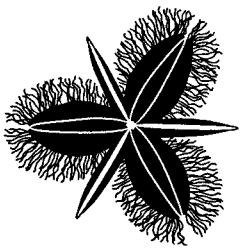

A PHYLOGENETIC ANALYSIS OF THE MONOCOTYLEDONS BASED ON MORPHOLOGICAL AND MOLECULAR CHARACTER SETS, WITH COMMENTS ON THE PLACEMENT OF *ACORUS* AND HYDATELLACEAE

Dennis W. Stevenson[A], *Jerrold I. Davis*[B], *John V. Freudenstein*[C], *Christopher R. Hardy*[AB], *Mark P. Simmons*[B] *and Chelsea D. Specht*[AD]

[A]New York Botanical Garden, Bronx, New York 10458, USA.
[B]L. H. Bailey Hortorium, Cornell University, Ithaca, New York 14853, USA.
[C]Department of Evolution, Ecology, & Organismal Biology, Ohio State University Columbus, Ohio 43210, USA.
[D]Department of Biology, New York University, New York 10012, USA.

Abstract

A phylogenetic analysis of the monocotyledons was conducted using 50 examplar monocot genera and four related dicot genera. Four parallel data sets were analyzed. The data sets represent restriction-site variation in the inverted repeat region of the chloroplast genome, nucleotide sequence variation in the chloroplast-encoded gene *rbc*L and in the mitochondrion-encoded gene *atp*A, and 152 morphological characters. These data sets were analyzed separately to assess characteristics of the data sets and in combined analyses to determine relationships among the sampled taxa. Of the four character sets, *rbc*L had the lowest retention index and data decisiveness. The combined analysis resulted in a single most parsimonious tree. Among the salient features of this tree are a monophyletic group consisting of *Acorus* and Alismatales that is sister to the rest of the monocots; the placement of the enigmatic *Trithuria* (Hydatellaceae) with *Xyris*; a monophyletic group containing the Arecanae, Bromelianae, Commelinanae, and Zingiberanae; and the Velloziales as sister to a clade consisting of Cyclanthanae plus Pandananae.

Key words: *Acorus*, *atp*A, Hydatellaceae, monocots, phylogeny, *rbc*L, *Trithuria*

INTRODUCTION

As a result of the first conference on the systematics and evolution of the monocots held at the Royal Botanic Gardens, Kew, numerous papers appeared that focused on phylogenetic analyses of relationships within the monocots. This work, coupled with subsequent research, has revealed numerous problematic taxa, numerous problems in our understanding of characters and character states, and problems concerning sources of data and methods of data analysis. As a result, we focused on building a data matrix by beginning with a few taxa for which multiple matching data sets could be developed.

We also decided to use trees produced in previously published analyses of monocots (Duvall *et al.* 1993a, 1993b; Chase *et al.* 1995; Stevenson and Loconte 1995; Davis 1995) as starting points to choose taxa, both as representatives of larger clades and as problematic taxa in term of disparate placements among the various trees produced by those analyses. We further decided to use these trees as elements of evidence in a larger process of reciprocal illumination. That is, we would use the trees to detect morphological characters that appeared to be homoplasious. When characters appear homoplasious as a result of a parsimony analysis, the structure of the tree indicates that the character state is not homologous. This inference, in turn, can be tested in some cases by splitting polymorphic terminals into monomorphic terminals, in other cases by filling in missing values with new observations, and still in others by examining ontogenetic or biosynthetic pathways that may provide further evidence of nonhomology.

In particular, we were concerned with the placement of *Acorus* and the Hydatellaceae. The former case was intriguing because of the disparate results that were obtained between analyses based

on *rbc*L gene sequences (Duvall *et al.* 1993a, 1993b; Chase *et al.* 1995) and those based on morphology (Stevenson and Loconte 1995; Chase *et al.* 1995). In the case of Hydatellaceae, the placement of the family is quite uncertain (Hamann 1998), and until now there was no available nucleotide sequence data for it. The analysis presented here represents the beginning of a long term study. The results of this paper will be used to direct further taxon sampling and character coding.

Materials and Methods

Because of space limitations the morphological matrix is not included; however, it is available upon request from the authors. The characters, character states, and codings are derived from Stevenson and Loconte (1995) and Chase *et al.* (1995) and are identical to those used in Rudall *et al.* (1999). Characters concerning root anatomy, particularly root hair development, have been re-examined for accuracy. Because of the paucity of data for most taxa (Stevenson and Loconte 1995), most root anatomy coding is the result of new observations. Data for the latter were obtained using epi-illumination light microscopy and scanning electron microscopy. For *Trithuria*, the observations on root, leaf, and stem anatomy and development are new.

Although the morphological data subset was analyzed separately for the study of data decisiveness, a separate morphology tree is not presented here but rather the results of morphological data are presented from the combined analysis. All multistate morphological characters were treated as nonadditive (unordered) in the phylogenetic analysis. Data was entered using Dada (Nixon 1998a) and trees and character optimization patterns on those trees were viewed with Clados (Nixon 1998b).

The taxon sample for the present analysis is the set of 53 genera in the restriction site analysis of Davis (1995), plus *Trithuria* J.D. Hook. The three molecular character sets are the original restriction site data, plus nucleotide sequence variation for two genes, the plastid-encoded *rbc*L and the mitochondrion-encoded *atp*A. Most of the *rbc*L sequences were taken from publicly available sources, though a few new ones were generated for this study. In contrast, we generated all of the *atp*A sequences not previously reported by Davis *et al.* (1998). All DNA sequences generated for the present study, except those from *Trithuria*, used the same accessions listed by Davis (1995) for the restriction site data. The voucher for *Trithuria* is *T. submersa* J.D. Hook., collected in Victoria, Australia (*Doust 1123*, MELU). Restriction site data were not generated for this accession, so the data matrix consists of morphological characters plus *rbc*L and *atp*A sequences for all 54 taxa, and restriction sites for all taxa except *Trithuria*. As this is a preliminary report of a larger study still in progress, final corrections have not yet been made on all DNA sequences, and GenBank accession numbers therefore are not yet available for the new sequences. The data matrix used in the analysis is available on request.

