1

A well-supported phylogenetic framework for the monocot order Alismatales reveals multiple losses of the plastid NADH dehydrogenase complex and a strong long-branch effect

William J. D. Iles, Selena Y. Smith and Sean W. Graham

1.1 Introduction

The order Alismatales is a cosmopolitan and enormously diverse clade of monocotyledons, comprising ~4500 extant species in 13 families, as currently defined (Stevens, 2001+; Janssen and Bremer, 2004; APG III, 2009). Some of the oldest monocot fossils (late Barremian and early Albian; 125–112 Ma) have been assigned to this lineage (Friis et al., 2004, 2010), and most phylogenetic studies (e.g. Chase et al., 2006; Givnish et al., 2006, 2010; Graham et al., 2006) resolve Alismatales as the sister group of all monocots except *Acorus* (Acorales: Acoraceae). Refining our understanding of the phylogenetic backbone of Alismatales will therefore be important for understanding the early evolutionary history of the monocots.

The overall composition of Alismatales remained relatively constant until a recent expansion to include Araceae and Tofieldiaceae (e.g. Dahlgren and Clifford, 1982; Tomlinson, 1982; Les et al., 1997; APG I, 1998; APG II, 2003; APG III, 2009;

Early Events in Monocot Evolution, eds P. Wilkin and S. J. Mayo. Published by Cambridge University Press. © The Systematics Association 2013.

Chase, 2004). This shift reflects substantial molecular systematic evidence (e.g. Duvall et al., 1993; Chase et al., 1995, 2000, 2006; Tamura et al., 2004a; Givnish et al., 2006; Graham et al., 2006) for a close relationship between Araceae, Tofieldiaceae and a clade of 'core alismatid' families that corresponds approximately to the order Helobiae (Engler, 1892) and subclass Alismatidae (Cronquist, 1988). Les and Tippery (Chapter 6, this volume) favour a narrower definition of the clade (as Alismatidae, with two orders, and excluding Araceae and Tofieldiaceae), but we find the broader circumscription of the order more appealing, because it underlines the evolutionary links among these diverse lineages. *Acorus* has also sometimes been recovered within Alismatales (e.g. Davis et al., 2004, 2006), but this placement may reflect substantial rate elevation in several mitochondrial genes (Petersen et al., 2006a, 2006b; Mower et al. 2007; Cuenca et al., 2010). There have been multiple morphological and molecular phylogenetic studies of individual families and major genera of Alismatales (e.g. Les et al., 1993, 1997, 2002a, 2002b, 2005, 2006, 2008, 2010; Tanaka et al., 1997, 2003; Waycott et al., 2002, 2006; Kato et al., 2003; Iida et al., 2004; Rothwell et al., 2004; Tamura et al., 2004b, 2010; Keener, 2005; Lehtonen, 2006, 2009; Lindqvist et al., 2006; Jacobson and Hedrén, 2007; Wang et al., 2007; Cabrera et al., 2008; Lehtonen and Myllys, 2008; Zhang et al., 2008; Ito et al., 2010; von Mering and Kadereit, 2010; Azuma and Tobe, 2011; Cusimano et al., 2011). However, only a few studies (e.g. Les et al., 1997) have surveyed the broad phylogenetic backbone of the order.

Les et al. (1997) provided the most comprehensive study of higher-order relationships in Alismatales. They sampled the plastid gene *rbc*L for exemplar taxa representing all families except Tofieldiaceae, and most of the genera except in Araceae. In addition to improving our knowledge of phylogenetic relationships in the order, and refining family-level circumscriptions, they were interested in reconstructing the evolution of characters that may be associated with hydrophilous (water-mediated) pollination. The core alismatid families are mostly fully aquatic (Les et al., 1997), and semi- to fully aquatic plants are also found in Araceae and Tofieldiaceae, consistent with an aquatic or semi-aquatic habit for the most recent common ancestor of the monocots (e.g. Chase, 2004; note that *Acorus* is also semi-aquatic). If so, terrestrial species in the order (i.e. most Araceae, some Tofieldiaceae) would therefore represent subsequent reversions in habit. The order encompasses all major aquatic life forms (i.e. emergent, floating-leaved, free-floating and submersed; Sculthorpe, 1967), and includes the only fully marine angiosperms, the seagrasses, a life form that evolved several times in the order (Les et al., 1997). Morphological features linked to hydrophily and an aquatic habit are expected to have an unusually high level of homoplasy, which may have contributed to the fluidity of earlier family-level classification schemes based on morphology (see Les and Haynes, 1995, Les et al., 1997).

