ROBUST INFERENCE OF MONOCOT DEEP PHYLOGENY USING AN EXPANDED MULTIGENE PLASTID DATA SET

SEAN W. GRAHAM,[1,6] JESSIE M. ZGURSKI,[2] MARC A. MCPHERSON,[2,7] DONNA M. CHERNIAWSKY,[2,8]
JEFFERY M. SAARELA,[1] ELVIRA S. C. HORNE,[2] SELENA Y. SMITH,[2] WINSON A. WONG,[2] HEATH E. O'BRIEN,[2,9]
VINCENT L. BIRON,[2,10] J. CHRIS PIRES,[3] RICHARD G. OLMSTEAD,[4] MARK W. CHASE,[5] AND HARDEEP S. RAI[1]

[1]*UBC Botanical Garden and Centre for Plant Research, University of British Columbia, 6804 SW Marine Drive,
Vancouver, British Columbia V6T 1Z4, Canada;* [2]*Department of Biological Sciences, University of Alberta,
Edmonton, Alberta T6G 2E9, Canada;* [3]*Department of Agronomy, University of Wisconsin,
Madison, Wisconsin 53706, USA;* [4]*Department of Biology, Box 351800, University of Washington,
Seattle, Washington 98195, USA;* [5]*Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AB, UK*
[6]*Corresponding author (swgraham@interchange.ubc.ca)*

ABSTRACT

We use multiple photosynthetic, chlororespiratory, and plastid translation apparatus loci and their associated noncoding regions (ca. 16 kb per taxon, prior to alignment) to make strongly supported inferences of the deep internal branches of monocot phylogeny. Most monocot relationships are robust (an average of ca. 91% bootstrap support per branch examined), including those poorly supported or unresolved in other studies. Our data strongly support a sister-group relationship between Asparagales and the commelinid monocots, the inclusion of the orchids in Asparagales, and the status of Petrosaviaceae as the sister group of all monocots except *Acorus* and Alismatales. The latter finding supports recognition of the order Petrosaviales. Also strongly supported is a placement of *Petermannia* disjunct from Colchicaceae (Liliales) and a sister-group relationship between Commelinales and Zingiberales. We highlight the remaining weak areas of monocot phylogeny, including the positions of Dioscoreales, Liliales, and Pandanales. Despite substantial variation in the overall rate of molecular evolution among lineages, inferred amounts of change among codon-position data partitions are correlated with each other across the monocot tree, consistent with low incongruence between these partitions. *Ceratophyllum* and Chloranthaceae appear to have a destabilizing effect on the position of the monocots among other angiosperms; the issue of monocot placement in broader angiosperm phylogeny remains problematic.

Key words: deep phylogeny, heterotachy, *Japonolirion,* monocot backbone, monocot sister group, monocot tree of life, Petermanniaceae, plastid genome evolution.

INTRODUCTION

Available sources of phylogenetic data do not allow for well-supported inference of all of the deep branches of monocot phylogeny. The internal branches that have proven most resilient to well-supported resolution are often relatively short (e.g., Chase et al. 2000), including portions of the backbone linking those major clades defined as orders in the classification of the Angiosperm Phylogeny Group II [APG II] (2003), and a number of poorly resolved branches deep within each order. To address these and other problematic areas of higher-order monocot phylogeny, we collected new data for a large number of exemplar monocot taxa from an expanded range of regions in the plastid genome, including many that have not been examined intensively in the monocots. The plastid regions surveyed consist of portions of ten photosystem II genes, two NADH dehydrogenase subunit genes, three ribosomal protein genes, *atp*B, *rbc*L, and a di-

verse collection of noncoding regions that span these and other genes from this genome. Collectively, these represent ca. 16 kb of DNA sequence data per taxon, or about a ninth of the nonduplicated information in the plastid genome. Most of the regions were sequenced using primers developed for the inference of deep nodes of angiosperm phylogeny (Graham and Olmstead 2000).

Discussions of strategies for large-scale phylogenetic inference are usually framed in terms of a trade off between taxon sampling vs. amount of data collected per taxon. Aiming for a relatively dense taxon sampling is generally a beneficial strategy for accurate inference of phylogenetic relationships among and within large groups of organisms (e.g., Hillis 1998; Pollock et al. 2002; Zwickl and Hillis 2002), and examining more characters per taxon can also be very useful (e.g., Graham et al. 1998; Soltis et al. 1998; Poe and Swofford 1999; Hillis et al. 2003). Both strategies have demonstrated their effectiveness in ongoing studies of monocot phylogeny (e.g., Chase et al. 1993, 1995*a, b,* 2000, 2005; Stevenson et al. 2000). While it is relatively clear which areas of monocot phylogeny are more (or less) sturdy, the level of data sampling appropriate for tackling the remaining areas of uncertainty is not clear in advance (e.g., Hillis et al. 2003). Fortunately, the economic trade off between taxa and characters is becoming less limiting. With the develop-

Present addresses: [7] Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, Alberta T6G 2P5, Canada; [8] Royal Alberta Museum, 12845-102 Avenue, Edmonton, Alberta T5N 0M6, Canada; [9] Department of Biology, Duke University, Box 90338, Durham, North Carolina 27708, USA; [10] Department of Medical Genetics, University of Alberta, Edmonton, Alberta T6G 2H7, Canada.

ment and application of methods for efficiently examining multiple genomic regions among distantly related taxa, we can collect and analyze substantially more data per taxon (and research dollar) than was possible at the time of the first two monocot conferences.

We therefore sampled a large number of characters for a broad sampling of monocot taxa. Our preliminary monocot study constitutes the largest to date, in terms of the amount of data examined per taxon. It also has a sufficient taxon density to represent all major branches of monocot phylogeny (excluding several achlorophyllous lineages), with several major clades examined at a taxon density comparable to other recent large studies. We report on the strong inferences that can be drawn from the current taxon sampling, highlight several remaining areas of uncertainty, and briefly review extensive among-lineage rate shifts in the context of the molecular evolutionary dynamics of different codon-position data partitions.

MATERIALS AND METHODS

Taxonomic and Genomic Sampling

The taxa surveyed are an expansion of the set examined in Graham et al. (2000) and McPherson et al. (submitted). Our sampling includes representatives of all ten clades recognized as monocot orders and 60 of 93 clades recognized at the rank of family in the classification scheme of APG II (2003), with Petermanniaceae accepted as a distinct family (discussed below). This sampling includes Petrosaviaceae and Dasypogonaceae, two families unplaced to order in the APG II (2003) classification scheme. We use 25 ''dicots'' as outgroup taxa (24 from Graham et al. [2000] and *Trimenia moorei* (Oliv.) W. R. Philipson [Austrobaileyales: Trimeniaceae; NSW 433770, GenBank nos. AY116652–AY116659]). The exemplars were chosen to represent a broad phylogenetic diversity of families, guided by the classification of APG II (2003) and other large-scale studies. Details of the species used, sample provenance, and GenBank accession information are presented in McPherson et al. (submitted) and in the Appendix. Ninety-four exemplar taxa are represented, 69 from the monocots. The latter number is an increase of 23 monocots from McPherson et al. (submitted).