The restriction site data set, as previously published by Davis (1995), comprises 89 cladistically informative sites and two informative length variants. The latter two characters, and all other structural variants in the plastid and mitochondrial genomes (as detected by restriction site mapping or nucleotide sequencing), were included in the 'morphological' data set. As described by Davis *et al.* (1998), restriction sites in regions that have been deleted in some taxa, for example sites within *ORF2280* which is deleted in *Oryza* and other taxa (Shinozaki *et al.* 1986; Hiratsuka *et al.* 1989; Shimada and Sugiura 1991; Downie *et al.* 1994; Davis 1995), were scored as unknown in those taxa. Similarly, nucleotides of *atp*A in regions that are deleted in some taxa were scored as unknown in those taxa.

The portion of *rbc*L used in the present analysis comprises 1,398 nucleotide positions, corresponding to positions 31 (relative to the first position of the start codon in *Oryza*) through 1,428. The portion of *atp*A used in this analysis comprises 1,272 aligned nucleotide positions, corresponding to 1,266 bp in the coding region of *Oryza* (positions 98 through 1,363; Kadowaki *et al.* 1990) plus 6 additional bp to account for inferred insertions relative to the *Oryza* sequence. New sequences were generated using standard automated sequencing procedures and previously published primers (Davis *et al.* 1998). Deletions of two sizes (3 and 6 bp) were detected in *atp*A of various taxa (see below) in the region between sites 585 and 603, as described by Davis *et al.* (1998). All observed 3-bp deletions are potentially alignable to the same positions, as are all 6-bp deletions, so both are potentially cladistically informative. Also, some of the potential aligned positions for the 3-bp deletions place them within some of the potential aligned positions for the 6-bp deletions, so it is possible to interpret deletions of each size as transformed states of deletions of the other size. Consequently, the deletions were coded for analysis as a single nonadditive (i.e., unordered) multistate character with three cladistically informative states (undeleted, 3-bp deletion, 6-bp deletion) and, as stated above, this character was included in the morphological character set. A few other taxa had autapomorphic insertions and deletions of various sizes in the same region, and additional states were used to accommodate those taxa.

The four character sets (morphology, restriction sites, *rbc*L, and *atp*A) were analyzed separately and as a combined matrix. Also, to facilitate comparisons of qualities of the four character sets, each of the three that had been scored for all 54 taxa (morphology, *rbc*L, and *atp*A) was analyzed separately with *Trithuria* excluded; this resulted in parallel analyses of the four character sets as scored for the same 53 taxa. Cladistic analyses were conducted with NONA (Goloboff 1993), using the default settings *amb-* (clades resolved only if support is unambiguous) and *poly=* (polytomies allowed), and *Nymphaea* was used as the outgroup for purposes of rooting (Nixon and Carpenter 1994). Each tree search involved 1,000 subsearches, with each of the subsearches involving construction of a Wagner tree using a random taxon entry sequence. This was followed by *tbr* swapping with up to 10 most-parsimonious trees retained from each replicate that had yielded most-parsimonious trees (*hold/10 mult*1000*). After all replicate searches had been conducted, shortest trees retained from the subsearches were then swapped to completion (*max**). The consistency index (CI; Kluge and Farris 1969) and retention index (RI; Farris 1989) for the entire data set and for each of the four character sets were obtained directly from NONA. Data decisiveness (DD; Goloboff 1991; also see Davis *et al.* 1998) was calculated for each of the four character sets using total character variation and minimum tree length as obtained from NONA,