Les et al. (1997) reconstructed the overall phylogenetic backbone of the order using a single plastid gene, and recovered multiple poorly to moderately supported branches underpinning the higher-order relationships. The monophyly and extent of several families were also unclear (this latter uncertainty was partly accommodated in the APG classification systems by the expanded circumscription of several families).

A few studies have revisited their *rbc*L data set, either alone or in combination with morphology (Chen et al., 2004a, 2004b; Li and Zhou, 2009), but no subsequent studies have sampled the order broadly using additional genes, with the exception of a suite of papers focussed primarily on mitochondrial gene evolution (Petersen et al., 2006b; Cuenca et al., 2010). Here we substantially expand the number of plastid genes sampled from exemplar species that represent the broad phylogenetic backbone of the order. Our major goal is to re-examine and further refine the overall backbone of Alismatales phylogeny recovered by Les et al. (1997) by considering more plastid data per taxon. This general approach has proved to be effective for the inference of broad-scale monocot phylogeny (e.g. Graham et al., 2006; Saarela et al., 2008; Givnish et al., 2010; Saarela and Graham, 2010). We confirm much of the broad phylogenetic backbone recovered by Les et al., (1997), with some notable exceptions. We also obtain substantially improved branch support in many cases. However, we demonstrate that too limited taxon sampling can lead to spurious inference of some local relationships when using plastid genes, which may be a consequence of elevated rates of evolution in a subset of regions examined. Finally, we document and characterize multiple independent losses of plastid genes that code for two subunits of the plastid NADH dehydrogenase chlororespiratory complex.

1.2 Materials and methods

1.2.1 Taxon sampling

Our main analyses focus on a set of 92 exemplar (representative) species comprising 31 species from Alismatales, 49 other monocots and 12 other angiosperms. We expanded taxon sampling in Alismatales by 26 species compared to our most recent broad study of monocot phylogeny (Saarela and Graham, 2010), and included all currently recognized families in the order (Appendix). Our overall taxon sampling for Alismatales is generally less dense than Les et al. (1997), but the included lineages constitute a highly representative subsample of the broad backbone of Alismatales phylogeny. As far as possible we included multiple representatives per family and targeted species within families that span their deepest phylogenetic splits, at least as defined in Les et al. (1997). We included the south-eastern Australian endemic *Maundia triglochinoides* because of a recent report that it lies outside Juncaginaceae (von Mering and Kadereit, 2010), rendering that family paraphyletic as currently circumscribed (Les et al., 1997). Our most complete generic sampling in the order is in Tofieldiaceae, with four of its five genera included (only *Isidrogalvia* is not sampled).

Outside Alismatales we excluded some taxa that were included previously (Graham and Olmstead, 2000; Graham et al., 2006; Saarela et al., 2007, 2008; Saarela and Graham, 2010) to facilitate maximum likelihood analysis, but our taxon sampling is broadly representative of Petrosaviidae (Cantino et al., 2007; this name was coined for the large clade that encompasses all monocots except *Acorus* and Alismatales). We also included new sequences for exemplar species from each of the following families: Acoraceae (Acorales), Bromeliaceae (Poales), Nartheciaceae (Dioscoreales), Nymphaeaceae (Nymphaeales), Orchidaceae (Asparagales), Philesiaceae and Rhipogonaceae (Liliales); see Appendix for details.

