	1	1	5	0	0	1	0	0	?	?	?	?	?	?	?	?	?	1	0	1	1	1	3	0	0	0	2	2	0	0	0	0	0	1	0	0	0	1
Jasarum	1	1	0	1	1	1	1	5	0	0	1	0	0	1	?	0	?	2	1	1	0	0	1	0	1	2	0	2	0	0	0	0	2	0	3	0	0	1	2	0	0	0	1
Lagenandra	1	1	0	1	1	1	1	5	0	0	1	1	0	1	0	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	0	1	0	0	1	0	1	0	2	1	1	1	0
Lasia	0	0	0	1	1	1	1	0	0	0	1	0	1	?	0	0	0	0	0	0	0	0	1	2	1	2	2	3	0	1	0	0	1	0	1	0	0	2	1	0	0	0	0
Lasimorpha	0	0	0	1	1	1	1	0	0	0	0	0	1	?	?	?	?	0	0	0	0	0	1	1	1	b	4	3	0	1	0	1	2	0	1	0	0	0	1	0	0	0	0
Lysichiton	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1	2	0	1	0	0	0	1	0	0	0	0
Mangonia	1	1	0	1	1	1	1	5	0	0	1	0	0	?	?	0	?	0	0	0	0	0	1	0	1	0	0	3	0	0	0	2	2	0	0	0	0	0	2	0	0	0	0
Monstera	0	1	1	1	1	1	0	2	0	1	1	0	1	1	0	0	0	0	0	0	0	0	1	0	1	1	0	1	1	0	0	0	3	0	0	0	0	2	1	0	0	0	0
Montrichardia	1	1	0	1	1	1	1	5	0	2	2	1	0	0	1	0	0	1	0	0	0	0	1	1	1	1	0	3	0	a	0	0	2	0	1	0	0	2	2	0	0	0	0
Nephthytis	1	1	0	1	1	1	1	5	0	0	1	0	1	1	0	0	0	1	0	0	0	0	0	1	1	2	4	3	0	a	0	1	2	0	0	0	0	0	0	0	0	0	0
Orontium	0	0	0	0	0	0	?	0	0	0	2	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	2	0	1	0	0	0	0	0	0	0	0
Pedicellarum	0	0	0	1	1	3	0	0	0	0	0	0	1	?	?	?	?	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
Peltandra	1	1	0	1	1	1	1	5	0	0	1	b	0	1	0	0	0	1	0	0	0	0	1	2	1	2	0	1	0	0	0	0	2	0	1	0	0	1	2	0	0	0	0
Philodendron	1	1	0	1	1	1	1	5	0	0	1	1	1	1	1	1	1	1	0	0	0	0	1	0	1	1	0	1	0	0	0	0	1	0	a	0	1	0	2	0	0	0	0
Phymatarum	1	1	0	1	1	1	1	5	0	0	0	1	0	?	0	0	0	1	0	0	0	0	1	0	1	1	0	1	0	0	0	0	1	0	4	1	0	1	1	0	0	0	0
Pinellia	1	1	0	1	1	1	1	5	0	2	1	e	0	1	0	0	0	1	0	0	0	0	0	2	1	2	2	3	0	0	0	2	2	0	0	0	0	0	2	1	0	0	1
Piptospatha	1	1	0	1	1	1	1	5	0	0	a	1	0	1	0	0	0	1	0	0	0	0	1	0	1	1	0	1	0	0	0	0	1	0	4	1	0	1	1	0	0	0	0
Pistia	1	1	0	1	1	1	1	5	0	0	1	3	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	3	0	0	0	0	1	0	2	0	0	0	2	1	0	1	1
Podolasia	0	0	0	1	1	1	1	0	0	0	0	0	1	?	0	0	0	0	0	0	0	0	0	1	1	0	4	3	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0
Pothoidium	0	0	0	1	1	3	0	0	0	0	1	0	1	?	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
Pothos	0	0	0	1	1	3	0	0	0	0	0	0	1	?	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
Protarum	1	1	0	1	1	1	1	5	0	0	0	0	0	?	?	0	?	2	0	0	0	0	0	2	1	2	2	2	0	0	0	2	2	0	0	0	0	1	2	0	0	0	0
Pseudodracontium	1	1	0	1	1	1	1	5	0	0	1	3	0	1	0	0	0	1	0	0	0	0	1	1	1	2	3	3	0	0	0	2	2	0	0	0	0	0	1	0	0	0	0
Pseudohydrosme	1	1	0	1	1	1	1	5	0	0	3	1	0	1	?	0	?	0	0	0	0	0	1	1	1	2	0	3	1	1	0	2	2	0	0	0	0	0	2	0	0	0	0
Pycnospatha	0	1	0	1	1	1	1	0	0	0	1	0	0	?	?	0	?	0	0	0	0	0	1	1	1	1	3	3	1	0	0	2	2	0	0	0	0	0	1	0	0	0	0
Remusatia	1	1	0	1	1	1	1	5	0	2	1	2	0	1	0	0	0	2	0	0	1	1	1	2	1	2	0	2	0	0	1	2	4	1	0	0	0	1	2	0	0	0	0
Rhaphidophora	0	1	1	1	1	1	0	1	0	0	1	0	1	?	0	0	0	0	0	0	0	0	1	0	1	1	0	1	1	0	0	0	3	0	0	0	0	2	1	0	0	0	0
Rhodospatha	0	1	1	1	1	1	0	2	0	0	1	0	1	1	0	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	0	0	3	0	0	0	0	2	1	0	0	0	0
Sauromatum	1	1	0	1	1	1	1	5	0	2	1	2	0	1	0	0	0	1	0	0	0	0	0	2	1	2	2	3	0	0	0	2	2	0	0	0	0	0	2	0	0	1	0
Scaphispatha	1	1	0	1	1	1	1	5	0	2	1	0	0	?	?	0	?	2	?	?	?	?	1	0	1	2	0	2	0	0	0	2	2	1	0	0	0	1	1	0	0	0	0
Schismatoglottis	1	1	0	1	1	1	1	5	0	0	0	1	0	a	0	0	0	1	0	0	0	0	1	0	1	1	0	1	0	0	0	0	1	0	f	a	0	1	2	0	0	0	0
Scindapsus	0	1	1	1	1	1	0	2	0	1	1	0	1	a	0	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	0	0	3	0	0	0	0	2	1	0	0	0	0