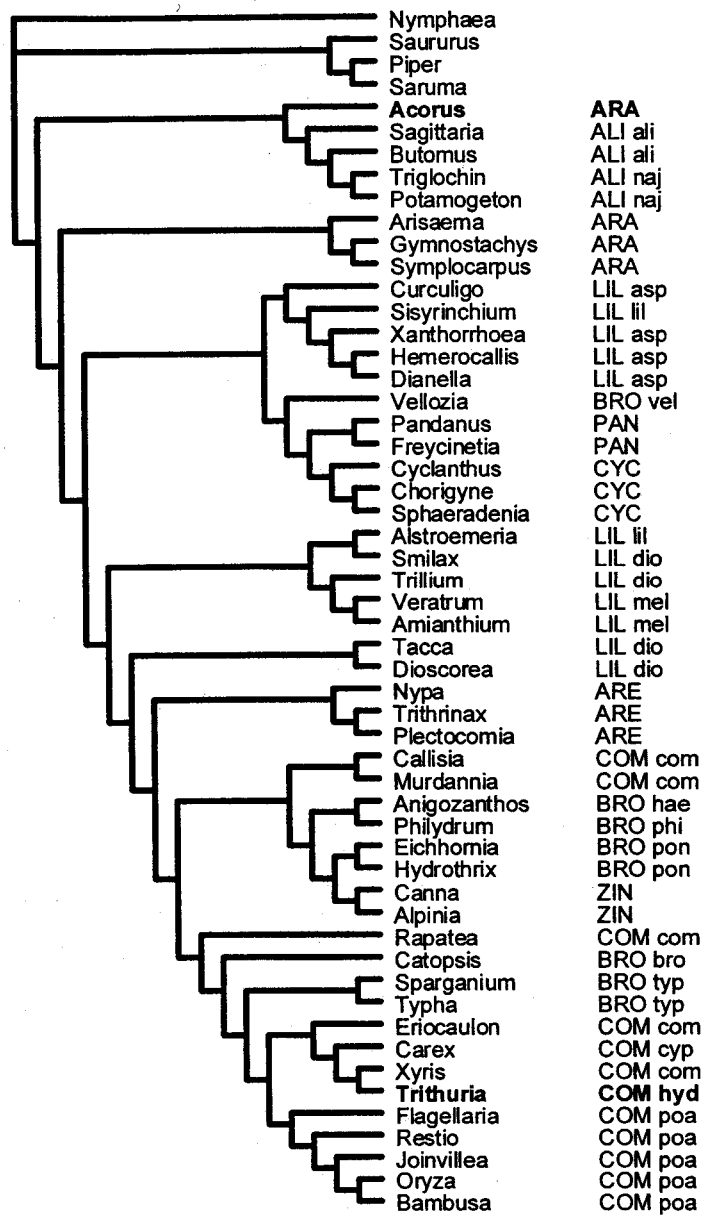

Fig. 1. Cladistic relationships among 50 genera of monocots and four genera of dicots as resolved by simultaneous analysis of morphological, restriction site, *rbc*L, and *atp*A character sets. *Acorus* and *Trithuria*, the relationships of which are discussed in the text, are in bold. Assignments of monocot genera to superorder and order in the system of Dahlgren *et al.* (1985) are signified by three-letter codes (capital letters for superorders; lower case letters for orders; ordinal placements indicated only for cases in which a superorder includes more than one order): ALI = Alismatanae; ARA = Aranae; ARE = Arecanae; BRO = Bromelianae; COM = Commelinanae; CYC = Cyclanthanae; LIL = Lilianae; PAN = Pandananae; ZIN = Zingiberanae; ali = Alismatales; asp = Asparagales; bro = Bromeliales; com = Commelinales; cyp = Cyperales; dio = Dioscoreales; hae = Haemodorales; hyd = Hydatellales; lil = Liliales; mel = Melanthiales; naj = Najadales; phi = Philydrales; poa = Poales; pon = Pontederiales; typ = Typhales; vel = Velloziales.

and mean lengths of 100,000 randomly constructed cladograms as obtained from PAUP (Swofford 1993).

Results and Discussion

For the set of 53 taxa (i.e., *Trithuria* excluded), 421 (30.1%) of the 1398 sampled *rbc*L nucleotides are cladistically informative, as are 260 (20.4%) of the 1272 of the inferred *atp*A sites. Of the 152 morphological characters, 125 (82.2%) are informative. The reason there are uninformative morphological characters is that this matrix contains a subset of taxa from a much larger morphological matrix and thus some of the characters are uninformative in this context. All of the characters were left in to accommodate taxa to be added in the future. The nucleotide characters, along with 125 informative morphological characters and 89 informative restriction site characters, resulted in a combined matrix of 895 characters for the 53 taxa. The consistency index (CI, based only on informative characters), retention index (RI), and data decisiveness (DD) for the four character sets are provided in Table 1. Inclusion of *Trithuria* in the data matrix caused 19 additional characters to become informative, resulting in a combined data set of 914 informative characters. Analysis of this matrix yielded one most-parsimonious tree of 4310 steps, CI 0.32, and RI 0.49 (Fig. 1).

Table 1. Characteristics of four character sets as scored for 49 monocot and four dicot taxa (*Trithuria* excluded; see text).

Character Set	Number of Informative Characters	Consistency Index	Retention Index	Data Decisiveness
*rbc*L	421	0.32	0.47	0.41
*atp*A	260	0.44	0.59	0.56
Morphology	125	0.28	0.59	0.54
Restriction sites	89	0.34	0.60	0.56

Of the four separate character sets, *atp*A has a CI of 0.44, while those of the other three character sets, ranging between 0.28 and 0.34, are substantially lower (Table 1). A somewhat different pattern is seen for RI and DD; again *rbc*L has among the lowest numbers, but in these cases all three other data sets have substantially higher scores. The RI and DD for *atp*A, morphology, and restriction sites range from 0.59 to 0.60 and 0.54 to 0.56, respectively, and substantially lower scores were obtained for *rbc*L (0.47 and 0.41, respectively). Thus, a bimodality is seen in each of these three indices, with *atp*A always in the class with highest score, *rbc*L always in the class with the lowest score, and morphology and restriction sites variously distributed.

There are more *rbc*L characters than there are for any of the other three character sets, but, as suggested by Davis *et al.* (1998), DD seems to reflect an intrinsic attribute of a data set, and does not seem to be affected substantially by number of characters, as long as a sufficient minimum number of characters is sampled. Also, all four character sets were drawn from an identical taxonomic sample. Thus, at least for this taxon sample, it seems appropriate to recognize the distinctions between these four character sets as valid. The implications of these distinctions, however, are not obvious. Goloboff (1991) proposed DD as a measure of the extent to which a data matrix favors one or more trees relative to other possible trees. Thus, a matrix that resolves a small number of trees that are many steps shorter than other possible trees is regarded as decisive, relative to one for which many or all possible trees are identical or similar in length. Decisiveness should not be equated with veracity, for it is surely possible for a data set to provide strong evidence of untrue relationships. However, we have suggested elsewhere that indecisive data sets may have a tendency not to be measurably incongruent with other data sets, because a data set that does not strongly favor any particular set of relationships is not likely to provide a strong enough signal to discernibly contradict that of another data set (Davis *et al.* 1998). Thus, we suggest that if a character set with a high DD is one that provides strong evidence for some set of relationships, then differences between the relationships supported by two different character sets of high DD should be easier to discern than differences between character sets of low DD. In short, DD may, in a general sense, provide an index of disprovability, and for that reason alone, data sets of high DD should be preferred to those of low DD. On these grounds, we note that *atp*A, morphology, and restriction sites compare favorably to *rbc*L, and that further exploration of these matters is in order.