1.2.2 Gene sampling

We extracted total genomic DNAs from silica-gel dried leaf material (Appendix) using standard protocols (Doyle and Doyle, 1987; Graham and Olmstead, 2000), or by using a DNeasy Plant Mini Kit (Qiagen Inc, Valencia, California, USA) for recalcitrant material. Some DNAs were provided by the Royal Botanic Gardens, Kew. In several cases we included sequences from GenBank (*Nuphar*, *Phalaenopsis*; several eudicots) or from other workers (*Vallisneria*; Appendix). In total we sampled 17 plastid genes and associated noncoding regions (omitting several noncoding regions from analysis, see below). These genes are involved in several different plastid functions: photosynthesis (*atp*B, *psb*B, *psb*C, *psb*D, *psb*E, *psb*F, *psb*H, *psb*J, *psb*L, *psb*N, *psb*T, *rbc*L), chlororespiration (*ndh*B, *ndh*F) and protein translation (*rpl*2, *rps*7, 3'-*rps*12). Our sample included the following multigene clusters: *psb*B-*psb*T-*psb*N-*psb*H (which we refer to here as *psb*BTNH), *psb*E-*psb*F-*psb*L-*psb*J (= *psb*EFLJ), and 3'-*rps*12-*rps*7-*ndh*B-*trn*L(CAA). We surveyed these regions using amplification and sequencing protocols noted in Graham and Olmstead (2000) and Saarela et al. (2008), and designed several modified primers for the *psb*BTNH region: modB60F (5'-CATACAGCTTTAGTTGCTGGTTGG), modB64R (5'-GGGATCAGGGATATTTCCAGCAAG), mod65R (5'-GGAAATGTTTCAAAAAAAG-TAGGC A) and modB71R (5'- CCCGGCGCCACTTTACCATATTC).

1.2.3 Data assembly

We carried out base-calling and contig assembly using Sequencher 4.2.2 (Gene Codes Corp., Ann Arbor, Michigan, USA), determining gene boundaries using tobacco and *Ginkgo* reference sequences (Saarela et al., 2008). We added the new sequences to an existing alignment (Saarela et al., 2008), which we adjusted manually using Se-Al 1.0 (Rambaut, 1998) following criteria described in Graham et al. (2000). We coded gaps as missing data. The total aligned length is 23 903 bp, a large portion of which consists of 'offset' noncoding regions that are unique to individual taxa (for a justification of this approach see Saarela and Graham, 2010). For comparison, the unaligned sequence lengths for the newly determined sequences range from 11 009 bp for *Najas* to 15 560 bp for *Stratiotes*. We recovered all 17 gene regions from most species (Appendix). However, for a subset of taxa the *ndh* genes appear to be pseudogenes (i.e. their reading frames are interrupted by stop codons, out-of-phase indels, or both; see below). We recovered a probable *ndh*F pseudogene from *Amphibolis*, and *ndh*B pseudogenes from *Amphibolis*, *Najas*, *Posidonia* and *Thalassia* (partial sequences in several cases, see below). We could not retrieve *ndh*F for *Najas*, *Posidonia* and *Thalassia*, despite extensive attempts at amplification. The apparently pseudogenized *ndh* genes were generally straightforward to align, and so we included them in the analyses. However, a possible *ndh*F pseudogene sequence for *Vallisneria* was so divergent that it could not be aligned reliably, and other *ndh* genes were not recovered for this taxon (M. Moore, Oberlin College, Ohio; pers. comm.).