Appendix 2 continued

	Spadix zonation	Spadix appendix	Basipetal flowering	Female zone length	Stamen thecae horns	Stamen connective	Stamen connation	Anther dehiscence	Staminodes in female zone	Ovary locules	Ovule type	Endosperm	Placentation	Chromosome base number (estimated)	Staminodes	Style	Ovule number	Perigon-tepal connation	Perigon - tepal apex	Male flowers	Leaf TS spongy aerenchyma	Leaf TS collenchyma	Leaf TS vascular bundles	Petiole TS collenchyma	Petiole TS ground tissue	Petiole TS vascular bundles	Cotyledon type	Cotyledon shape	First leaf	Hypocotyl presence	Primary root presence	Sporopollenin	Roots	Root branching	Leaf veins	Anther thecae	Rhaphides	Biforines
	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81
Acorus	0	0	0	–	0	0	0	0	–	0	3	0	1	0	–	0	0	0	0	–	2	5	0	6	2	0	0	0	1	0	0	0	0	0	0	0	0	–
Aglaodorum	1	0	0	0	0	1	0	2	1	0	0	1	2	a	0	0	1	–	–	0	e	2	1	2	2	1	?	?	?	?	?	1	0	0	0	0	1	1
Aglaonema	1	0	0	0	0	1	0	2	0	1	0	1	3	a	0	0	0	–	–	1	d	2	1	2	2	b	1	–	0	1	1	1	0	0	0	0	1	1
Alloschemone	0	0	0	–	0	0	0	0	–	1	0	?	3	2	–	0	1	–	–	–	1	0	1	6	1	0	?	?	?	?	?	0	0	0	0	0	1	0
Alocasia	4	1	0	0	0	1	2	1	0	1	0	0	3	2	0	0	0	–	–	0	d	q	b	4	3	2	0	0	0	0	1	0	0	0	0	1	1
Ambrosina	4	0	0	1	0	0	2	0	0	1	3	0	3	3	0	0	0	–	–	0	0	4	1	4	3	1	?	?	?	?	?	1	0	0	0	0	1	0
Amorphophallus	3	1	0	0	0	0	1	2	0	0	0	1	3	b	0	0	1	–	–	0	0	4	1	4	2	2	1	–	0	1	1	1	0	0	0	0	1	0
Amydrium	0	0	0	–	0	0	0	0	–	1	0	0	3	6	–	0	1	–	–	–	1	0	1	o	0	a	?	?	?	?	?	0	0	0	0	0	1	0
Anadendrum	0	0	0	–	0	0	0	0	–	1	0	1	3	6	–	0	1	1	0	–	0	0	0	p	0	0	1	–	0	1	0	0	0	0	0	0	1	0
Anaphyllopsis	0	0	1	–	0	0	0	1	–	1	0	0	3	1	–	0	1	0	1	–	0	5	0	?	?	?	?	?	?	?	?	?	0	0	0	0	1	0
Anaphyllum	0	0	0	–	0	0	0	1	–	1	0	1	2	1	–	0	1	0	1	–	3	5	1	0	3	0	?	?	?	?	?	?	0	0	0	0	1	0
Anchomanes	1	0	0	0	0	1	0	1	0	1	0	1	3	a	0	0	1	–	–	0	0	0	1	2	3	1	1	–	0	1	1	1	0	0	0	0	1	1
Anthurium	0	0	0	–	0	0	0	0	–	0	0	0	0	6	–	0	1	0	1	–	a	0	0	0	0	0	0	0	a	?	?	0	0	0	0	1	0
Anubias	1	0	0	0	0	1	2	0	0	0	0	0	0	c	0	0	0	–	–	0	d	0	0	o	e	b	0	1	1	0	0	1	0	0	0	0	1	1
Aridarum	g	a	0	0	1	0	a	2	0	1	2	0	q	0	0	0	0	–	–	0	b	a	1	0	j	b	?	?	?	?	?	1	0	0	0	0	1	1
Ariopsis	2	0	0	0	0	1	2	2	0	1	3	0	2	2	0	0	0	–	–	0	1	4	2	4	0	2	?	?	?	?	?	1	0	0	0	0	1	1
Arisaema	3	2	0	0	0	0	1	a	0	1	3	0	3	2	1	0	0	–	–	0	a	4	b	4	j	b	0	0	a	0	0	1	0	0	0	0	1	0
Arisarum	3	2	0	0	0	0	1	0	0	1	3	0	3	2	0	0	0	–	–	0	b	4	b	4	j	b	0	0	0	0	0	1	0	0	0	0	1	0
Arophyton	1	0	0	0	0	0	0	1	e	1	1	3	1	3	?	0	0	1	–	–	0	0	4	2	4	2	2	1	–	0	1	1	1	0	0	0	0	1	1
Arum	4	2	0	0	0	0	0	1	0	1	3	0	2	2	1	0	0	–	–	0	b	4	1	4	2	1	0	0	a	0	0	1	0	0	0	0	1	0
Asterostigma	1	0	0	0	0	0	1	2	2	1	0	0	0	3	8	0	0	1	–	–	1	0	5	a	0	2	2	?	?	?	?	?	1	0	0	0	0	1	0
Biarum	4	2	0	0	0	0	0	0	a	0	1	3	0	3	1	1	0	1	–	–	0	e	4	1	4	2	1	?	?	?	?	?	?	1	0	0	0	0	1	0
Bognera	2	0	0	0	0	0	1	2	0	0	1	0	?	3	8	0	0	1	–	–	0	?	?	?	?	?	?	?	?	?	?	?	1	0	0	0	0	1	?