General Tree Topology and Groupings

A brief discussion of relationships supported by the single most-parsimonious tree follows. We use the taxonomic system of Dahlgren *et al.* (1985) as the principal basis for comparison of our results with previous systems, with the general distinction that superorders are designated with the suffix -anae. This summary emphasizes a few groupings of particular interest, which are discussed further below.

The deepest branch within the monocots is between a clade that includes *Acorus* as sister of a monophyletic Alismatanae (including Alismatales and Najadales, as represented by *Sagittaria*, *Butomus*, *Triglochin*, and *Potamogeton*), and a second clade that includes all other monocots. Within the latter group, a monophyletic grouping of three genera of Araceae is sister of all remaining monocots. Dahlgren *et al.* (1985) included *Acorus* within Araceae, and the present analysis, in placing this genus with Alismatanae, suggests that it must be removed from Araceae, Arales, and Aranae if those groups are to be monophyletic. Other analyses have placed *Acorus* as sister of all other monocots (e.g. Duvall *et al.* 1993a, b; Chase *et al.* 1993, 1995), and that position also is inconsistent with the placement of *Acorus* in Araceae, Arales, or Aranae.

Among the remaining monocots there is a pectinate arrangement of three major clades that together include all sampled taxa of Lilianae, Cyclanthanae, and Pandananae, plus *Vellozia*, which was included in Bromelianae by Dahlgren *et al.* (1985). Lilianae, as circumscribed by Dahlgren *et al.*, includes more than 50 families, and the present sampling is insufficient to establish details of relationships in this group. However, it is useful to mention a few key points. The position among the pectinate, non-monophyletic Lilioids (i.e., the Lilianae of Dahlgren *et al.*) of a clade that includes Cyclanthanae, Pandanae, and Velloziales has been observed previously (e.g. Duvall *et al.* 1993b; Chase *et al.* 1993; Davis *et al.* 1995). It now seems well established that Velloziales should be removed from Bromelianae, and that little information is conveyed by the continued placement of Cyclanthaceae and Pandanaceae in separate superorders or orders. The APG (1998) placed Cyclanthaceae, Pandanaceae, Velloziaceae, and Stemonaceae (the latter family not sampled in the present analysis) in a single order, Pandanales, and we concur with their general decision to group these taxa.

Monophyly of Asparagales, Dioscoreales, and Liliales, as circumscribed by Dahlgren *et al.* (1985), is challenged by the relationships resolved here. Although the two sampled representatives of Melanthiales in the present study are resolved as a monophyletic group, other analyses (e.g. Chase *et al.* 1993) have resolved several other genera from this order as distantly related. Thus, we concur with the APG (1998) that three or more major monophyletic groups of families among the Lilioids should be recognized, and that precise relationships among these groups remain to be worked out.

In the present analysis, all remaining monocots are resolved as a large clade that includes all members of Arecanae, Bromelianae (except the aforementioned Velloziales), Commelinanae, and Zingiberanae, as circumscribed by Dahlgren *et al.* (1985). This clade, which we have previously designated the 'ABCZ' clade, corresponds in most respects with the 'commelinoid' group recognized by the APG (1998). Within the ABCZ clade, Arecanae are sister of a group that includes all of the other taxa, and the latter fall into two major clades. The Zingiberanae are resolved as a monophyletic group within one of these clades, but both of the clades include members of Bromelianae and Commelinanae. Thus, any taxonomic structure for the ABCZ clade that is consistent with the present results, and with the demand for monophyletic taxa, would involve major realignments relative to the system of Dahlgren *et al.* (1985). One now-familiar (and nomenclaturally bothersome) aspect of the relationships resolved in this group is the resolution of a clade that includes most members of Commelinanae (*sensu* Dahlgren *et al.* 1985), but that excludes Commelinaceae itself (and one or more other families). Because this includes most members of the familiar Commelinanae, but not Commelinaceae, it must be either accommodated within another taxon or assigned another name. Some of the necessary realignments within the ABCZ clade are accounted for in the system proposed by the APG (1998), but those authors also left six families unaligned (including Bromeliaceae, Mayacaceae, and Rapateaceae). Within the ABCZ clade, the present results place *Trithuria* with Xyridaceae, Cyperaceae, and Eriocaulaceae, thus combining representatives of three different orders of Commelinanae (Commelinales, Cyperales, and Hydatellales) as circumscribed by Dahlgren *et al.* (1985). In the present analysis, a sister-group relationship between this clade and one that corresponds to the Poales of Dahlgren *et al.* (1985) is resolved. Members of both of these groups were included by the APG (1998) within a much more broadly circumscribed Poales that accommodates many of the families that Dahlgren *et al.* (1985) had included in Commelinales, and that, in light of the placement of Commelinaceae, could not be included in an order with that name unless it was so broadly circumscribed that it included all families of the ABCZ clade except Arecaceae.

MORPHOLOGICAL CHARACTERS AND SPECIFIC CLADE SUPPORT

The following discussion of morphological characters is limited to a few characters supporting particular nodes, a discussion of the placement of *Trithuria*, a member of the Hydatellaceae which is a family of uncertain affinity (Dahlgren *et al.* 1985; Hamann 1998), and the placement of *Acorus*.