Details of DNA extraction, amplification, and sequencing protocols are provided in Graham and Olmstead (2000) and McPherson et al. (submitted). The data set considered for phylogenetic analysis here spans 17 protein-coding plastid genes (*atp*B, *ndh*B, *ndh*F, *psb*B, *psb*C, *psb*D, *psb*E, *psb*F, *psb*H, *psb*J, *psb*L, *psb*N, *psb*T, *rbc*L, *rpl*2, *rps*7, 3′-*rps*12), and includes six intergenic regions in two of the photosystem II gene clusters (*psb*E–*psb*F–*psb*L–*psb*J and *psb*B–*psb*T–*psb*N–*psb*H), two intergenic regions in a cluster of genes spanning 3′-*rps*12, *rps*7 (two ribosomal small subunit genes) and *ndh*B, and three introns (one each in *ndh*B, *rpl*2 and 3′-*rps*12). Most monocots were sampled for a larger portion of *ndh*F (ca. l2.08 kb, representing most of the locus) than the outgroup taxa (ca. 1.29 kb from the relatively slowly evolving 5′-end of the gene). Three additional regions that are largely noncoding were also sampled for Asparagales and Liliales (and several other taxa; Appendix): an intergenic region between *ndh*B and *trn*L (CAA), an intergenic region

between *atp*B and *rbc*L, and a contiguous region spanning *trn*L (UAA) and *trn*F (GAA) that includes an intron in the former gene and an intergenic spacer region. In a few cases closely related alternative taxa were sampled for the latter region (see Appendix). With these and a few other minor exceptions, all taxa were completely represented for all regions. Further details, including methods of data compilation, are provided in Graham and Olmstead (2000) and McPherson et al. (submitted)

The genes examined include a mix of those involved in photosynthesis (*atp*B, *rbc*L, and the photosystem II genes), chlororespiration (the *ndh* genes) and plastid translation (the *trn, rps,* and *rpl* loci). Those situated in the plastid inverted repeat regions (*rpl*2, *rps*7, 3′-*rps*12, *ndh*B and associated noncoding regions) are exceptionally slowly evolving (e.g., Graham et al. 2000), and the single-copy genes include slowly and rapidly evolving protein-coding and noncoding regions. The *ndh*F locus, for example, includes slowly and rapidly evolving portions (Olmstead and Sweere 1994; Kim and Jansen 1995); the 3′-end of *ndh*F evolves substantially more rapidly (per nucleotide) than some of the single-copy noncoding regions examined here, such as the *trn*L–*trn*F region (Graham and Barrett 2004; Saarela et al. 2005). The characters examined represent most of the spectrum of plastid evolutionary rates.

Alignment and Phylogenetic Analysis

We added taxa to previously published alignments (Rai et al. 2003) using criteria set out in Graham et al. (2000). The unaligned total DNA sequence length obtained in monocot exemplars ranges from ca. 12.7 kb in *Burmannia* L. to ca. 17.9 kb in *Coelogyne* Lindl., with a mean unaligned length in monocots of ca. 15.9 kb. Some monocots (such as *Burmannia*) lack sequence data for one or more major regions (see Appendix). The combined alignment is nearly 32 kb in length (31,900 base pairs [BP]), approximately twice the unaligned length of any individual taxon. This size is a result of fairly extensive gaps and/or unalignable regions, including some in nonangiosperm taxa that were part of the overall alignment, but that were not considered in the current study. Those noncoding regions that were too difficult to align were set aside in staggered gapped regions, with the staggered elements largely restricted to single taxa. Because they are unique, such single-taxon elements are parsimony uninformative and have no influence on the analysis. This allowed us to avoid defining character exclusion sets in hard to align regions, a substantial undertaking for a matrix of this size. However, in a subset of cases, we were able to align small blocks of taxa within otherwise unalignable regions. Variable characters within these elements can therefore contribute to tree searching (for a comparable example, see Steane et al. 1999). All gap cells are treated as missing data. We did not attempt to score insertion/deletion (indel) events. Graham et al. (2000) provide an overview of the (relatively limited) utility of indels inferred from several plastid inverted-repeat noncoding regions for inference of deep angiosperm phylogeny.

In total, 5617 aligned sites are potentially parsimony informative in the full 94-taxon data set, or 4798 sites in the monocots alone. We analyzed the matrix using heuristic

maximum-parsimony (MP) searches with PAUP* vers. 4.0b10 (Swofford 2002), using tree-bisection-reconnection (TBR) branch swapping, and 100 random-addition replicates. Branch support was estimated using bootstrapping (Felsenstein 1985) with 100 bootstrap replicates. No tree limits were set, and all other settings used were the default ones. *Ceratophyllum* L. and Chloranthaceae have variable and poorly supported placements in basal angiosperm phylogeny (e.g., Qiu et al. 1999; Graham and Olmstead 2000; Graham et al. 2000; Soltis et al. 2000; Hilu et al. 2003; Davis et al. 2004), possibly because of a long, undivided subtending branch (*Ceratophyllum*) and short ''basal'' branches in angiosperm phylogeny (Chloranthaceae; e.g., Fig. 3 in Graham et al. 2000). We therefore performed a separate bootstrap analyses with these two taxa excluded, to examine their influence on phylogenetic analysis.

McPherson et al. (submitted) used simulated data sets based on several hypotheses of relationship for *Aphyllanthes monspeliensis* L., the sole member of Aphyllanthaceae (Asparagales), to demonstrate that there is substantial scope for systematic error in analyses that include this taxon, at least for the plastid regions examined here. Their analyses showed that a subset of hypotheses are recovered rarely when used as model trees, while others may be recovered frequently when they are not used to simulate data (i.e., high potential for type I and II errors). We therefore performed two basic sets of analyses with *Aphyllanthes* L. included or excluded from consideration (for more on this problematic taxon, see Fay et al. [2000]; Chase et al. 2005; Pires et al. 2005]; *Ceratophyllum* and Chloranthaceae were included in both cases.

To examine whether the rate of molecular evolution shifts in parallel among different codon position classes across the monocot backbone, we set up character sets (CHARSETs) in PAUP* that classify each nucleotide in the protein-coding regions into one of the three codon positions (reading frames were defined with respect to *Nicotiana tabacum* L. and *Ginkgo biloba* L. sequences). The short overlap between *pdb*D and *psb*C was ignored, as individual sites in this region have different codon positions in these two genes. Parsimony-inferred changes from each codon position were noted for each branch in monocot phylogeny, based on one of the most-parsimonious trees (see Fig. 2). We determined whether length estimates for the first two codon positions were correlated across branches with those for the third codon position, using JMP vers. 4.0.1 (SAS Institute 2000).

<div style="text-align:center">RESULTS</div>

Most aspects of outgroup relationship are discussed elsewhere (Graham and Olmstead 2000; Graham et al. 2000). *Amborella trichopoda* Baill. (Amborellaceae) was considered to be the sister group of all remaining angiosperms, following the current consensus result of recent large-scale analyses that included seed-plant outgroups (Mathews and Donoghue 1999; Parkinson et al. 1999; Qiu et al. 1999; Soltis et al. 1999; Soltis et al. 2000; Barkman et al. 2000; Graham and Olmstead 2000; Graham et al. 2000; Zanis et al. 2002; Hilu et al. 2003; Stefanovic et al. 2004). A recent result placing *Amborella* Baill. apart from the basal angiosperm split (Goremykin et al. 2003, 2004) is likely to be an artifact of extremely low taxon density, combined with ex-

ceptionally long branches in the exemplar taxa used by these authors for monocots, all of which are members of Poaceae (Soltis and Soltis 2004; Stefanovic et al. 2004; S. W. Graham unpubl. data).

We typically use family (or higher taxon) names to describe terminal taxa (for a rationale, see Chase et al. 2005). A detailed consideration of relationships recovered in Asparagales is provided in McPherson et al. (submitted); the results presented here for this order focus primarily on two taxa not considered previously for these plastid regions (*Agapanthus africanus* [Agapanthaceae] and *Doryanthes palmeri* [Doryanthaceae]). With *Aphyllanthes* excluded, most random-addition replicate searches found the same pair of trees (Fig. 1, 2) for 96 of 100 random-addition replicates, with a tree length of 35,826 steps, consistency index (CI), including all sites, of 0.362, and retention index (RI) of 0.474. With *Aphyllanthes* included, a single tree was found (a portion is shown in Fig. 3) in 95 of 100 random-addition replicates, with length 36,276 steps (CI = 0.360; RI = 0.473). The latter tree is topologically equivalent to one of the former trees when *Aphyllanthes* is pruned from it. We considered well-supported (''strongly supported'' or ''robust'') branches to have bootstrap support of ca. 95% and more, and poorly supported branches to have ca. 75% or less bootstrap support. By this criterion, most branches of monocot phylogeny are well supported by the current data (Fig. 1). To simplify the presentation of the results, we focus initially on presenting those branches that are poorly to moderately supported, and then address the well-supported relationships. All bootstrap values below refer to analyses with *Aphyllanthes* excluded, and *Ceratophyllum* and Chloranthaceae included, unless otherwise stated.