Bucephalandra	4	1	0	0	1	0	0	2	0	1	3	0	3	1	1	0	0	–	–	0	0	0	1	4	1	1	?	?	?	?	?	1	0	0	0	0	1	0
Caladium	2	0	0	0	0	1	2	1	0	0	0	0	2	h	0	0	0	–	–	0	e	4	1	4	j	2	0	1	1	0	0	1	0	0	0	0	1	1
Calla	0	0	0	–	0	0	0	0	–	1	0	0	3	9	–	0	1	0	–	–	2	5	a	6	2	0	0	0	a	0	0	1	0	0	0	0	1	0
Callopsis	1	0	0	0	0	0	0	2	0	1	0	0	3	9	0	0	1	–	–	0	0	5	a	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0
Carlephyton	1	0	0	0	0	0	1	1	1	1	3	1	3	d	0	0	1	–	–	0	b	4	1	4	j	2	?	?	?	?	?	1	0	0	0	0	1	1
Cercestis	1	0	0	0	0	1	0	1	0	1	0	1	3	b	0	0	1	–	–	0	a	1	0	0	b	a	1	–	0	1	1	1	0	0	0	0	1	0
Chlorospatha	2	0	0	0	0	1	0	0	2	0	1	3	0	3	1	0	0	0	–	–	0	b	4	2	4	j	2	?	?	?	?	?	1	0	0	0	0	1	0
Colletogyne	1	0	0	0	0	0	0	0	1	1	3	1	3	d	0	0	1	–	–	0	2	4	1	4	2	2	1	–	0	1	1	1	0	0	0	0	1	1
Colocasia	4	2	0	0	0	1	2	1	0	1	2	0	2	2	0	0	0	–	–	0	j	4	b	4	j	b	0	1	1	?	?	1	0	0	0	0	1	1
Croatiella	1	0	0	0	0	1	b	0	1	2	3	?	3	8	0	0	1	–	–	0	?	?	?	?	?	?	?	?	?	?	?	1	0	0	0	0	1	?
Cryptocoryne	4	0	0	0	1	0	0	2	0	1	3	0	3	9	0	0	0	–	–	0	0	4	2	4	2	1	0	0	0	?	?	1	0	0	0	0	1	0
Culcasia	1	0	0	0	0	1	0	1	0	0	0	1	3	b	0	0	1	–	–	0	b	0	0	0	b	0	1	–	0	0	0	1	0	0	0	0	1	1
Cyrtosperma	0	0	1	–	0	0	0	0	–	1	1	0	2	1	–	0	0	0	1	–	k	0	0	0	j	0	0	0	0	0	0	0	0	0	0	0	1	0
Dieffenbachia	2	0	0	0	0	1	2	1	1	0	0	1	3	8	0	0	1	–	–	0	2	m	1	1	2	b	1	–	0	1	1	a	0	0	0	0	1	1
Dracontioides	0	0	1	–	0	0	0	1	–	0	0	0	0	1	–	0	1	0	1	–	0	5	0	6	3	1	0	0	0	0	0	0	0	0	0	0	1	0
Dracontium	0	0	1	–	0	0	0	1	–	0	0	0	0	1	–	0	1	0	1	–	l	0	a	0	3	a	?	?	?	?	?	0	0	0	0	0	1	0
Dracunculus	4	2	0	0	0	0	0	2	0	1	3	0	4	2	1	0	0	–	–	0	a	4	1	4	2	b	0	0	0	0	0	1	0	0	0	0	1	0
Eminium	4	2	0	0	0	0	0	1	0	1	3	0	3	2	1	0	1	–	–	0	b	4	1	4	2	a	?	?	?	?	?	1	0	0	0	0	1	0
Epipremnum	0	0	0	–	0	0	0	0	–	1	0	0	2	6	–	0	0	–	–	–	a	5	0	o	0	0	?	?	?	?	?	0	0	0	0	0	1	0
Filarum	3	2	0	0	0	0	0	2	0	1	?	1	?	4	0	0	1	–	–	0	0	4	1	4	0	1	?	?	?	?	?	1	0	0	0	0	1	0
Furtadoa	2	0	0	0	0	1	0	0	1	1	0	0	3	a	0	0	0	–	–	1	0	0	a	0	0	1	?	?	?	?	?	1	0	0	0	0	1	0
Gearum	2	0	0	0	0	1	2	2	1	0	3	1	3	8	0	0	1	–	–	0	e	0	1	0	2	1	1	–	0	1	1	1	0	0	0	0	1	1
Gonatopus	1	0	0	0	0	0	1	2	0	0	0	1	3	8	0	0	1	0	1	1	0	5	1	o	0	1	1	–	0	1	1	0	0	0	0	1	0
Gorgonidium	1	0	0	0	0	1	a	b	1	0	3	0	3	8	0	0	1	–	–	1	0	0	1	?	?	?	?	?	?	?	?	1	0	0	0	0	1	0
Gymnostachys	0	0	0	–	0	0	0	0	–	1	3	0	1	0	–	0	1	0	0	–	0	5	0	6	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Hapaline	2	2	0	0	0	1	2	1	0	1	0	1	3	b	0	0	1	–	–	0	b	4	1	4	j	1	1	–	0	1	1	1	0	0	0	0	1	1
Helicodiceros	4	1	0	0	0	0	0	1	0	1	3	0	4	2	1	0	0	–	–	0	a	4	1	4	2	1	0	0	0	0	0	1	0	0	0	0	1	0
Heteropsis	0	0	0	–	0	0	0	1	–	0	0	1	3	b	–	0	1	–	–	–	a	5	0	6	0	0	1	–	0	0	0	0	0	0	0	0	1	0
Holochlamys	0	0	0	–	0	0	0	0	–	1	0	0	3	6	–	0	1	1	0	–	b	0	1	0	0	1	?	?	?	?	?	0	0	0	0	0	1	0
Homalomena	2	0	0	0	0	1	0	1	1	0	0	0	0	a	0	0	0	–	–	0	d	0	1	0	e	1	?	?	?	?	?	1	0	0	0	0	1	1
Incarum	1	0	0	0	0	1	1	0	1	2	3	0	3	8	0	0	1	–	–	0	?	?	?	?	?	?	?	?	?	?	?	1	0	0	0	0	1	?