1.2.4 Phylogenetic analyses

We focussed on coding regions and several conservative noncoding regions from the plastid IR region for the main analysis, following Saarela et al. (2007) and Graham and Iles (2009); the included noncoding regions are intergenic spacers in the contiguous region spanning 3'-*rps*12, *rps*7, *ndh*B and *trn*L, and single introns in each of *rpl*2, 3′-*rps*12 and *ndh*B. We performed heuristic maximum parsimony (MP) searches using PAUP* (Swofford, 2002) with 100 random addition replicates and tree-bisection-reconnection branch swapping, and otherwise using default settings. We used RAxML version 7.2.6 (Stamatakis, 2006; Stamatakis et al., 2008) at the Bioportal website (www.bioportal.uio.no) to perform maximum likelihood (ML) analyses. jModelTest (Posada 2008) was employed to infer the optimal DNA substitution model using the AICc (the Akaike Information Criterion, correcting for sample size) considering the full matrix or subpartitions (see below) for Alismatales only. The GTR+Γ or GTR+Γ+I models were selected in all cases (GTR is the general-time-reversible model, the gamma distribution [Γ] accounts for among-site rate heterogeneity, and the 'I' parameter accommodates invariable sites). Previous analyses of the same gene set across monocots as a whole (e.g. Saarela et al., 2008; Saarela and Graham, 2010) favoured the GTR+Γ+I model. We omitted the I parameter here, as it may be adequately accounted for using the gamma distribution alone (Yang, 2006). We initiated the ML search from 104 random MP starting trees (multiples of eight are required on the Bioportal website), retaining the tree with the highest likelihood score across all searches. We also performed a partitioned ML analysis by distinguishing four partitions, one for each codon position and a separate one for the set of noncoding regions included here, but otherwise using the same settings and general DNA substitution model. We evaluated branch support using the nonparametric bootstrap (Felsenstein, 1985). We considered 500 (MP) or 104 (ML) bootstrap replicates using the search settings described above, but with 10 random addition replicates (MP) or a single random starting tree (ML) per bootstrap replicate. We use the terms 'weak', 'moderate,' and 'strong' to refer to bootstrap support values recovered in the ranges <70%, 70–89%, and ≥90% respectively (Graham et al., 1998).

In earlier unpublished analyses using fewer taxa in Alismatales we noticed that inferred phylogenetic relationships among three families (Alismataceae, Butomaceae and Hydrocharitaceae) depended strongly on the regions and phylogenetic criteria used, and sometimes conflicted with the main results reported here. To explore the possibility that this effect was related to taxon sampling, phylogenetic method or rate heterogeneity in plastid genes, we performed multiple ML and MP analyses for different gene and taxon subsamplings. Specifically, we ran ML and MP analyses on various subsets of the plastid genes, in addition to the full set of regions, using the search settings described above (although in some MP bootstrap analyses we set a MaxTrees limit of 1000 trees). We repeated these analyses for two different taxon densities in Alismatales, a 'reduced' taxon set of 11 exemplar taxa, vs. a 'dense' taxon set comprising all 31 exemplar species. We used two outgroups for these analyses: *Acorus calamus* (Acoraceae) and *Japonolirion osense* (Petrosaviaceae).

1.3 Results

1.3.1 The phylogenetic backbone of Alismatales inferred with 17 plastid genes

Outside Alismatales, the backbone relationships inferred from the full combined data set for 92 taxa are broadly similar to previous estimates using these genes (Fig 1.1, cf. Graham et al., 2006), and so we do not discuss them further here. A portion of the (unpartitioned) ML tree representing Alismatales is presented in Fig 1.2; considering four data partitions in ML analysis did not result in a substantially different topology (not shown: the latter scheme differed in one poorly supported branch inside Araceae). The MP analysis yielded a single most parsimonious tree (tree length = 26 194 steps) that is also highly congruent with the unpartitioned ML tree for Alismatales (not shown). Unpartitioned ML and MP bootstrap values are noted beside individual branches in Fig 1.2; partitioned ML values are indicated in Table 1.1. To facilitate comparisons across analyses and to other studies we have also tabulated support values from the various ML and MP analyses for a subset of branches (Table 1.1; labelled with letters in Fig 1.2); these correspond to interfamilial relationships in the order, in addition to two branches that contradict the monophyly of Juncaginaceae and Cymodoceaceae, respectively. These major backbone relationships in Alismatales are generally strongly supported (ML) or strongly to moderately supported (MP) by the 17-gene data (summarized in the second major column in Table 1.1).