Poorly to Moderately Supported Relationships

We infer the eudicots plus *Ceratophyllum* to be the sister group of the monocots, and Chloranthaceae to be the sister group of the magnoliids (see also Graham et al. 2000). Both placements have weak bootstrap support from bootstrap analysis (67% and 60% BP, respectively; Fig. 1A). When *Ceratophyllum* and Chloranthaceae are removed from phylogenetic analysis, this appears to stabilize the inference of some of the remaining basal angiosperm relationships, with substantially improved bootstrap support for monocot monophyly (100% vs. 84% BP; Fig. 1A), and marginal improvement in support for a close relationship between the monocots and eudicots (73% vs. 67% BP; Fig. 1A). The bootstrap values shown in the monocots (Fig. 1B) are from analyses with *Ceratophyllum* and Chloranthaceae included. Excluding these two taxa has very little effect on support values within the monocots (data not shown).

A few deep branches of monocot phylogeny are not strongly supported. These include the precise order of splits at the base of Alismatales: Araceae are found to be the sister group to a clade of three sampled alismatid families (83% BP). The arrangement of Poales, Commelinales–Zingiberales, Arecales, and Dasypogonaceae at the base of the commelinid monocots is also unclear; the two branches supporting the relationship observed here each have <50% BP. A sister-group relationship between Liliales and Asparaga-

Fig. 1.—One of two most-parsimonious trees inferred from a large plastid data set (*atp*B, *ndh*B, *ndh*F, ten photosystem genes, *rbc*L, *rpl*2, *rps*7, 3′-*rps*12, and various introns and noncoding regions; see text).—1A. Relationships among the major angiosperm lineages.—1B. Phylogenetic relationships in the monocots. Numbers above branches are results of bootstrap analyses using all taxa; those below branches are results when *Ceratophyllum* and Chloranthaceae are excluded from consideration (only the former are shown only for Fig. 1B). *Aphyllanthes monspeliensis* is excluded from both analyses (but see Fig. 3). An arrowhead (Fig. 1B) indicates a branch not seen on

les–commelinids is only weakly supported (70% BP), as is that between Dioscoreales and Pandanales (63% BP).

Melanthiaceae are found to be the sister group to the other families of Liliales sampled here, but with only 39% BP. Of the two members of Asparagales new to this study, Agapanthaceae are clearly a member of a clade of three closely related families (Agapanthaceae, Alliaceae s.s., Amaryllidaceae). However, inferred interrelationships among these three taxa are poorly supported, with only 37% support for one of the arrangements shown here (Alliaceae–Amaryllidaceae; Fig. 1B). An alternative arrangement, Alliaceae–Agapanthaceae, is found on the other most parsimonious tree, with 56% BP (Fig. 3). We find Sparganiaceae–Typhaceae and Bromeliaceae to be successive sister groups of the remaining families of Poales, but with only 45% BP for this arrangement. However, these two lineages are strongly supported as emerging from the base of Poales (Fig. 1B). An arrangement of Flagellariaceae and Restionaceae as successive sister groups of the remaining sampled graminid taxa (Ecdeiocoleaceae and Poaceae) has moderately strong support (87% BP).

Strongly Supported Relationships

The remaining relationships are nearly all strongly supported by bootstrap analysis (ca. 95% BP and higher; Fig. 1B). Ignoring the depression in support for monocot monophyly when *Ceratophyllum* and Chloranthaceae are included (84% vs. 100%; Fig. 1A), *Acorus* L. is robustly inferred to be the sister group of all other monocots. Alismatales and Petrosaviaceae are (respectively) the next successive sister groups of the remaining monocots. Asparagales and the commelinid monocots are well supported as sister taxa. All orders consisting of more than one family (sensu APG II 2003) are well supported at the taxon samplings used here. Several multi-ordinal clades are also strongly supported, including Commelinales–Zingiberales, and the commelinid monocots as a whole (Arecales, Commelinales–Zingiberales, Poales, Dasypogonaceae).

Most relationships within orders are also robustly supported. The (partly) mycoheterotrophic family Burmanniaceae is strongly supported as the sister group of Dioscoreaceae (only these two members of Dioscoreales were sampled). Apart from the basal split in Liliales, all relationships in this order are well supported. Petermanniaceae are distinct from Colchicaceae, whose sister group is Alstroemeriaceae of taxa sampled. Within Alismatales, the inferred relationships of Alismataceae, Butomaceae, and Scheuchzeriaceae are all well supported (the former two are sister groups with respect to the taxa sampled). Within Pandanales, both internal branches are well supported; Velloziaceae are the sister group of the remaining Pandanales, and Pandanaceae and Cyclanthaceae are sister taxa. Pontederiaceae are the sister group of Haemodoraceae of families sampled in Commelinales. Apart from the base of Poales, relationships among the remaining taxa of Poales are almost all well supported. Cyperaceae–Xyridaceae are the sister group of Mayacaceae. This cyperid clade is the sister group of the graminid families. Ecdeiocoleaceae and Poaceae are sister groups with respect to the taxa included.

The position of Doryanthaceae (a member of Asparagales not sampled by McPherson et al. (submitted) is quite well supported (92% BP). It is inferred to be the sister group of a large clade consisting of Iridaceae, Xeronemataceae, Alliaceae s.l., Asparagaceae s.l., and Xanthorrhoeaceae s.l. The latter three taxa (Alliaceae s.l., Asparagaceae s.l., and Xanthorrhoeaceae s.l.) are indicated with numerals in Fig. 1B, and represent the more inclusive versions of these families in APG II (2003). Iridaceae are sister to a clade consisting of Xeronemataceae, Alliaceae s.l., Asparagaceae s.l. and Xanthorrhoeaceae s.l. A well-supported clade comprised of Ixioliriaceae and Tecophilaeaceae is the sister group to all of these taxa. Bootstrap support for this major part of the backbone of Asparagales [Asparagales base, ((Ixioliriaceae–Tecophilaeaceae), (Doryanthaceae, (Iridaceae, (Xeronemataceae, (Xanthorrhoeaceae s.l., (Asparagaceae s.l.–Alliaceae s.l.))))))] is thus robustly supported by our data.

Effect of Inclusion of Aphyllanthes monspeliensis (Aphyllanthaceae)

When *Aphyllanthes* is included in the analysis, it is resolved as the sister group of Agavaceae, but with poor support (50% BP, Fig. 3; see also McPherson et al. submitted). Its inclusion does not impinge on other inferred relationships, but it does moderately depress bootstrap values for five of ten branches (by 10–25%) in a local cluster of families corresponding to Asparagaceae s.l., plus Alliaceae s.l. (Fig. 3). There is strong support (100% BP) for the inclusion of *Aphyllanthes* within the major clade consisting of Alliaceae s.l. and Asparagaceae s.l., and bootstrap support outside this local cluster does not appear to be adversely affected by its inclusion (data not shown).

DISCUSSION

The plastid gene set considered here has been used for phylogenetic inference of other deep and difficult phylogenetic problems (Graham and Olmstead 2000; Rai et al. 2003) and provides strong support for a number of deep monocot relationships that were poorly supported or unresolved in earlier studies. Average bootstrap support for the monocot portion of the tree from our data is ca. 91% per branch, based on 65 internal branches involving 68 monocot taxa (*Aphyllanthes* excluded). Three-quarters of the internal monocot branches have at least 95% bootstrap support, and 90% have at least 70% bootstrap support. Only six monocot branches on either shortest tree have less than 50% bootstrap support (Fig. 1B). Some of these improvements in support compared to earlier studies are paralleled in other recent