Jasarum	2	0	0	0	0	1	2	2	0	1	2	0	3	3	0	0	0	–	–	0	0	5	2	4	2	2	1	–	0	1	1	1	0	0	0	0	1	0
Lagenandra	4	0	0	0	1	0	0	2	0	1	3	0	3	9	0	0	0	–	–	0	0	4	1	4	d	2	0	0	0	0	0	1	0	0	0	0	1	0
Lasia	0	0	1	–	0	0	0	0	–	1	0	0	1	1	–	0	1	0	1	–	m	0	0	0	3	0	?	?	?	?	?	0	0	0	0	0	1	0
Lasimorpha	0	0	1	–	0	0	1	1	–	1	1	0	3	1	–	0	0	0	1	–	m	0	0	0	3	0	?	?	?	?	?	0	0	0	0	0	1	0
Lysichiton	0	0	0	–	0	0	0	0	–	0	3	1	0	2	–	0	1	0	0	–	2	3	0	3	2	0	1	–	0	0	0	0	0	0	0	0	1	0
Mangonia	3	0	0	0	0	1	1	b	1	0	0	0	0	8	0	0	1	–	–	0	0	0	a	0	2	a	0	0	0	0	0	1	0	0	0	0	1	0
Monstera	0	0	0	–	0	0	0	0	–	0	0	1	3	6	–	0	0	–	–	–	a	5	1	o	m	1	1	–	0	0	0	0	0	0	0	0	1	0
Montrichardia	1	0	0	0	0	1	0	1	0	1	0	1	3	c	0	0	1	–	–	0	b	0	a	6	2	2	1	–	0	1	1	1	0	0	0	0	1	0
Nephthytis	1	0	0	0	0	1	0	2	0	1	0	1	3	a	0	0	1	–	–	0	1	d	1	2	0	b	1	–	0	1	1	1	0	0	0	0	1	1
Orontium	0	0	0	–	0	0	0	0	–	1	0	1	3	1	–	0	1	0	0	–	2	5	0	6	2	0	1	–	0	0	0	0	0	0	0	0	1	0
Pedicellarum	0	0	0	–	0	0	0	0	–	0	0	1	3	0	–	0	1	1	0	–	0	5	0	6	0	0	?	?	?	?	?	0	0	0	0	0	1	0
Peltandra	4	1	0	0	0	1	2	2	1	1	2	1	2	2	0	0	0	–	–	0	2	4	b	4	2	2	1	–	0	0	0	1	0	0	0	0	1	1
Philodendron	2	0	0	0	0	1	0	b	0	0	0	0	0	c	0	0	0	–	–	0	b	0	1	0	n	d	0	1	1	0	0	1	0	0	0	0	1	1
Phymatarum	4	1	0	0	1	0	0	2	0	1	3	0	3	1	0	0	0	–	–	0	2	0	1	0	2	1	?	?	?	?	?	1	0	0	0	0	1	0
Pinellia	4	2	0	0	0	0	0	2	0	1	0	0	3	1	0	0	1	–	–	0	1	4	2	4	3	2	0	0	1	0	0	1	0	0	0	0	1	0
Piptospatha	g	a	0	0	0	0	0	2	a	1	j	0	j	1	0	0	0	–	–	0	a	0	1	0	1	1	0	1	1	0	0	1	0	0	0	0	1	1
Pistia	4	0	0	1	0	0	2	0	0	1	3	0	3	2	0	0	0	–	–	0	2	4	b	4	j	1	0	0	1	1	1	1	0	0	0	0	1	1
Podolasia	0	0	1	–	0	0	0	0	–	1	0	0	3	1	–	0	1	0	1	–	j	5	0	6	j	0	?	?	?	?	?	0	0	0	0	0	1	0
Pothoidium	0	0	0	–	0	0	0	0	–	1	0	1	3	0	–	0	1	0	0	–	0	5	0	6	0	0	?	?	?	?	?	0	0	0	0	0	1	0
Pothos	0	0	0	–	0	0	0	0	–	0	0	1	3	0	–	0	1	0	0	–	1	5	0	6	e	0	1	–	0	0	0	0	0	0	0	0	1	0
Protarum	4	2	0	0	0	1	2	2	1	1	2	0	3	2	0	0	0	–	–	0	1	4	1	?	?	?	?	?	?	?	?	1	0	0	0	0	1	1
Pseudodracontium	3	1	0	0	0	0	0	1	0	1	0	1	3	1	0	0	1	–	–	0	0	4	b	4	2	2	?	?	?	?	?	1	0	0	0	0	1	0
Pseudohydrosme	1	0	0	0	0	1	0	2	0	0	0	1	3	a	0	0	1	–	–	0	0	2	1	?	?	?	1	–	0	1	1	1	0	0	0	0	1	1
Pycnospatha	0	0	1	–	0	0	0	0	–	1	0	0	3	1	–	0	1	–	–	–	m	0	0	6	3	1	?	?	?	?	?	0	0	0	0	0	1	0
Remusatia	2	0	0	0	0	1	2	1	0	1	j	0	j	2	0	0	0	–	–	0	e	4	1	4	2	b	?	?	?	?	?	0	0	0	0	0	1	0
Rhaphidophora	0	0	0	–	0	0	0	0	–	0	0	0	0	6	–	0	0	0	–	–	a	5	a	o	0	0	?	?	?	?	?	0	0	0	0	0	1	0
Rhodospatha	0	0	0	–	0	0	0	0	–	0	0	0	0	2	–	0	0	0	–	–	a	0	0	6	1	0	?	?	?	?	?	0	0	0	0	0	1	0
Sauromatum	4	2	0	0	0	0	0	1	0	1	3	0	3	1	1	0	0	0	–	–	0	a	4	b	4	0	2	0	1	1	0	0	1	0	0	0	1	?
Scaphispatha	1	0	0	0	0	1	2	1	0	1	0	0	3	2	0	0	0	0	–	–	0	0	4	2	1	2	2	?	?	?	?	?	1	0	0	0	0	1	1
Schismatoglottis	4	1	0	0	0	0	0	2	a	1	0	0	2	1	0	0	0	–	–	0	k	0	b	0	j	1	0	1	1	0	0	1	0	0	0	0	1	1
Scindapsus	0	0	0	–	0	0	0	0	–	1	0	0	3	6	–	0	1	–	–	–	0	0	a	o	e	a	0	0	0	0	0	0	0	0	0	1	0	0