The Position of *Acorus*

Acorus has been variously placed in cladistic analyses depending upon the type of data. In an analysis of the monocots based upon morphology alone (Stevenson and Loconte 1995), *Acorus* was placed with the Typhales along with the Hydatellaceae. In striking contrast, analyses of only *rbc*L nucleotide sequence data and of *rbc*L data plus morphological data (Chase *et al.* 1995) place *Acorus* as the sister taxon to all other monocots. However, both of these results are peculiar in a morphological context. The placement of *Acorus* as the sister group to the rest of the monocotyledons in the combined analysis of Chase *et al.* (1995) resulted in only one morphological synapomorphy for the rest of the monocotyledons, namely the presence of monocotyledonous anther wall formation as opposed to dicotyledonous anther wall formation reported for *Acorus*. However, the report in the literature for the dicotyledonous anther wall formation in *Acorus* can be neither documented nor confirmed (see Rudall and Furness 1997 for a thorough discussion). Thus, from recent work (Rudall and Furness 1997) one must conclude that there are no known morphological synapomorphies that support the sister group relationship of *Acorus* with the rest of the monocotyledons.

Some of the synapomorphies that group *Acorus* with Hydatellaceae plus Typhaceae in the Stevenson and Loconte (1995) and Chase *et al.* (1995) analyses of morphology alone are easily understood as parallelisms, such as apical placentation. One other character that supported that topology was the presence of perisperm as a storage tissue in seeds. However, more recent investigations on perisperm by Rudall (1997) and Rudall and Furness (1997) have revealed that the developmental pathway of perisperm in *Acorus* is different from that of the Typhaceae or Hydatellaceae. Thus, the perisperm of *Acorus* apparently is not homologous with the perisperm found in other monocots but rather is autapomorphic for *Acorus*. There are two morphological synapomorphies, the presence of collar rhizoids (Tillich 1998) and the presence of a 3-bp deletion in *atp*A, that support *Acorus* as the sister taxon to the Alismatids in this analysis. The 3-bp deletion occurs in *Acorus*, *Sagittaria*, *Butomus*, *Triglochin*, and *Potamogeton*. It also occurs in a few taxa within the ABCZ clade. The 6-bp deletion occurs in two places in the tree, as a synapomorphy of Zingiberanae (i.e., in *Canna* and *Alpinia*), and as a synapomorphy of Pandananae and Cyclanthanae (i.e., in all five taxa sampled from these two superorders, though not in the sister of this group, *Vellozia*). Thus, outside the ABCZ clade, the 3-bp deletion occurs only as an unreversed synapomorphy of *Acorus* + Alismatanae, and the 6-bp deletion occurs only as an unreversed synapomorphy of Pandananae + Cyclanthanae, so neither appears to be a transformed state of the other. Therefore, unless *Acorus* is placed within the ABCZ clade, the 3-bp deletion unambiguously favors its placement within a clade that also includes all Alismatanae. We conclude that the problem of the various placements of *Acorus* in prior analyses stems from lack of data, poor character coding, and the autapomorphic nature of *Acorus* in both morphological and molecular features (e.g. Davis *et al.* 1998). Areas that are in particular need of investigation on broad comparative bases for *Acorus*, Alismatids, and Aroids are inflorescence development, inflorescence topology, root anatomy, and vegetative branching.

The Position of *Trithuria* (Hydatellaceae)

Another problematic taxon is the Hydatellaceae (Hamann 1998). In a morphological analysis of monocot families (Stevenson and Loconte 1995), Hydatellaceae was placed as the sister family to a clade composed of Typhaceae and Sparganiaceae with *Acorus* sister to all of these (*Acorus* (Hydatellaceae (Typhaceae Sparganiaceae))). Because molecular data were not available for Hydatellaceae, Chase *et al.* (1995) did not include it in their *rbc*L analysis but did include it in the combined *rbc*L and morphology analysis. Not surprisingly, the position of Hydatellaceae in that analysis was the

same as in the analysis of morphology alone. Three morphological synapomorphies supported the sister group relationship of Hydatellaceae and Typhales in these previous analyses. These are porate pollen, diclinous flowers, and a pseudomonomerous gynoecium. However, the pollen of Hydatellaceae is polymorphic, with both sulcate and porate types occurring, and with the latter known only for *Hydatella inconspicua*. The synapomorphy of porate pollen in this instance is based upon an optimization of a polymorphic coding for the entire family. We discuss the general problems of polymorphism coding, and the advantages of exemplar sampling, in a section below.

Diclinous flowers and pseudomonomery are both common within the monocots as a whole, and are highly homoplasious (Stevenson and Loconte 1995) because of our failure to really understand how to code them. These characters are among many that, to be coded properly, need to be continually reconsidered as is facilitated by the reciprocal illumination made possible from cladistic analysis. In this fashion, developmental bases for similarities and differences in final morphologies may be discovered and coded as disussed above for the 'perisperm' of *Acorus*. Because of the anatomical and morphological simplicity of the Hydatellaceae, presumably because of the highly reduced nature of these small aquatics, the species of Hydatellaceae are often missing codable states (Stevenson and Loconte 1995). The addition of molecular data in this analysis has resulted in support for a sister group relationship between *Trithuria* (Hydatellaceae) and *Xyris* (Xyridaceae). It should be noted that this relationship could not have been found in the combined analysis of Chase *et al.* (1995) because molecular data was lacking for both Hydatellaceae and Xyridaceae. The hypothesis of a sister group relationship between *Trithuria* and *Xyris* (Fig. 1) is supported by two morphological synapomorphies, both of which have a CI of 1.0. These characters are the presence of latrorse anther dehiscence and the presence of an embryostega (stopper or seed operculum) that is derived from radial growth of the inner integument. This type of embryostega also occurs in *Hydatella* (Hamann 1998), the only other genus of the Hydatellaceae. It is interesting to note that the only other case of an inner integumentary embryostega in monocots is in *Mayaca* (Venturelli and Bouman 1986), and that this character served as a synapomorphy for *Mayaca* and Xyridaceae in a previous analysis (Stevenson and Loconte 1995). In contrast, the embryostega of the Commelinaceae is derived from the outer integument (Grootjen 1983; Grootjen and Bouman 1981).