←

both shortest trees (the alternative arrangement with bootstrap values is shown in Fig. 3). Order and family placements and designations follow APG II (2003), using the optional "bracketed" system in Asparagales ("1" = Asparagaceae s.l.; "2" = Alliaceae s.l.; "3" = Xanthorrhoeaceae s.l.; Fig. 1B), except that Petermanniaceae are also recognized as a family (see text). Petrosaviaceae and Dasypogonaceae are currently unplaced to order. Tree length and fit statistics are provided in the main text.
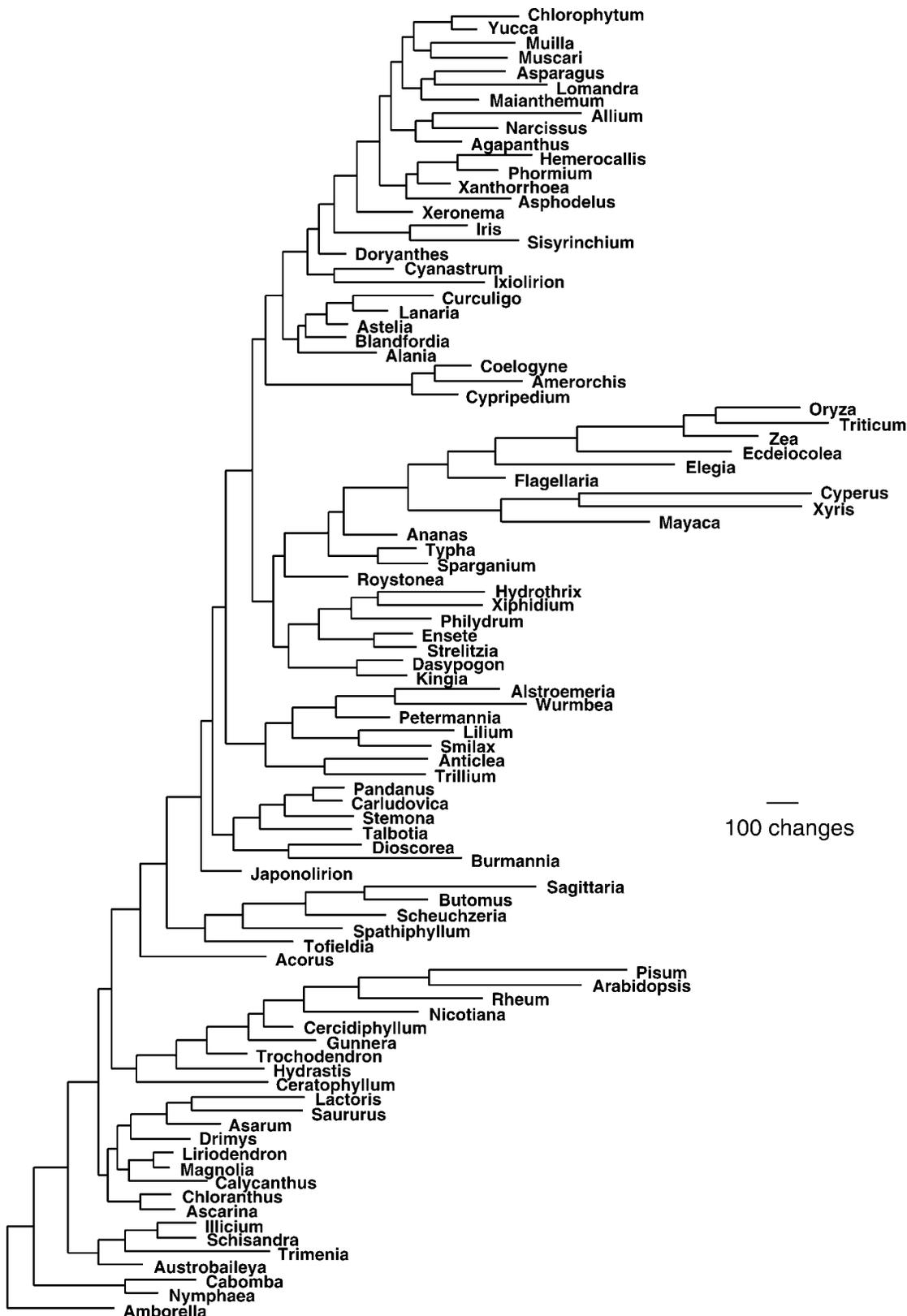
Fig. 2.—One of two most-parsimonious trees (Fig. 1), presented as a phylogram (ACCTRAN optimization). The scale bar indicates 100 changes.
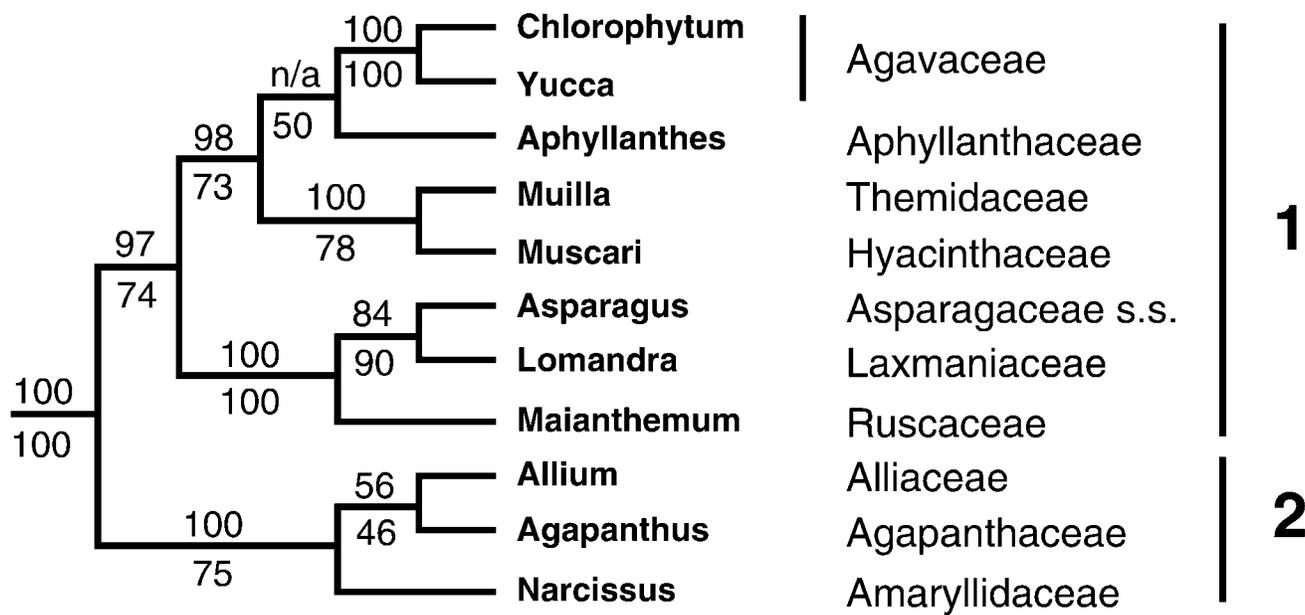
Fig. 3.—Region of disruption of bootstrap values in Asparagales when *Aphyllanthes monspeliensis* (Aphyllanthaceae) is included in analysis ("1" = Asparagaceae s.l.; "2" = Alliaceae s.l.). Numbers above branches are bootstrap values with *Aphyllanthes* excluded, those below branches are bootstrap values with it included. The tree shown is a portion of the most-parsimonious tree found with *Aphyllanthes* included (see text); this tree is otherwise identical to one of the two trees found with *Aphyllanthes* excluded (the other is shown in Figs. 1, 2), with comparable bootstrap values for the rest of the tree.

studies, including the two-gene sampling of Tamura et al. (2004), and the multigene studies of Chase et al. (2005) and Pires et al. (2005). A few clades are poorly or moderately well supported here, but are nonetheless congruent with other studies (e.g., the poorly supported position of Melanthiaceae relative to other sampled members of Liliales, which is consistent with that in Tamura et al. [2004], Givnish et al. [2005], and one of the analyses in Chase et al. [2005]).

Assessing the Strength of Support of Deep Monocot Relationships

Sampling error will tend to reduce estimated branch support when too few data are collected per taxon ("not enough characters examined"), or when too few characters define individual branches ("rapid radiations"). However, even when deep branches are short, simulation results using available models of DNA sequence evolution (e.g., Hillis 1998) indicate that maximum parsimony can yield very accurate reconstructions for relatively small amounts of data per taxon (a few thousand kb of DNA sequence data), provided a sufficiently dense taxon sampling is employed. Nonetheless, empirical analyses based on real DNA sequence data sets that are densely sampled and of this order of size (several kb long) are incompletely congruent with each other concerning major and minor points of relationship, and have numerous areas with weak statistical support (e.g., Soltis et al. 2000; Davis et al. 2004; Tamura et al. 2004; Chase et al. 2005; Pires et al. 2005), or exclude many problematic taxa (Tamura et al. 2004). The existence of these uncertain areas (e.g., the composition or relative arrangements of Asparagales, Dioscoreales, Liliales, Pandanales, and the major commelinid lineages) provides a continuing impetus for expansion of the amount of data collected per taxon.