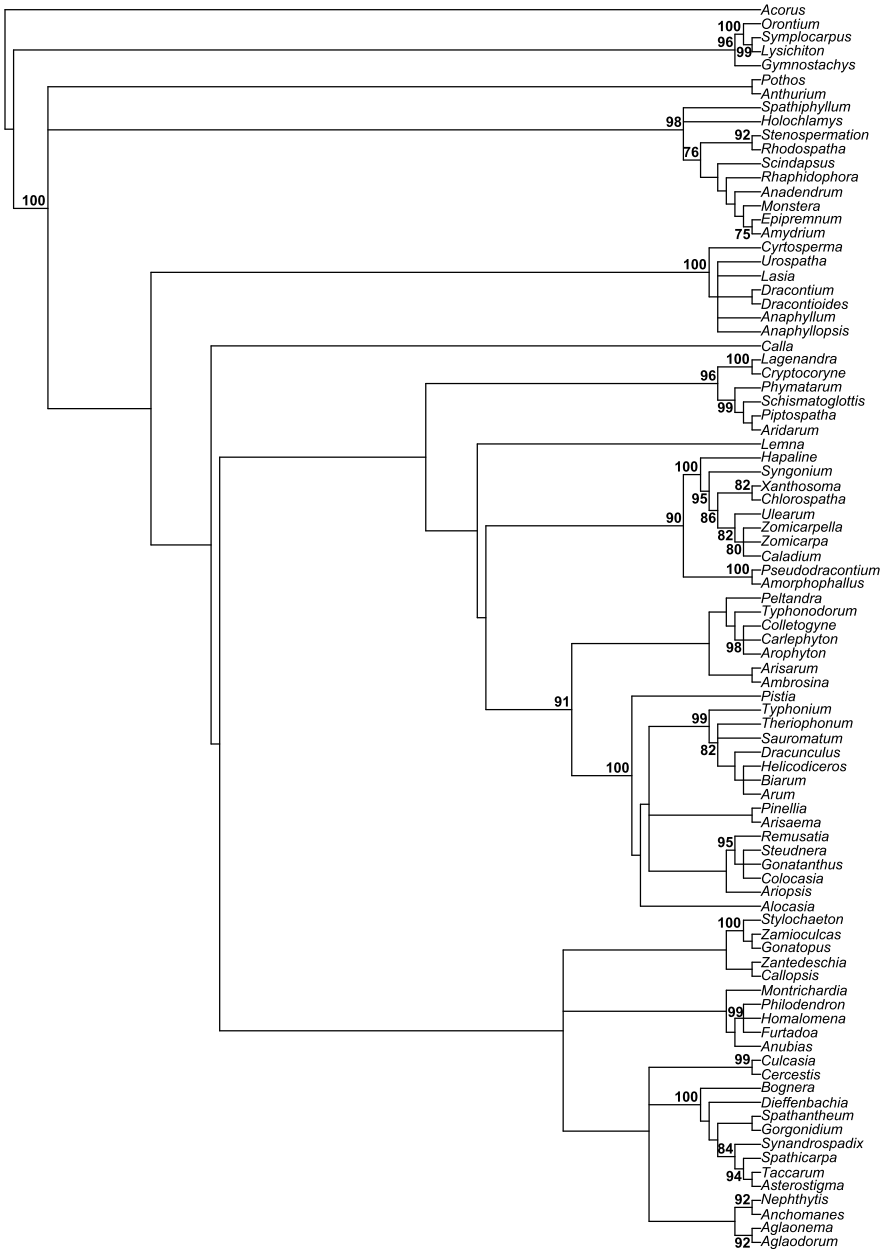
Appendix 2 continued

	Flower sexuality	Perigon	Trichosclereids	Spathe structure	Inflorescence stipe/peduncle	Shoot architecture	Phyllotaxy	Pollen aperture	Pollen units	Pollen shape	Mean pollen size	Pollen exine surface	Petiole geniculum	Pollen starch	Sclerotic hypodermis in roots	Endothecial thickenings	Resin canals in roots	Laticifers	Latex S bodies	Latex rubber (r) particles	Latex Y bodies	Latex R bodies	Primary leaf venation midrib	Primary leaf venation basal ribs	Leaf/petiole differentiation	Primary leaf venation marginal venation	Primary leaf venation anterior division or ultimate leaf lobes	Marginal form of leaf	Fine venation	Leaf fenestration	Prickles on stem or petiole	Stem producing erect shoots with bulbil	Stem type - thickening	Stem type - habit	Peltate leaves	Aquatic habit	Petiole ligule	Infravaginal squamules	Spathe behaviour
Spathantheum	1	1	0	1	1	1	1	5	0	0	2	1	0	1	0	0	0	0	0	0	0	1	2	1	1	0	3	0	0	0	0	2	2	0	0	0	0	0	0
Spathicarpa	1	1	0	1	1	1	1	5	0	0	1	1	0	1	0	0	0	1	0	0	0	0	1	0	1	0	3	0	0	0	0	1	2	0	0	0	0	0	0
Spathiphyllum	0	0	2	1	1	1	0	5	0	0	1	3	1	0	0	0	0	0	0	0	0	0	1	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0
Stenospermation	0	1	1	1	1	1	0	2	0	1	1	0	1	1	0	0	0	0	0	0	0	0	1	0	1	1	1	0	1	0	0	0	0	3	0	0	0	0	2
Steudnera	1	1	0	1	1	1	1	5	0	0	0	0	0	1	0	0	0	2	0	0	1	1	1	2	1	2	0	2	0	0	0	0	1	1	0	0	0	0	1
Stylochaeton	1	0	0	1	1	1	1	6	0	0	2	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	3	0	0	0	0	2	2	0	0	0	0	0
Symplocarpus	0	0	0	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1	2	2	0	1	0	0	0
Synandrospadix	1	1	0	1	1	1	1	5	0	2	1	2	0	1	0	0	0	1	0	0	0	0	1	0	1	1	0	3	0	0	0	0	2	2	0	0	0	0	0
Syngonium	1	1	0	1	1	1	1	5	0	0	1	d	0	1	0	0	0	2	1	1	0	0	1	2	1	2	0	2	0	0	0	0	3	0	0	0	0	0	1
Taccarum	1	1	0	1	1	1	1	5	0	0	2	e	0	1	0	0	0	1	0	0	0	0	1	1	1	1	3	3	0	0	0	0	2	2	0	0	0	0	1
Theriophonum	1	1	0	1	1	1	1	5	0	2	1	e	0	1	?	0	?	1	0	0	0	0	1	0	1	2	0	3	0	0	0	0	2	2	0	0	0	0	0
Typhonium	1	1	0	1	1	1	1	5	0	2	1	e	0	1	0	0	0	1	0	0	0	0	0	2	1	2	2	3	0	0	0	0	2	2	0	0	0	0	0
Typhonodorum	1	1	0	1	1	1	1	5	0	0	1	0	0	1	0	0	0	1	0	0	0	0	1	2	1	2	0	1	0	0	0	0	2	0	1	0	0	0	1
Ulearum	1	1	0	1	1	1	1	5	0	2	1	2	0	?	?	0	?	2	0	0	0	0	0	0	1	0	0	3	0	0	0	0	0	0	0	0	0	0	0
Urospatha	0	0	0	1	1	1	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	1	1	1	4	3	0	0	0	1	2	0	1	0	0	0	0
Xanthosoma	1	1	0	1	1	1	1	5	a	0	1	0	0	1	0	0	0	2	1	1	0	0	1	2	1	2	0	2	0	0	0	0	b	b	0	0	0	0	1
Zamioculcas	1	0	0	1	1	1	1	1	0	0	2	0	1	1	0	0	0	0	0	0	0	0	1	0	1	1	1	3	0	0	0	0	1	2	0	0	0	0	0
Zantedeschia	1	1	0	1	1	1	1	5	0	0	1	1	0	1	0	0	0	1	0	0	0	0	1	0	1	1	0	1	0	0	0	0	b	2	0	a	0	0	0
Zomicarpa	1	1	0	1	1	1	1	5	0	2	1	2	0	1	0	0	0	2	0	0	0	0	2	1	2	g	3	0	0	0	0	2	2	2	0	0	0	0	0
Zomicarpella	1	1	0	1	1	1	1	5	0	2	?	2	0	?	?	0	?	2	0	0	0	0	0	0	1	0	0	3	0	0	0	0	0	0	0	0	0	0	1
Scheuchzeria	0	0	0	0	0	–	0	5	2	2	?	0	0	?	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	1	0	1	0
Tofieldia	0	0	0	0	0	–	0	3	0	2	a	0	0	?	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2	0	1	0	0	0
Hypanc	0	0	0	0	0	0	0	0	0	0	?	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lemna	0	1	0	0	–	–	–	7	0	2	0	2	–	?	0	?	0	0	–	–	–	–	0	0	1	0	0	4	0	–	0	3	–	0	2	0	0	–		
Spirodela	0	1	0	0	–	–	–	7	0	2	0	2	–	?	0	?	0	0	–	–	–	–	0	0	1	0	0	4	0	–	0	3	–	0	2	0	0	–		
Landoltia	0	1	0	0	–	–	–	7	0	2	0	2	–	?	0	?	0	0	–	–	–	–	0	0	1	0	0	4	0	–	0	3	–	0	2	0	0	–		
Wolffia	0	1	0	2	–	–	–	7	0	2	0	2	–	?	–	?	–	0	–	–	–	–	–	–	2	–	–	–	–	–	0	3	–	–	2	–	0	–		
Wolffiella	0	1	0	2	–	–	–	7	0	2	0	2	–	?	–	?	–	0	–	–	–	–	–	–	2	–	–	–	–	–	0	3	–	–	2	–	0	–		