Well-supported clades include Alismatales as a whole (branch a), the core alismatid clade (branch c), a 'petaloid' clade (branch d, comprising three core alismatid families; Les and Tippery, Chapter 6, this volume, refer to this subclade as Alismatales), a 'tepaloid' clade (branch f, comprising the remaining eight core alismatid families; Les and Tippery refer to this subclade as Potamogetonales) and most other branches (branches e and h-m). Core alismatids were distinguished as having either petaloid or tepaloid perianths by Posluszny and Charlton (1993); note that taxa lacking obvious perianths (e.g. *Halodule* and *Najas*) belong to both clades (Fig 1.2). In some cases MP bootstrap support values are marginally (10–20%) weaker than the corresponding ML values (i.e. branch g, which defines the first split in the tepaloid clade above its root node; branch i, which rejects monophyly of Juncaginaceae by placing *Maundia* as the sister group of five families in the tepaloid clade; branch j, for the clade comprising these five families). Two of these three are moderately strongly supported by MP (branches i, j), but all three have strong support (94–100%) from unpartitioned and partitioned ML analyses.

A branch that contradicts the monophyly of Cymodoceaceae here (branch n, which links *Ruppia*, Ruppiaceae, a monogeneric family, with one of the two sampled genera of Cymodoceaceae, *Halodule*; Fig 1.2) is only weakly to moderately supported by all three methods. All other families with multiple exemplar species are strongly supported as monophyletic at the taxon sampling here, and several families with denser sampling also have well-supported internal phylogenetic structure. Specifically, all three internal branches in Tofieldiaceae have strong support, including a placement of *Pleea* as the sister group of the remaining genera, and of *Harperocallis* as the sister group of *Tofieldia-Triantha*; two of four internal branches in Hydrocharitaceae are strongly supported, including a placement of *Stratiotes* as the sister group of other Hydrocharitaceae (Fig 1.2) at the current sparse taxon sampling for this family.

The only major relationship that is not well supported in Alismatales concerns the relative arrangement of its three major subclades: Araceae, Tofieldiaceae and the core clade of alismatid families. Branch b, recovered in the best ML trees here, depicts Araceae as the sister group of the core alismatid families (hence, Tofieldiaceae are the sister group to these two clades, as the order as a whole is also strongly supported). However, this arrangement receives relatively weak support from all three phylogenetic criteria (i.e. 62–66% support, Table 1.1). The two other

Cambridge University Press
978-1-107-01276-9 - The Systematics Association Special Volume 83: Early Events in Monocot Evolution
Edited by Paul Wilkin and Simon J. Mayo
Excerpt
More information