Empiricists do not have access to the correct tree of monocot phylogeny, but instead use a variety of statistical methods, such as bootstrap analysis, to assess the degree of confidence in phylogenetic inference (e.g., Hillis and Bull 1993; Felsenstein 2004). There is no clear consensus on what the cutoff should be for considering a clade to be strongly supported ("robust"), and there is some disagreement about why different methods provide different pictures of clade support. Bayesian phylogenetic inference, for example, is thought to suffer from inflated clade posterior probability estimates (e.g., Suzuki et al. 2002; Erixon et al. 2003; Lemmon and Moriarty 2004), and empirical and theoretical studies have demonstrated that bootstrap analysis can provide biased measures of support (e.g., Felsenstein and Kishino 1993; Hillis and Bull 1993). Small differences in clade support are also demonstrable for different resampling methods (jackknife, bootstrap), or different implementation strategies for these methods (Davis et al. 2004). However, the numerical value that we accepted to indicate well-supported clades in bootstrap analysis (ca. 95%) may often be on the conservative side for bootstrap analysis (Felsenstein and Kishino 1993; Hillis and Bull 1993).

A caveat for bootstrap analysis (and other methods for statistical inference of branch support) is that there are conditions, such as the oft-cited phenomenon of "long-branch attraction," under which phylogenetic inferences can be misleading (Felsenstein 1978; Hendy and Penny 1989). Long-branch attraction has been invoked for the placement of several problematic monocot taxa (e.g., the positions of Ixioliriaceae and Aphyllanthaceae in Asparagales; Fay et al. 2000). Parametric phylogenetic methods such as maximum likelihood and Bayesian analysis have been found to be less prone to the distorting effects of long-branch attraction on

phylogenetic analysis (e.g., Chang 1996; Swofford et al. 2001), unless there are discordant changes in evolutionary rates among characters (''heterotachy'')—in which case, maximum parsimony may be more reliable (Kolaczkowski and Thornton 2004).

The overall rate of molecular evolution in monocot plastid genomes observed here is quite variable among lineages (Fig. 2), in line with previous studies (Wilson et al. 1990; Bousquet et al. 1992; Gaut et al. 1992, 1996). The effect of this rate variation on phylogenetic analysis is unclear. However, we can at least rule out substantial heterotachy between two codon position partitions in our protein-coding regions—the first two vs. third codon positions. The former should predominantly reflect nonsynonymous substitutions, and the latter synonymous ones (e.g., Sanderson et al. 2000). We might therefore expect these to have different substitution dynamics among different lineages, based on theoretical considerations (Kimura 1968, 1983; Ohta 1992). In practice, however, the rate of synonymous and nonsynonymous substitutions are correlated in a broad variety of organisms (e.g., Sharp 1991; Wolfe and Sharp 1993; Akashi 1994; Gaut et al. 1996). We examined the amount of change in the first two vs. third codon positions, using this information as a rough proxy of nonsynonymous vs. synonymous changes. Despite inconstancy in the overall rate of molecular evolution (Fig. 2), changes in these two functionally defined codon-position classes are strongly correlated across the sam-

pled branches of the monocot tree ($r = 0.9036$; $P < 0.0001$; Fig. 4).

We will not enter further into the debate about different measures of statistical support. However, a different framework for documenting the reliability of phylogenetic results is to demonstrate that analyses involving the same taxa for different genomic regions depict similar relationships (e.g., Penny et al. 1982). If there is instead well-supported incongruence among different genetic linkage groups, this may reflect deviations of particular gene trees from a consensus organismal pattern (e.g., Maddison 1997) or various phenomena that lead to strong systematic biases in the data (Naylor and Brown 1998). In general, any misleading effects of these on inferences of higher-order relationships using plastid data alone are expected to be rather small (Savolainen et al. 2002; see also Chase et al. 2005). However, determining whether this is indeed the case will at least require the availability of trees from multiple linkage groups (the plastid genome is a single linkage group), with each tree as well supported as possible.

The correlated change we observe among codon positions (Fig. 4) may explain why inferences made from subpartitions of our data (different codon position sets) are generally highly congruent with each other (e.g., McPherson et al. submitted). Our results are also generally highly congruent with other studies (e.g., Davis et al. 2004; Tamura et al. 2004; Chase et al. 2005; Pires et al. 2005). Where incongruent, the
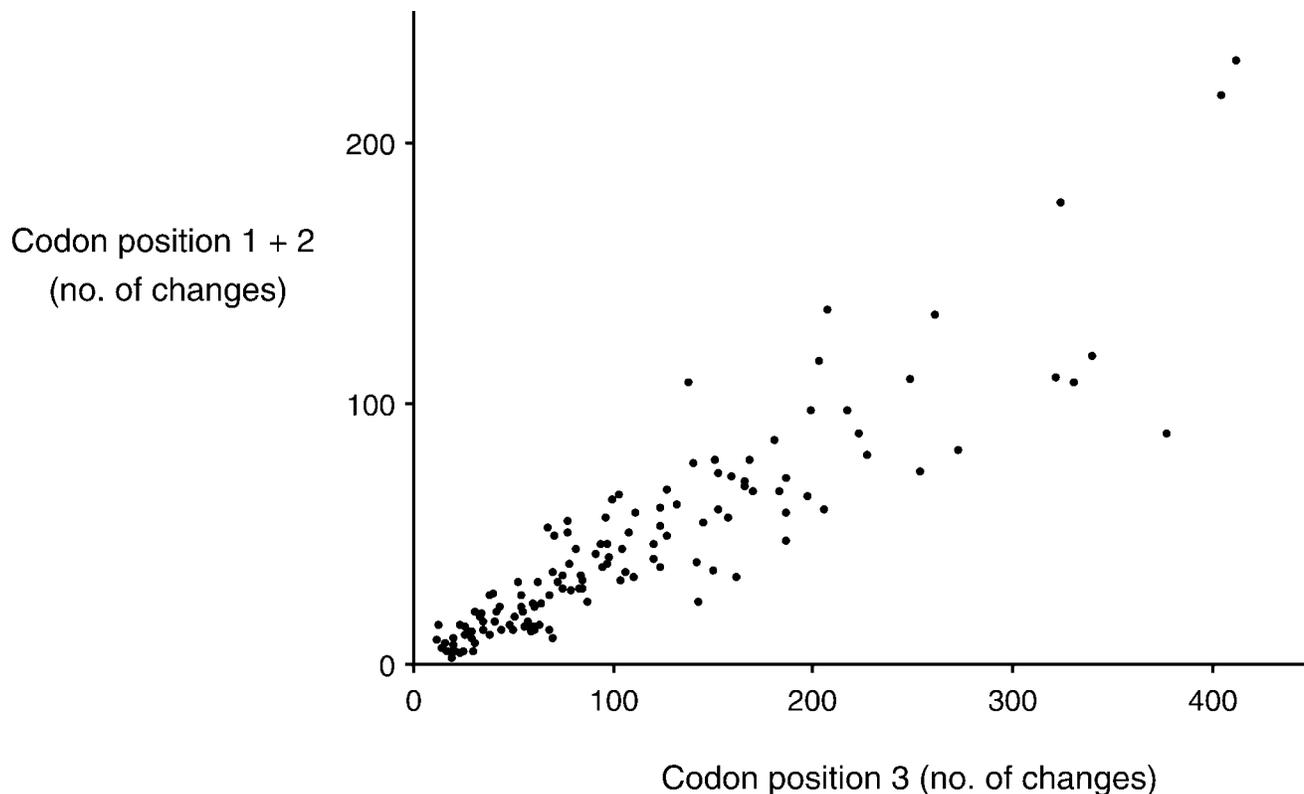

Fig. 4.—Demonstration of parallel substitution rates in codon positions 1 and 2, vs. codon position 3 across monocot phylogeny. Data points for each codon set represent individual branch length estimates, summed across relevant sites in protein-coding regions. All terminal and internal branches in monocot phylogeny were considered, including the branch immediately subtending the monocots (Fig. 2). Changes along each branch were computed using ACCTRAN optimization for the relevant character set.

other studies generally have poor support for the conflicting relationship (see below for several exceptions). Using the phylogenetic data presented here, we have therefore come close to, but not yet fully attained, a phylogenetic backbone of monocot relationships that is well supported across all deep nodes. Other work (in progress) will address more of the details of this plastid framework by adding currently unsampled families.