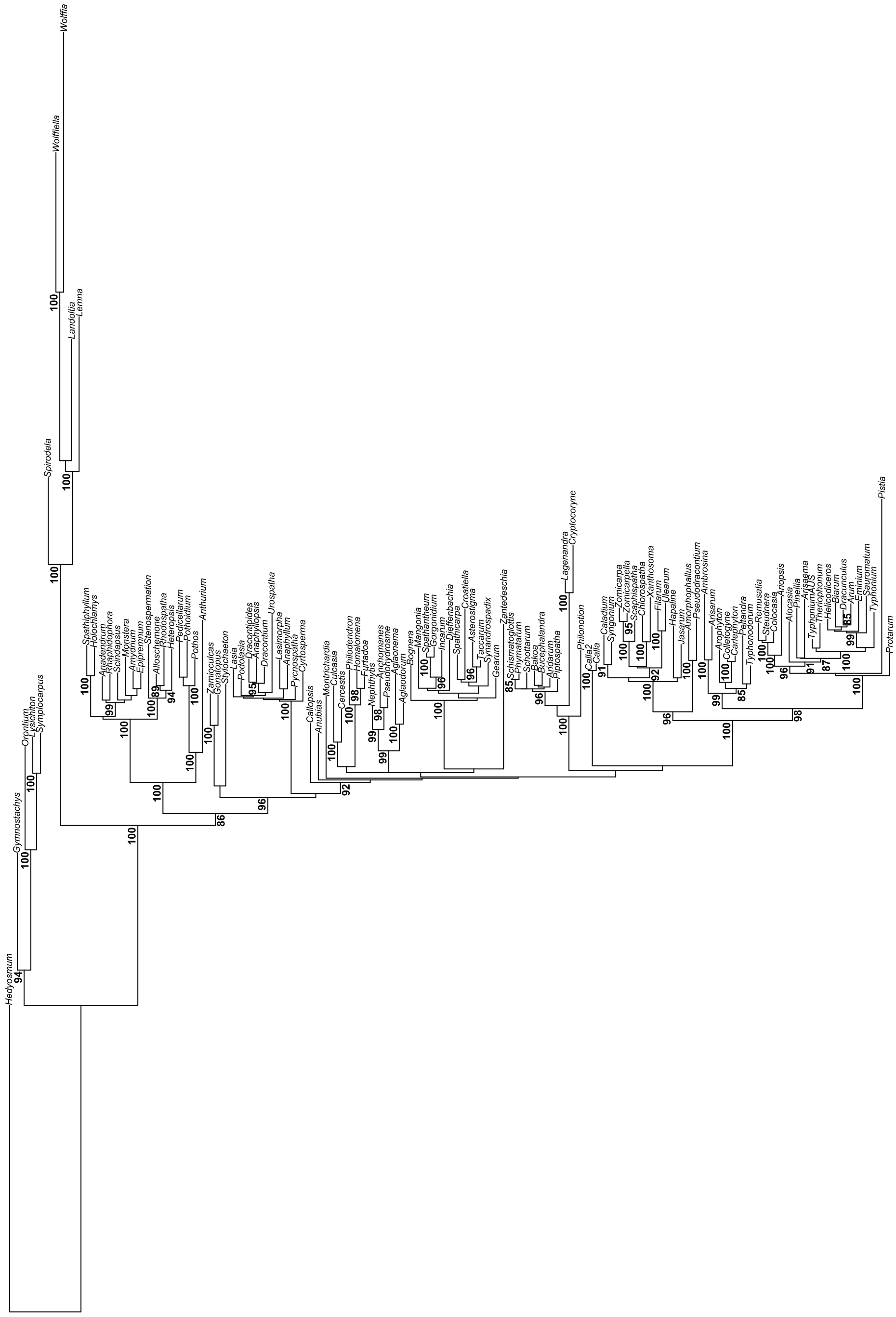
	Spadix zonation	Spadix appendix	Basipetal flowering	Female zone length	Stamen thecae horns	Stamen connective	Stamen connation	Anther dehiscence	Staminodes in female zone	Ovary locules	Ovule type	Endosperm	Placentation	Chromosome base number (estimated)	Staminodes	Style	Ovule number	Perigon-tepal connation	Perigon - tepal apex	Male flowers	Leaf TS spongy aerenchyma	Leaf TS collenchyma	Leaf TS vascular bundles	Petiole TS collenchyma	Petiole TS ground tissue	Petiole TS vascular bundles	Cotyledon type	Cotyledon shape	First leaf	Hypocotyl presence	Primary root presence	Sporopollenin	Roots	Root branching	Leaf veins	Leaf shape	Anther thecae	Rhaphides	Biforines
Spathantheum	1	0	0	0	0	1	2	e	1	2	2	0	3	8	0	0	1	–	–	1	a	0	1	0	2	b	0	0	0	0	0	0	0	0	1	0		
Spathicarpa	1	0	0	0	0	1	2	2	1	1	3	0	3	8	0	0	1	–	–	1	0	0	1	1	0	b	0	0	1	0	0	1	0	0	0	1	0		
Spathiphyllum	0	0	0	–	0	0	0	0	–	0	0	0	0	6	–	0	0	0	0	–	0	0	0	0	k	a	0	0	1	0	0	0	0	0	1	0		
Stenospermation	0	0	0	–	0	0	0	0	–	0	0	0	3	2	–	0	0	–	–	–	0	5	a	0	b	1	0	0	0	0	0	0	0	0	1	0		
Steudnera	1	0	0	0	0	1	2	1	1	1	0	0	2	2	0	0	0	–	–	0	b	4	1	4	2	b	0	0	1	0	0	1	0	0	1	0		
Stylochaeton	1	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	1	1	1	2	0	1	0	d	a	?	?	?	?	?	?	1	0	0	0	1	1	
Symplocarpus	0	0	0	–	0	0	0	0	–	1	3	1	1	6	–	0	1	0	0	–	2	0	0	o	2	0	1	–	0	0	0	0	0	0	0	0	1	0	
Synandrospadix	2	0	0	0	0	1	2	0	1	0	3	0	3	8	0	0	1	–	–	1	e	0	1	0	2	b	0	0	0	0	0	1	0	0	0	0	1	0	
Syngonium	2	0	0	0	0	1	2	1	0	1	0	1	3	2	0	0	1	–	–	0	b	4	1	4	j	b	1	–	0	?	?	1	0	0	0	0	1	1	
Taccarum	1	0	0	0	0	1	2	b	1	0	0	0	0	8	0	0	1	–	–	1	a	0	2	0	b	b	?	?	?	?	?	?	1	0	0	0	1	0	