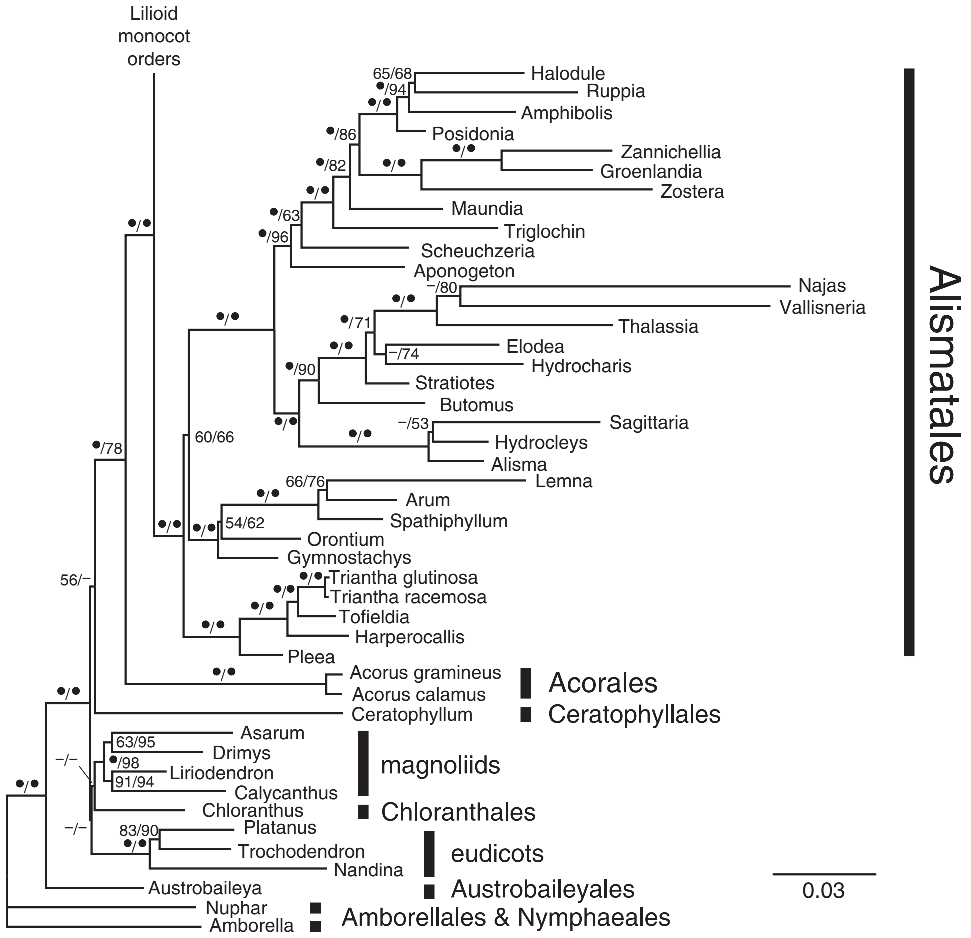

Fig 1.1 Higher-order monocot phylogeny inferred from 17 plastid genes and associated noncoding regions (see text for details). Several outgroup taxa are included. This is the best likelihood tree from an unpartititioned ML analysis (–ln$L$ = 158 904.167). A portion of this tree is shown in magnified form in Fig 1.2. Support values based on bootstrap analyses are noted beside branches (left-hand value = unpartitioned ML, right-hand value = MP); filled circles indicate 100% bootstrap support, dashes <50% bootstrap support. Scale: substitutions per site. An expanded view of Alismatales is presented in Fig 1.2.

possible arrangements for these three clades have been recovered elsewhere with weak to strong support (compare column 2 with columns 4–7 in Table 1.1, which summarize relevant support values in Alismatales across several other studies). One of these alternative possibilities (Araceae sister to Tofieldiaceae; clade b3) has negligible support here, but the other (Tofieldiaceae sister to core alismatids; clade b2) has poor but non-negligible support (i.e. 31–40% of bootstrap replicates from the 17-gene data).