Remaining Problematic Relationships in Monocot Phylogeny

A broad circumscription of the magnoliids (''eumagnoliids'') in Soltis et al. (1999) and Soltis et al. (2000) includes the monocots, based on a weakly supported clade found in their analysis. In contrast, APG II (2003) was more wary, accepting only Canellales, Laurales, Magnoliales, and Piperales as magnoliids. Although a sister-group relationship between the magnoliids (in a narrow sense) and monocots cannot be ruled out from our bootstrap analysis, classifications that depend on this or other placements should be treated very cautiously. Duvall et al. (2005) use various optimality criteria to analyze four concatenated genes and find several placements of the monocots in angiosperm phylogeny; for example, their maximum parsimony analysis depicts the eudicots as the sister group of monocots plus *Ceratophyllum* (with weak support for these relationships), while their Bayesian analysis places the monocots as the sister group of the magnoliids, with posterior probabilities of 0.97–1.00 for all relevant branches. Given the potential for Bayesian support values to be inflated or misleading (discussed above), the relatively uncertain position of the monocots observed in our analysis, and their variable position in other studies (e.g., Hilu et al. 2003; Davis et al. 2004; Chase et al. 2005), the jury is thus still out on where the monocots belong in flowering-plant phylogeny. We also observed a moderately depressive effect on the strength of support for monocot monophyly with Ceratophyllaceae and Chloranthaceae included in analysis. Although straightforward to demonstrate, the cause of this intriguing phenomenon is not clear, although it is conceivably associated with whatever is causing the uncertain placement of these two families in all current angiosperm-wide studies.

The monophyly of the monocots, and the position of *Acorus* as the sister group of all other extant monocots are both strongly supported here. *Acorus* has not been uniformly supported as the sister group of all monocots in all recent studies (e.g., Qiu et al. 1999), but when it is not, this generally is a function of poor statistical support in individual studies. However, Duvall and Ervin (2004) and Duvall et al. (2005), documented problems with the nuclear 18S rDNA locus for this taxon. A further exception to this uniform picture is a recent study by Davis et al. (2004) who examined monocot deep phylogeny using *rbc*L and the mitochondrial locus *atp*A. In combined analyses of these two genes, they found a strongly supported sister-group relationship between Acoraceae and a major clade of Alismatales consisting of all the taxa they sampled in this order, except Araceae and Tofieldiaceae. Their analysis of *rbc*L alone depicted *Acorus* as the sister group to all other monocots, but that of *atp*A alone depicted Acoraceae and associated alismatid taxa on very long branches, in a position nested deep in monocot phylogeny, as part of a small clade that included several taxa from Asparagales (Ixioliriaceae, Iridaceae, and a member of Agavaceae; see Fig. 4 in Davis et al. 2004, one of their most-parsimonious trees). This result, and the discordant placement of *Acorus* in their combined analyses, may be an artifact (see also Chase 2004).

Very short branches are more prone to the effects of sampling error, which may consequently contribute to poor support for some deep internal branches in monocot phylogeny (e.g., Chase et al. 2000, 2005). However, some of the ''short'' branches referred to in earlier studies still are weakly supported in our expanded data sampling and in comparable recent studies (e.g., Chase et al. 2005), despite not being clearly different in inferred length from neighboring branches that have strong support (compare, for example, the lengths and support for branches subtending Asparagales vs. Dioscoreales–Pandanales; Figs. 1B, 2). Relative unevenness of branch lengths can potentially contribute to erroneous or unstable phylogenetic inference due to long-branch attraction (Felsenstein 1978; Hendy and Penny 1989). However, if there are any problematic long branches remaining in the current data set, their effect here may be primarily to destabilize local estimates of bootstrap support, rather than to lead to erroneous placement of the affected clades. Although it is not always possible to do so, additional taxon sampling may often (although not always; e.g., Rannala et al. 1998; Poe and Swofford 1999) help ameliorate the effects of long branches, by breaking them up.

Nonetheless, disparity in rates of evolution may contribute to difficulties in inferring some relationships accurately. The elevated rate in the grasses commented on by other workers (e.g., Gaut et al. 1992) is evidently not unique to them within Poales (rate elevation compared to other monocots is evident from visual inspection of branches subtending Cyperaceae, Ecdeiocoleaceae, Mayacaceae, Poaceae, Restionaceae, and Xyridaceae; Fig. 2). However, the branches immediately subtending Bromeliaceae, Sparganiaceae–Typhaceae, and Flagellariaceae are short relative to other members of Poales (Fig. 2). This disparity might explain why the relative positions of Bromeliaceae and Sparganiaceae–Typhaceae are unclear with our current taxon sampling (Fig. 1B), and why the backbone relationships inferred with regards to Flagellariaceae, Restionaceae, and the other graminid families show moderately strong conflict with those inferred in Chase et al. (2005; they see the reciprocal relative arrangement of Flagellariaceae and Restionaceae).

A few other higher-order groupings are in moderately strong disagreement with clades reported in Chase et al. (2005). These include the relationships among three major clades of Alismatales. We find Tofieldiaceae to be the sister group of a moderately well-supported clade consisting of Araceae and the remaining Alismatales. This result was also observed in some analyses of the three-gene data set of Chase et al. (2000), but conflicts with a well-supported linkage between Tofieldiaceae and alismatid families in Chase et al. (2005). In addition, we observe Mayacaceae to be the sister group of Cyperaceae and Xyridaceae of taxa sampled in the cyperid clade, with strong support. Chase et al. (2005) instead find moderate support for a closer relationship between Mayacaceae and Cyperaceae. Addressing such con-

flicts may require improved taxon sampling from the current data, work that is currently in progress.

Fay et al. (2000) posited that addition of more data per taxon would be required to solve the problematic positions of Aphyllanthaceae and Ixioliriaceae in Asparagales, as both had labile positions in their analyses and both are relatively isolated taxa. Despite the addition of four more genes to the complement employed by Fay et al. (2000), the position of *Aphyllanthes* in the analysis of Chase et al. (2005) and Pires et al. (2005) is still labile and weakly supported. McPherson et al. (submitted) used simulation studies to demonstrate that inference of the phylogenetic position of *Aphyllanthes* in Asparagales has a high error rate. Uncertainty in the placement of *Aphyllanthes* in Asparagaceae s.l. may be a function of it being on a long terminal branch (e.g., Fay et al. 2000; McPherson et al. submitted). Its inclusion in analysis does not appear to affect the underlying relationships inferred for other taxa in this family, although it depresses bootstrap support values in the local clade that includes it (Asparagaceae s.l.–Alliaceae s.l.; Fig. 3).

We inferred strong support for a sister-group relationship between Ixioliriaceae and Tecophilaeaceae, a result that was seen with poor support in some trees inferred by Fay et al. (2000), and with moderate support in the analysis of Pires et al. (2005). This clade is contradicted by the four- and seven-gene analyses of Chase et al. (2005), and the two-gene analysis by Davis et al. (2004), who instead find strong to moderate support for a sister-group relationship between Ixioliriaceae and Iridaceae. Our analyses robustly resolved this midpoint of the Asparagales backbone (Fig. 1B), but the strongly discordant positions of Ixioliriaceae among studies clearly requires further attention. Further taxon sampling (to density levels comparable to Pires et al. 2005) among the relatives of Ixioliriaceae and Aphyllanthaceae may help clarify their phylogenetic status.