Theriophonum	4	2	0	0	0	0	0	0	b	0	1	3	0	4	e	1	0	0	–	–	0	0	4	1	4	2	1	?	?	?	?	?	?	1	0	0	0	1	0	
Typhonium	4	2	0	0	0	0	0	0	b	0	1	3	0	3	1	1	0	1	–	–	0	a	4	2	4	0	b	?	?	?	?	?	?	1	0	0	0	1	0	
Typhonodorum	4	1	0	0	0	0	1	2	1	1	1	3	1	3	2	0	0	1	–	–	0	2	4	2	4	2	2	1	–	0	a	a	1	0	0	0	0	1	1	
Ulearum	4	2	0	0	0	0	0	0	2	0	1	0	1	3	4	0	0	1	–	–	0	0	4	b	q	0	1	?	?	?	?	?	?	1	0	0	0	1	0	
Urospatha	0	0	1	–	0	0	0	0	1	–	0	0	0	0	1	–	0	0	0	1	–	e	0	0	0	3	0	0	0	0	0	0	0	0	0	0	1	0	
Xanthosoma	2	0	0	0	0	1	2	1	0	0	0	0	0	1	0	1	0	–	–	0	b	4	a	4	j	b	0	1	1	1	1	0	1	0	0	0	1	1	
Zamioculcas	1	0	0	0	0	0	0	0	0	0	0	1	3	8	0	0	1	0	1	1	0	0	0	0	0	1	?	?	?	?	?	?	0	0	0	0	1	0	
Zantedeschia	1	0	0	0	0	0	1	0	2	a	0	0	0	0	7	0	0	0	–	–	0	2	1	2	1	2	2	0	0	1	0	0	1	0	0	0	1	1	
Zomicarpa	3	2	0	0	0	0	0	0	2	0	1	0	0	3	5	0	0	0	–	–	0	1	4	b	4	b	2	0	0	1	0	0	0	0	0	0	1	1	
Zomicarpella	3	2	0	0	0	0	0	0	2	0	1	0	0	3	1	0	0	0	–	–	0	a	4	1	4	0	b	?	?	?	?	?	0	0	0	0	1	1	
Scheuchzeria	0	0	0	–	0	0	0	0	–	0	0	1	3	3	–	0	1	0	0	–	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	0	0	0	–	
Tofieldia	0	0	0	–	0	0	0	0	–	0	a	0	0	6	–	0	0	0	0	–	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	0	0	0	–	
Hypanc	0	0	0	–	0	0	0	0	0	0	?	–	0	0	0	0	0	0	0	–	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	0	0	0	–	
Lemna	–	–	–	–	0	0	0	0	–	1	0	0	3	5	–	0	a	–	–	–	2	5	?	6	–	–	0	0	1	1	1	0	0	1	0	0	1	0		
Spirodela	–	–	–	–	0	0	0	0	–	1	0	0	3	5	–	0	1	–	–	–	2	5	?	6	–	–	0	0	1	1	1	0	0	1	0	0	1	0		
Landoltia	–	–	–	–	0	0	0	0	–	1	0	0	3	5	–	0	1	–	–	–	?	5	?	6	–	–	0	0	1	1	1	0	0	1	0	0	?	?		
Wolffia	–	–	–	–	0	0	0	0	–	1	3	0	3	5	–	0	1	–	–	–	2	5	?	6	–	–	?	?	?	?	?	?	0	1	–	1	1	1	0	
Wolffiella	–	–	–	–	0	0	0	0	–	1	3	0	3	5	–	0	1	–	–	–	2	5	?	6	–	–	?	?	?	?	?	?	0	1	–	1	1	1	0	