Cyclanthanae, Pandananae, and Velloziales

As mentioned earlier, a clade that includes Cyclanthanae, Pandananae, and Velloziales has been observed previously (e.g. Duvall *et al.* 1993b; Chase *et al.* 1993, 1995; Davis 1995). The sister group relationship of Pandanaceae and Cyclanthaceae is strongly supported by these analyses, a morphological analysis with seven unambiguous synapomorphies for this group (Stevenson and Loconte 1995), and a combined analysis (Chase *et al.* 1995). The relationship of the Velloziaceae as sister to Pandanaceae and Cyclanthaceae was not found in previous analyses of morphological data alone (Stevenson and Loconte 1995; Chase *et al.* 1995) but was present in a combined analysis (Chase *et al.* 1995) as it is here. There are two morphological synapomorphies of this clade. They are the presence of basally connate stamens and the presence of starchy endosperm.

Basally connate stamens are a rather common feature within the monocots. However, an interesing aspect of this with respect to the Pandanales (*sensu* APG 1998) is that the connate stamens form a stemonophore structure at the common base. Thus, the character is not so much basal connation of the stamens but rather basal connation into a stemonophore instead of a tube. This redefinition, of course, serves to remove an instance of presumed homoplasy within the monocots.

Starchy endosperm is a feature of the Bromelianae, Commelinanae, and Zingiberanae (the BCZ subclade of the aforementioned ABCZ clade). This represents, in part, the historical group known as the Farinosae (Engler 1892), which did not include the Pandanales or Cyclanthales. Therefore, the presence of starchy endosperm is a homoplasy that serves as a local (parallel or secondary) synapomorphy for two different major groups. In turn, the structure of the cladogram suggests that starchy endosperm is not homologous across all occurrences; and, thus, that it has developed at least twice from a non-starchy condition within the monocots. This would further suggest that different biochemical pathways may be involved, and perhaps a different final chemical composition, which may be reflected in the different forms of starch grains that are known within the monocots. However, there is a paucity of information concerning the chemical and microscopic nature of this starch. It can come in many forms, for example as single grains or compound grains, and there is also variation in shape and size (Dahlgren and Clifford 1982). Unfortunately, very little is known about the diversity and biosynthetic pathways of starch outside a few economically important plants (Badenhuizen 1959; Radley 1968). However, examination of the tree suggests that there are differences that have yet to be discerned. This is but another case where lack of understanding of characters and character states can lead to inappropriate and/or mistaken codings.

Exemplar Sampling and Polymorphism Coding

In the discussion of *Trithuria*, above, we noted an instance in which a polymorphism can influence the structure of a tree. This raises the general problem in phylogenetic analysis of how to sample diverse multi-species taxa for phylogenetic analyses of large groups such as the monocots. Specifically, if all taxa within such a group cannot be sampled (e.g. all members of groups such as Orchidaceae, Poaceae, Arecaceae, or Araceae in a study of monocot relationships), there exists a range of possible ways to define terminals for analysis. At one extreme, a number of individual species are sampled, and relationships among the larger groups are inferred from relationships among the exemplar species; this is exemplar sampling. At the other extreme, highly inclusive terminals are used, each of which is intended to combine observations from more than one species, and thus to represent a broad range of diversity. Between these extremes lies a series of intermediate approaches. Studies based on DNA sequence variation, particularly those that utilize a single gene, involve an exemplar approach, though with the increasing use of multiple genes, sequences from different taxa often are combined. For example, the terminal labelled '*Trillium*' in the

present study includes an *rbc*L sequence from one species of this genus and an *atp*A sequence from another. Studies based on morphology frequently involve complex multi-taxon terminals. For example, the analysis by Stevenson and Loconte (1995) included 103 monocot terminals, most of which represented variation within whole families.

One problem that immediately arises, when observations from two or more species are aggregated into a single terminal for analysis, is that of how character variation within the terminal should be expressed. One approach is to score that terminal as polymorphic (two or more states present) for all characters that are known to vary within it. Thus, a terminal that is meant to accommodate variation in Arecaceae might be scored as having ovules basal *and* axile, since both states occur within the family. Consolidation of this sort has been advocated by various authors (e.g. Donoghue 1994), with a variety of rationalizations, including reduction in the number of terminals in an analysis and the desire to include as much information as possible. However, one obvious problem with such an approach is that it represents an assumption that the constituent taxa within each of the aggregate terminals in an analysis do belong together. Stevenson and Loconte (1995), realizing that *Acorus* is not likely to be nested within a monophyletic Araceae (Grayum 1987), included *Acorus* in their study as a separate terminal; however, if there are other taxa that are conventionally assigned to Araceae, and that do not actually belong within that group, but are represented in the character scoring for a terminal that is intended to represent the family, the analysis could be affected in unpredictable ways. Recent studies have demonstrated the polyphyly of traditionally accepted groups such as Melanthiaceae (e.g. Chase *et al.* 1995), and it is difficult to know which other currently accepted taxa also are unnatural assemblages.