Unusual and disparate placements of Burmanniaceae were observed in early analyses of monocot *rbc*L data. Gaut et al. (1992) found Burmanniaceae nested in what is now referred to as the commelinid monocots, and Duvall et al. (1993) found it nested in Asparagales. Both placements may have been due to the long branch associated with this family, coupled with the relatively limited taxon sampling used in both studies. A relatively long branch also subtends our exemplar taxon from Burmanniaceae (*Burmannia capitata*; Fig. 2). We find strong bootstrap support for its position as the sister group of Dioscoreaceae, of taxa examined here (Fig. 1B). Some members of Burmanniaceae are completely mycoheterotrophic and achlorophyllous, including some taxa in *Burmannia.* Although photosynthesis has not been characterized physiologically in *B. capitata,* this species is chlorophyllous (Imhof 1999). It also has uninterrupted reading frames for all 16 protein-coding regions examined here (*ndh*F was not examined for this species), including all ten photosystem II genes, *atp*B and *rbc*L. This suggests that these loci produce gene products that are functional in photosynthesis. Nonetheless, partial heterotrophy (if present in this taxon) may contribute to its relatively long subtending branch. Whatever the cause, this long branch could also result in misleadingly high bootstrap support. However, the position we inferred for Burmanniaceae (Fig. 1B) is congruent with other recent phylogenetic studies based on molec-

ular and morphological data, which lends more credence to the idea that this taxon has been correctly placed among the deep branches of monocot phylogeny. Chase et al. (1995b, 2000, 2005) and Caddick et al. (2002) found Burmanniaceae to be nested in a redefined Dioscoreales (APG 1998, APG II 2003), although in contrast to our study they found only poor support for this clade as a whole. Improved taxon sampling in Dioscoreales using the plastid regions sampled here should be valuable for further clarifying relationships among the constituent taxa in this order.

Contributions to Our Knowledge of Monocot Higher-Order Relationships

One of our most significant findings is the well-supported placement of Petrosaviaceae in monocot phylogeny. This family (represented here by *Japonolirion osense*; see Cameron et al. 2003) is strongly supported as the sister group of all monocots except *Acorus* and Alismatales, and is thus (apparently) the sole extant descendant of a very early split in monocot phylogeny. This supports the idea that the family should be recognized in its own order (Petrosaviales Takht.) in rank-based classifications (see also Chase 2004). The family's position was only partly resolved in the plastid-based study of Chase et al. (2000), but the moderate to strong support for the relationship observed here is also seen in the seven-gene sampling of Chase et al. (2005), and the two-gene sampling of Tamura et al. (2004). However, Tamura et al. (2004) used only two outgroup taxa in total, and this sparse sampling may have inflated support levels for their basal monocot inferences.

Additional major findings include a well-supported relationship between Commelinales and Zingiberales (see also Tamura et al. 2004; Chase et al. 2005; Pires et al. 2005), and strong support for a sister-group relationship between Asparagales and the commelinid monocots. The latter relationship has been seen in several other studies with comparable taxon samplings, but with only poor support (e.g., Tamura et al. 2004; Chase et al. 2005). Our most densely sampled major clade is Asparagales. All ''unbracketed'' families of Asparagales (in the sense of APG II 2003) have now been sampled for the regions considered here, and most of their interrelationships are inferred with strong support (Fig. 1B). The spine of inferred relationships in Asparagales largely parallels other recent studies (Fay et al. 2000; Chase et al. 2005; Givnish et al. 2005; Pires et al. 2005), but is generally better supported here. The orchids are well supported here as a member of Asparagales (see also Tamura et al. 2004; Chase et al. 2005), and their inclusion helps define the deepest nodes in that order (McPherson et al. submitted).

The position of Doryanthaceae (as the sister group of a large group of Asparagales that includes Alliaceae s.l., Asparagaceae s.l., Iridaceae, Xanthorrhoeaceae s.l., and Xeronemataceae; Fig. 1B) is strongly supported in our analysis, and is also found with weak support in Fay et al. (2000) and Pires et al. (2005). Although their precise interrelationships are unclear, Agapanthaceae, Alliaceae s.s., and Amaryllidaceae are linked in a strongly supported clade (Alliaceae s.l.), a relationship seen by Meerow et al. (1999), but without strong support. Fay et al. (2000) found this relationship with strong support, although a reanalysis using standard parsi-

mony by McPherson et al. (2004) found only moderate support using the same data. Pires et al. (2005) also find this clade with strong support, and find good support for Agapanthaceae as the sister group of a clade consisting of Alliaceae s.s. and Amaryllidaceae s.s. The larger clade (Alliaceae s.l.) is unperturbed by the inclusion of *Aphyllanthes* in our analyses, although its level of support is somewhat adversely affected (Fig. 3). Other aspects of Asparagales phylogeny inferred from our data are discussed in more detail in McPherson et al. (submitted).

Most other relationships that are strongly supported are also largely or completely congruent with the comparable taxonomic sampling by Chase et al. (2005). These include most relationships within Liliales, those among the four sampled families of Pandanales (Cyclanthaceae, Pandanaceae, Stemonaceae, and Velloziaceae) and most of the relationships within the commelinids, including relationships within Commelinales, and most relationships within Poales. Our sequences of *Petermannia* F. Muell. (Petermanniaceae) were not derived from the misidentified sample included in Rudall et al. (2000); see Chase et al. (2005). The distinct position of *Petermannia* in Liliales could be dealt with in the APG system by recognition of Petermanniaceae (Fig. 1B), or perhaps by a substantial expansion in the circumscription of Colchicaceae. Further inference of phylogenetic relationships within these orders will be addressed in more detail elsewhere using improved taxon sampling.

The phylogenetic status of Daspyogonaceae is still unclear, although our data and others (e.g., Chase et al. 2000, 2005; Davis et al. 2004; Givnish et al. 2005) indicate that it is the sole extant representative of a lineage that diverged very early in the history of the commelinid monocots. As in other recent studies, we did not find strong support for any particular arrangement of this family and Arecales, Commelinales–Zingiberales, and Poales at the base of the commelinid clade (Fig. 1B). Various arrangements of these four lineages have been observed (e.g., Davis et al. 2004; Chase et al. 2005; Givnish et al. 2005). Tamura et al. (2004) found moderate support for a sister-group relationship between Poales and Commelinales–Zingiberales, but they did not sample Dasypogonaceae. The relationships observed here, with Dasypogonaceae as the sister group of Commelinales–Zingiberales, and Arecales as the sister group of Poales, have very poor support (38% and 33% BP, respectively). However, if correct, this scenario would either require elevation of Dasypogonaceae to ordinal status, or a substantial reworking of current ordinal boundaries in the commelinid monocots. A sister-group relationship between Commelinales and Zingiberales has strong support, so it would not be acceptable to sink Dasypogonaceae in either order. If Dasypogonaceae, Commelinales, and Zingiberales were instead combined in a single order, Commelinales is the name at the rank of order with the earliest use (see APG II 2003). Most other arrangements at the base of the commelinid monocots would also require recognition of Dasypogonaceae as a distinct order in the APG system of classification. However, sister-group relationships between Dasypogonaceae and Poales or Arecales have minor bootstrap support (36% BP for the former, 11% BP for the latter), and if either relationship is eventually shown to be correct, Dasypogonaceae could be included in the respective order. While Dasypogonaceae are confirmed here to be part of a deep-diverging split in the commelinid monocots, it is clear that more data are needed to satisfactorily resolve the phylogenetic and taxonomic status of this problematic family.

Looking Forward

The data presented here should contribute to an ongoing renaissance in our understanding of monocot systematics and evolution, which was sparked by the morphological work of Rolf Dahlgren and colleagues (e.g., Dahlgren et al. 1985) and further promoted by the first large-scale molecular studies (e.g., Chase et al. 1993). Our data provide a more robust framework for making inferences about molecular and morphological evolution, and should help with the fine tuning of higher-order monocot classification. The largely robust framework of monocot deep phylogeny presented here, which includes exemplar taxa from all major chlorophyllous clades and many of the most problematic taxa, demonstrates clearly the value of an expanded plastid genomic sampling, by yielding results that are both well supported and congruent with those inferred in other recent studies. We are continuing taxon sampling using the current genomic set in undersampled groups (such as Alismatales, Liliales, and Poales). We need to improve our understanding of relationships within each of these major clades, and to address the remaining weak nodes along the spine of monocot phylogeny (particularly the relative positions of Asparagales–commelinids, Dioscoreales, Liliales, and Pandanales). Improved taxon sampling outside the monocots may also help address the position of the monocots in the angiosperms. We expect that at least some of these problems will require collection of further data per taxon (from the plastid and/or other genomes), and likely substantially more data in a subset of cases (perhaps of the order of whole plastid genome sampling).