6.7 Online Supporting Material

Supplementary Figure S1: 90% majority rule consensus tree of 10000 most parsimonious trees obtained from maximum parsimony analysis of the chloroplast restriction site data of French et al. (1995; 488 char.) with bootstrap values obtained from 1000 replicates indicated above nodes.



Supplementary Figure S2: Maximum likelihood phylogeny of 115 species obtained with RAxML from combined chloroplast data (4156 nt) with bootstrap values obtained from 1000 replicates indicated above nodes.



Danksagung

Zunächst möchte ich meiner Doktormutter Susanne Renner danken, dass sie mir die Möglichkeit gegeben hat, die Dissertation unter ihrer Leitung durchzuführen. Ihr Engagement in der Betreuung, die Möglichkeit, zu jeder Tages- und Nachtzeit mit Fragen, Problemen oder für anregende Diskussion an ihre Tür zu klopfen, aber vor allem ihr Anspruch haben die Zeit in München für mich zu einer spannenden Herausforderung und lehrreichen Zeit gemacht.

Ich danke den Ureinwohnern der Merxmüllerwohnung Stefanie Volz, Hanno Schäfer, Boris Schlumpberger und Christoph Heibl, dass sie mich damals so toll aufgenommen haben und, zusammen mit den neu Dazugekommenen Ryan Oyama, Norbert Holstein, Patrizia Sebastian, Pablo Duchen, Ingo Michalak und Lars Nauheimer meine Zeit in München zu einem Erlebnis gemacht haben, für die nahezu immer gute Arbeitsatmosphäre, die besonderen Mittagspausen und Hilfe, vor allem von Ryan, bei allen Problemen! Bei Christoph Heibl möchte ich mich noch zusätzlich dafür bedanken, dass er mich mit R und LaTeX bekannt gemacht hat, und für seinen Enthusiasmus beim Erkunden neuer Methoden zur Datenanalyse, was mich sehr motiviert hat und von dem ich all die Zeit profitiert habe.

Josef Bogner verdanke ich, dass die Araceen die Pflanzenfamilie war, mit der ich mich die letzten dreieinhalb Jahre beschäftigt habe. Jederzeit war er bereit zu einem Gespräch über "seine" Araceen, hatte Antworten auf alle meine Fragen und versorgte mich immer mit der aktuellsten Literatur.

I want to thank Ned Friedman for the possibility to spend 3 inspiring months in his lab to learn not only scientific theory and practice, but also for life; his Lab Group for the warm welcome; Noah Greenberg and Amy Rosenblum for giving me a home in this period; and finally Kirsten Ryerson for the nice time spent together, and especially for the awesome trip to Utah.

Hanna Weiss-Schneeweiss möchte ich für die viele Zeit danken, die sie sich genommen hat, um mich in ihrem Labor in die Methode der Fluoreszenz *in situ* Hybridisierung einzuarbeiten und um bei Fragen zur Theorie von Chromosomenevolution zu helfen.

Wilbert Hetterscheid danke ich, weil er mir bereitwillig Blatt-, aber auch Lebendmaterial von allen Arten seiner umfangreichen *Typhonium*-Sammlung zur Verfügung gestellt hat, und darüber hinaus sein Wissen und seine Daten bei allen Fragen zu der Gattung.

I thank Simon Mayo because he offered me the possibility to collaborate in the analysis of the morphological data matrix and the resulting inspiring discussions.

Mila Vosyka und Eva Facher danke ich für ihre Hilfe bei Problemen jedweder Art, bei Autotransporten, beim Mikrotomieren und Mikroskopieren, bei der Laborarbeit, bei der Laborsuche und -einrichtung, bei Organisatorischem und beim Zählen von Chromosomen.

Allen in den "Botanischen Staatsanstalten" Verbliebenen, sowie den im Garten tätigen Biologen und anderen Mitarbeitern für die Zusammenarbeit in den verschiedensten Bereichen und das allgemeine tägliche Miteinander.

Ich danke Dirk Metzler für die Begutachtung der Arbeit.

I thank Bob Ricklefs for helpful comments.

Tanja Stadler danke ich für die Hilfe beim Installieren ihres Programms Cass zur Simulation von Bäumen.

Ich danke Alexander Kocyan für die Einarbeitung in das Labor.

Ich danke Matthias Erben für die stets schnelle Hilfe bei allen bürokratischen und das Gewächshaus betreffenden Angelegenheiten und das Gießen.

Wolfram Lobin und Alan Galloway danke ich für die Bereitstellung von Blattmaterial.

Herrn Schottler danke ich für die Pflege meiner *Typhonium*-Sammlung.

Ich danke meinen Eltern, meinen Geschwistern und meiner Oma, die immer für mich da sind und mir viel Freude geben, sowie meiner Schwiegermama, die mich über lange Zeit zu Hause verwöhnt hat!

Ευχαριστὼ τον ὰνδρα μου Martin που èιναι πὰνδα μαζὶ μου. Σ'αγαπὰω.