Another problem that arises when multi-taxon assemblages are combined within a single terminal is that spurious relationships may be supported even if the taxa that are combined within the artificial terminal constitute a monophyletic group (Nixon and Davis 1991). If an artificial terminal is polymorphic for two binary characters (i.e., states 0 and 1 occur in the group for both characters), there are four possible character combinations, yet fewer than all four of these combinations may actually occur among the constituent taxa. In other words, superior and inferior ovaries may occur among the species of a monophyletic family, and opposite and alternate leaves also may both occur, but there may be no species that has inferior ovaries and alternate leaves. When a terminal is coded so as to represent these polymorphisms, and it is included in a cladistic analysis, all possible combinations of states among polymorphic characters are subjected to the parsimony criterion, and the most parsimonious trees may imply combinations that do not exist in any species, and that may never have existed in the history of that group. Each polymorphism in a terminal multiplies the number of possible character combinations, so a moderate number of such characters may allow an enormous number of combinations. For example, it is unlikely that there has ever been an angiosperm with monocotyledonous embryos, opposite leaves, monosulcate pollen, poricidal anther dehiscence and betalain pigments, yet such a combination would be allowed by a phylogenetic analysis in which a single terminal stood for the angiosperms and was scored as polymorphic for all characters that vary. On the basis of these considerations, Nixon and Davis (1991) argued that polymorphism coding to accommodate variation among constituent taxa of a presumptive higher-level taxon should be regarded as an expedient, and that 'taxa' of this sort should be broken up as much as possible so that the only character combinations present in a matrix are those that have been observed to occur in nature. This, in fact, is our goal in shifting from the use of families as terminals (Stevenson and Loconte 1995) towards an exemplar approach, as in the present study. Known instances of character polymorphism within families have guided our sampling, and the recognized families with greater numbers of such polymorphisms will continue to be sampled more deeply than those in which polymorphisms are fewer.

In contrast with these views, Doyle and Donoghue (1992) and Donoghue (1994) have asserted that the separation of taxa that previously had been grouped within a single terminal, in order to remove polymorphic codings, involves an assumption of monophyly of each of the resulting subgroups. These objections have been discussed by Nixon (1996), who has pointed out that splitting polymorphic terminals reduces the number of *a priori* assumptions, and thus allows the cladistic analysis to test the monophyly of groups, while the consolidation of separate terminals into a single polymorphic taxon prevents the investigator from discovering that the group may be, in fact, an unnatural assemblage. Furthermore, the problems associated with the introduction of unobserved character combinations into a matrix, as discussed above, have not been addressed by those who advocate the extensive use of polymorphism coding in favor of exemplar sampling.

While morphological characters and character states may be difficult to assess and code *a priori*, they do lend them themselves to testing via cladistic analysis and reciprocal illumination, i.e., *a posteriori*. Non-equivalency in coding can be demonstrated with developmental studies. What is initially viewed as homoplasy in a character state that occurs, for example, twice on a tree may, in fact, be precisely what the tree indicates, which is that this is not the same character state but rather two different things. Once this is incorporated, we may then have two different characters and/or character states with each being a synapomorphy with no homoplasy. We will then have learned much more about the plants under study. For now, sequence data does not lend itself to this approach.

ACKNOWLEDGEMENTS

The authors thank Rob DeSalle of the American Museum of Natural History for access to and use of sequencing facilities, and Victor Albert, Mark Chase, Timothy Evans, and Donald Les for providing unpublished *rbc*L sequences. We also wish to thank the organizers of Monocots II for a superb conference and their extraordinary patience.

REFERENCES

APG [Angiosperm Phylogeny Group] (1998). An ordinal classification for the families of flowering plants. *Annals of the Missouri Botanical Garden* **85**, 531–553.

Badenhuizen, N. P. (1959). 'Chemistry and Biology of the Starch Granule.' (Springer-Verlag: Wien.)

Chase, M. W., Soltis, D. E., Olmstead, R. G., Morgan,D., Les, D. H., Mishler, B. D., Duvall, M. R., Price, R. A., Hills, H. G., Qiu, Y.-L., Kron, K. A., Rettig, J. H., Conti, E. , Palmer, J. D., Manhart, J.R., Sytsma, K. J., Michaels, H. J., Kress, W. J., Karol, K. J., Clark, W. D., Hedrén, M., Gaut, B. S., Jansen, R. K., Kim, K.-J., Wimpee, C. F., Smith, J. F., Furnier, G. R., Strauss, S. H., Xiang, Q.-Y., Plunkett, G. M., Soltis, P. S., Swensen, S. M., Williams, S. E., Gadek, P. A., Quinn, C. J., Eguiarte, L. E., Golenberg, E., Learn, G. H. Jr., Barrett, S. C. H., Dayanandan, S., and Albert, V. A. (1993). Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbc*L. *Annals of the Missouri Botanical Garden* **80**, 528–580.

Chase, M., Stevenson, D., Wilkin, P., and Rudall, P. (1995). Monocot systematics: a combined analysis. In 'Monocotyledons: Systematics and Evolution'. (Eds P. Rudall, P. Cribb, D. Cutler, and C. Humphries.) pp. 685–730. (Royal Botanic Gardens: Kew.)

Dahlgren, R. M. T., and Clifford, H. T. (1982). 'The Monocotyledons: A Comparative Study.' (Academic Press: London.)

Dahlgren, R. M. T, Clifford, H. T., and Yeo, P. F. (1985). 'The Families of Monocotyledons: Structure, Evolution and Taxonomy.' (Springer-Verlag: New York.)

Davis, J. I. (1995). A phylogenetic structure of the monocotyledons, as inferred from chloroplast DNA restriction site variation, and a comparison of measures of clade support. *Systematic Botany* **20**, 503–527.

Davis, J. I., Simmons, M. P., Stevenson, D. W., and Wendel, J. F. (1998). Data decisiveness, data quality, and incongruence in phylogenetic analysis: an example from the monocotyledons using mitochondrial *atp*A sequences. *Systematic Biology* **47**, 282–310.

Donoghue, M.J. (1994). Progress and prospects in reconstructing plant phylogeny. *Annals of the Missouri Botanical Garden* **81**, 405–418.