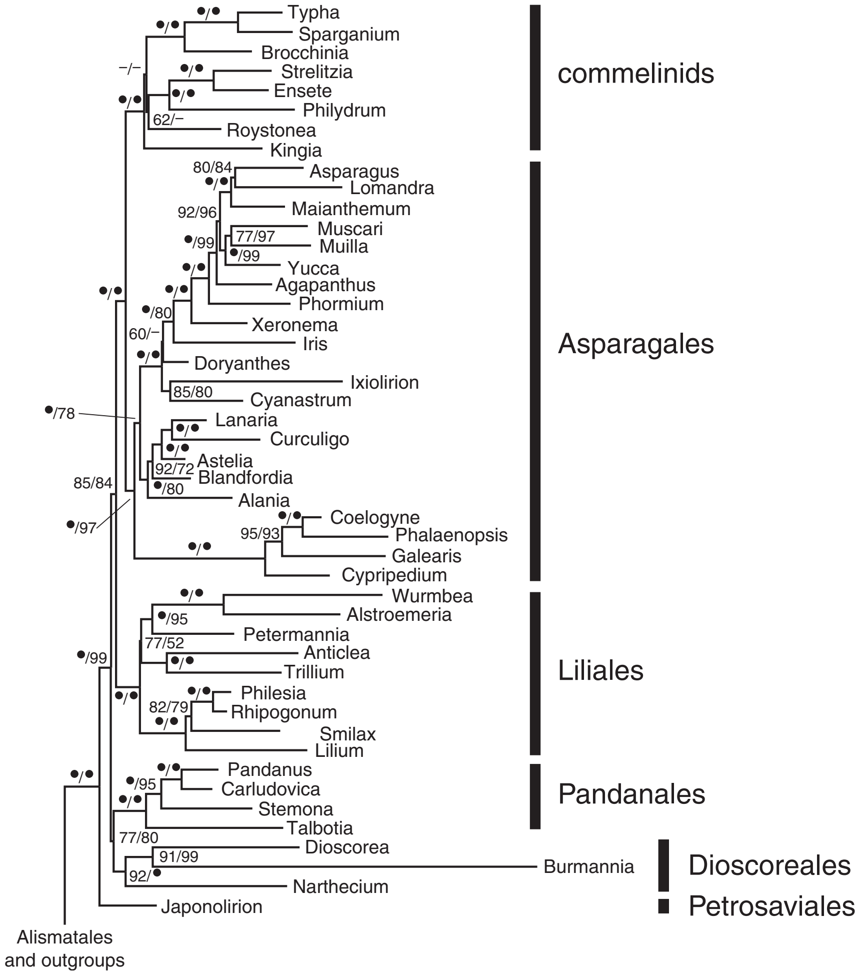

Fig 1.1 (*cont.*)

1.3.2 Effect of taxon density on branch support and conflict

Our reduced taxon set (for the full 17-gene set) has only 11 exemplar taxa from Alismatales, and so several major clades are no longer applicable when compared to the full taxon set (i.e. clades i-n; Fig 1.2 and Table 1.1). We did not run a partitioned ML analysis for this taxon set. The unpartitioned ML and MP bootstrap values are lower for several clades compared to the full taxon sampling (cf. columns 2 and 3 in Table 1.1). These reductions in support are more acute for MP than ML for two clades (i.e. for e and f). However, two clades saw marginal increases with fewer taxa sampled: clade b for ML and MP (this corresponds to Araceae + core alismatids), and clade g for MP only (this is a seven-family clade of core alismatids that corresponds to all of the tepaloid families except Aponogetonaceae; Fig 1.2).

Table 1.1 Comparison of the support values (bootstrap or jackknife) for interfamilial relationships (or for clades i and n, support for the relationships disrupting family monophyly). Author abbreviations: Les et al. (1997; their Fig 2) = L97; Chase et al. (2006; their Fig 2) = C06; Davis et al. (2006; their Fig 2) = D06; von Mering and Kadereit (2010; their Fig 3) = vM&K10. ML-p = partitioned ML; pt = plastid. A dash ('–') means support was not noted or assessed in the corresponding study; a '<' means the branch had <50% support (<70% in von Mering and Kadereit, 2010); 'na' = not applicable due to disrupted monophyly. Clade labels are depicted in Fig 1.2, except for those with a number (b2, b3, etc).

		Full taxon set[b]	**Reduced taxon set[b]**	**L97**	**C06**	**D06**	**vM&K10**
	No. of exemplar species (Alismatales):	31 species	11 species	78 species	13 species	12 species	37 species
	No. of genes:	17 pt genes	17 pt genes	1 gene[c]	7 genes[d]	4 genes[e]	1 gene[c]
	Branch support determined using:	ML, ML-p (MP)	ML (MP)	MP	MP	MP	ML (MP)
Clade label and description[a]							
a	Alismatales	100, 100 (100)	100 (99)	–	100	na	– (–)
b	Araceae + core alismatids	60, 62 (66)	85 (82)	–	–	na	– (–)
b2	Tofieldiaceae + core alismatids	40, 34 (31)	15 (16)	–	99	na	– (–)
b3	Tofieldiaceae + Araceae	<5, <5 (<5)	<5, <5	–	–	na	72 (<)
c	Core alismatid clade	100, 100 (100)	100 (100)	96	100	100	93 (95)
d	Petaloid clade	100, 100 (100)	100 (100)	88	87	<	92 (96)
e	Hydrocharitaceae + Butomaceae	100, 100 (90)	69 (<5)	31	–	<	<(<)
e2	Hydrocharitaceae + Alismataceae	<5, <5 (<5)	31 (99)	<	<	<	– (–)