Doyle, J. A., and Donoghue, M. J. (1992). Fossils and seed plants phylogeny reanalysed. *Brittonia* **44**, 89–106.

Downie, S. R., Katz-Downie, D. S., Wolfe, K. H., Calie, P. J., and Palmer, J. D. (1994). Structure and evolution of the largest chloroplast gene (*ORF2280*): Internal plasticity and multiple gene loss during angiosperm evolution. *Current Genetics* **25**, 367–378.

Duvall, M. R., Learn, G. H. Jr., Eguiarte, L. E., and Clegg, M. T. (1993a). Phylogenetic analysis of *rbc*L sequences identifies *Acorus calamus* as the primal extant monocotyledon. *Proceedings of the National Academy of Sciences, USA* **90**, 4641–4644.

Duvall, M. R., Clegg, M. T., Chase, M. W., Clark, W. D., Kress, W. J., Hills, H. G., Eguiarte, L. E., Smith, J. F., Gaut, B. S., Zimmer, E. A., and Learn, G. H. Jr. (1993b). Phylogenetic hypotheses for the monocotyledons constructed from *rbc*L sequence data. *Annals of the Missouri Botanical Garden* **80**, 607–619.

Engler, A. (1892). 'Die Systematische Anordnung der Monokotyledoneen Angiospermen.' (Riemer: Berlin.)

Farris, J. S. (1989). The retention index and the rescaled consistency index. *Cladistics* **5**, 417–419.

Goloboff, P. A. (1991). Homoplasy and the choice among cladograms. *Cladistics* **7**, 215–232.

Goloboff, P. A. (1993). 'NONA, version 1.5.1.' Distributed by the author. (Tucuman: Argentina.)

Grayum, M.H. (1987). A summary of evidence and arguments supporting the removal of *Acorus* from the Araceae. *Taxon* **36**, 723-729.

Grootjen, C. J. (1983). Development of the ovule and seed in *Cartonema spicata* (Cartonemataceae). *Australian Journal of Botany* **31**, 297–305.

Grootjen, C. J., and Bouman, F. (1981). Development of the ovule and seed in *Stanfieldiella imperforata* (Commelinaceae). *Acta Botanica Neerlandica* **30**, 265–275.

Hamann, U. (1998). Hydatellaceae. In 'The Families and Genera of Vascular Plants, Vol. 4.' (Ed. K. Kubitzki). pp. 231–234. (Springer-Verlag: New York.)

Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C.-R., Meng, B.-Y., Li, Y.-Q., Kanno, A., Nishizawa,Y., Hirai, H., Shinozaki, K., and Sugiura, M. (1989). The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Molecular Genetics and Genetics* **217**, 185–194.

Kadowaki, K. I., Kazama, S., and Suzuki, T. (1990). Nucleotide sequence of the F-1 Atpase aα-subunit gene from rice mitochondria. *Nucleic Acids Research* **18**, 1302.

Kluge, A. G., and Farris, J. S. (1969). Quantitative phyletics and the evolution of anurans. *Systematic Zoology* **18**, 1–32.

Nixon, K. C. (1996). Paleobotany in cladistics and cladistics in paleobotany: enlightenment and uncertainty. *Review of Paleobotany and Palynology* **90**, 361–373.

Nixon, K. C. (1998a). 'Dada Version 2.0001.' Distributed by the author. (Ithaca: New York.)

Nixon, K. C. (1998b). 'Clados Version 1.40009.' Distributed by the author. (Ithaca: New York.)

Nixon, K. C., and Carpenter, J. M. (1994). On outgroups. *Cladistics* **9**, 413–426.

Nixon, K. C., and Davis, J. I. (1991). Polymorphic taxa, missing values and cladistic analysis. *Cladistics* **7**, 233–241.

Radley, J. A. (1968). 'Starch and its Derivatives.' 4th ed. (Chapman and Hall: London.)

Rudall, P.J. (1997). The nucellus and chalaza in monocotyledons. *Botanical Review* **63**, 140–181.

Rudall, P.J., and Caddick, L.R. (1994). Investigation of the presence of phenolic compounds in monocot cell walls, using UV fluorescence microscopy. *Annals of Botany* **75**, 483–491.

Rudall, P.J., and Furness, C. (1997). Systematics of *Acorus*: ovule and anther. *International Journal of Plant Sciences* **158**, 640–651.

Rudall, P. J., Stevenson, D. W., and Linder, H. P. (1999). Structure and systematics of *Hanguana*, a monocotyledon of uncertain affinity. *Australian Systematic Botany* **12**, 311–330.

Shimada, H., and Sugiura, M. (1991). Fine structural features of the chloroplast genome: Comparison of the sequenced chloroplast genomes. *Nucleic Acids Research* **19**, 983–995.

Shinozaki, K., M. Ohme, Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.-Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J. Takaiwa, F., Kato, A., Todoh, N., Shimada, H., and Sugiura, M. (1986). The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO Journal* **5**, 2043–2049.

Stevenson, D. W., and Loconte, H. (1995). Cladistic analysis of monocot families. In 'Monocotyledons: Systematics and Evolution'. (Eds P. Rudall, P. Cribb, D. Cutler, and C. Humphries.) pp. 543–578. (Royal Botanic Gardens: Kew.)

Swofford, D. L. (1993). 'PAUP: Phylogenetic Analysis using Parsimony, version 3.1.' (Illinois Natural History Survey: Champaign.)

Tillich H. J. (1998). Development and organization. In 'The Families and Genera of Vascular Plants, Vol. 3.' (Ed. K. Kubitzki). pp. 1–19. (Springer-Verlag: New York.)

Venturelli, M., and Bouman, F. (1986). Embryology and seed development in *Mayaca fluviatilis* (Mayacaceae). *Acta Botanica Neerlandica* **35**, 497–516.