Our study demonstrates the benefits of continued expansion in plastid genome sampling for addressing unresolved problems in monocot deep phylogenetics. While there are obvious benefits to examining other genomic regions (e.g., for evidence of intergenomic incongruence, or for finding strongly supported placements of achlorophyllous mycoheterotrophic monocot groups such as Triuridaceae) and morphological characters (e.g., for finding characters that support new groupings implied by molecular data), the time has not yet come to decelerate sampling in the plastid genome for monocot systematics. Indeed, the rate of increase in plastid-based studies at all levels of plant phylogeny has not yet leveled off (e.g., Shaw et al. 2005). This genome has proven to be a workhorse of modern monocot systematics, and we predict that it will likely remain so for the foreseeable future.

LITERATURE CITED

AKASHI, H. 1994. Synonymous codon usage in *Drosophila melanogaster*: natural selection and translation accuracy. *Genetics* **136**: 927–935.

ANGIOSPERM PHYLOGENY GROUP [APG]. 1998. An ordinal classification for the families of flowering plants. *Ann. Missouri Bot. Gard.* **85**: 531–553.

ANGIOSPERM PHYLOGENY GROUP II [APG II]. 2003. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG II. *Bot. J. Linn. Soc.* **141**: 399–436.

BARKMAN, T. J., G. CHENERY, J. R. MCNEAL, J. LYONS-WEILER, W. J. ELLISENS, G. MOORE, A. D. WOLFE, AND C. W. DEPAMPHILIS. 2000. Independent and combined analysis of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. *Proc. Natl. Acad. Sci. U.S.A.* **97**: 13166–13171.

BOUSQUET, J., S. H. STRAUSS, A. H. DOERKSEN, AND R. A. PRICE. 1992. Extensive variation in the evolutionary rate of *rbc*L genes sequences among seed plants. *Proc. Natl. Acad. Sci. U.S.A.* **89**: 7844–7848.

CADDICK, L. R., P. J. RUDALL, P. WILKIN, T. A. HEDDERSON, AND M. W. CHASE. 2002. Phylogenetics of Dioscoreales based on combined analyses of morphological and molecular data. *Bot. J. Linn. Soc.* **138**: 123–144.

CAMERON, K. M., M. W. CHASE, AND P. J. RUDALL. 2003. Re-circumscription of the monocotyledonous family Petrosaviaceae to include *Japonolirion. Brittonia* **55**: 214–225.

CHANG, J. T. 1996. Full reconstruction of Markov models on evolutionary trees: identifiability and consistency. *Math. Biosci.* **137**: 51–73.

CHASE, M. W. 2004. Monocot relationships: an overview. *Amer. J. Bot.* **91**: 1645–1655.

———, M. R. DUVALL, H. G. HILLS, J. G. CONRAN, A. V. COX, L. E. EGUIARTE, J. HARTWELL, M. F. FAY, L. R. CADDICK, K. M. CAMERON, AND S. HOOT. 1995*a*. Molecular phylogenetics of Lilianae, pp. 109–137. *In* P. J. Rudall, P. J. Cribb, D. F. Cutler, and C. J. Humphries [eds.], Monocotyledons: systematics and evolution. Royal Botanic Gardens, Kew, Richmond, Surrey, UK.

———, M. F. FAY, D. S. DEVEY, O. MAURIN, N. RØNSTED, J. DAVIES, Y. PILLON, G. PETERSEN, O. SEBERG, M. N. TAMURA, C. B. ASMUSSEN, K. HILU, T. BORSCH, J. I DAVIS, D. W. STEVENSON, J. C. PIRES, T. J. GIVNISH, K. J. SYTSMA, M. A. MCPHERSON, S. W. GRAHAM, AND H. S. RAI. 2005. Multigene analyses of monocot relationships: a summary, pp. xx–xx. *In* J. T. Columbus, E. A. Friar, J. M. Porter, L. M. Prince, and M. G. Simpson [eds.], Monocots: comparative biology and evolution, 2 vols. Rancho Santa Ana Botanic Garden, Claremont, California, USA.

———, D. E. SOLTIS, R. G. OLMSTEAD, D. MORGAN, D. H. LES, B. D. MISHLER, M. R. DUVALL, R. A. PRICE, H. G. HILLS, Y.-L. QIU, K. A. KRON, J. H. RETTIG, E. CONTI, J. D. PALMER, J. R. MANHART, K. J. SYTSMA, H. J. MICHAELS, W. J. KRESS, K. G. KAROL, W. D. CLARK, M. HEDRÉN, B. S. GAUT, R. K. JANSEN, K.-J. KIM, C. F. WIMPEE, J. F. SMITH, G. R. FURNIER, S. H. STRAUSS, Q.-Y. XIANG, G. M. PLUNKETT, P. S. SOLTIS, S. M. SWENSEN, S. E. WILLIAMS, P. A. GADEK, C. J. QUINN, L. E. EGUIARTE, E. GOLENBURG, G. H. LEARN JR., S. W. GRAHAM, S. C. H. BARRETT, S. DAYANANDAN, AND V. A. ALBERT. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. 1993. *Ann. Missouri Bot. Gard.* **80**: 528–580.

———, ———, P. S. SOLTIS, P. J. RUDALL, M. F. FAY, W. H. HAHN, S. SULLIVAN, J. JOSEPH, M. MOLVRAY, P. J. KORES, T. J. GIVNISH, K. J. SYTSMA, AND J. C. PIRES. 2000. Higher-level systematics of the monocotyledons: an assessment of current knowledge and a new classification, pp. 3–16. *In* K. L. Wilson and D. A. Morrison, [eds.], Monocots: systematics and evolution. CSIRO Publishing, Collingwood, Victoria, Australia.

———, D. W. STEVENSON, P. WILKIN, AND P. J. RUDALL. 1995*b*. Monocot systematics: a combined analysis, pp. 685–730. *In* P. J. Rudall, P. J. Cribb, D. F. Cutler, and C. J. Humphries [eds.], Monocotyledons: systematics and evolution. Royal Botanic Gardens, Kew, Richmond, Surrey, UK.

DAHLGREN, R. M. T., H. T. CLIFFORD, AND P. F. YEO. 1985. The families of the monocotyledons: structure, evolution and taxonomy. Springer-Verlag, Berlin, Germany. 520 p.

DAVIS, J. I., D. W. STEVENSON, G. PETERSEN, O. SEBERG, L. M. CAMPBELL, J. V. FREUDENSTEIN, D. H. GOLDMAN, C. R. HARDY, F. A. MICHELANGELI, M. P. SIMMONS, C. D. SPRECHT, F. VERGARA-SILVA, AND M. GANDOLFO. 2004. A phylogeny of the monocots, as inferred from *rbc*L and *atp*A sequence variation, and a comparison of methods for calculating jackknife and bootstrap values. *Syst. Bot.* **29**: 467–510.

DUVALL, M. R., M. T. CLEGG, M. W. CHASE, W. D. CLARK, W. J. KRESS, H. G. HILLS, L. E. EGUIARTE, J. F. SMITH, B. S. GAUT, E. A. ZIMMER, AND G. H. LEARN, JR. 1993. Phylogenetic hypotheses for the monocotyledons constructed from *rbc*L sequence data. *Ann. Missouri Bot. Gard.* **80**: 607–619.

———, AND A. B. ERVIN. 2004. 18S gene trees are positively misleading for monocot/dicot phylogenetics. *Molec. Phylogen. Evol.* **30**: 97–106.

———, S. MATHEWS, N. MOHAMMAD, AND T. RUSSELL. 2005. Placing the monocots: conflicting signal from trigenomic analyses, pp. xx–xx. *In* J. T. Columbus, E. A. Friar, J. M. Porter, L. M. Prince, and M. G. Simpson [eds.], Monocots: comparative biology and evolution, 2 vols. Rancho Santa Ana Botanic Garden, Claremont, California, USA.

ERIXON P., B. SVENNBLAD, T. BRITTON, AND B. OXELMAN. 2003. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Syst. Biol.* **52**: 665–673.

FAY, M. F., P. J. RUDALL, S. SULLIVAN, K. L. STOBART, A. Y. DE BRUIJN, G. REEVES, F. QAMARUZ-ZAMAN, W. P. HONG, J. JOSEPH, W. J. HAHN, J. G. CONRAN, AND M. W. CHASE. 2000. Phylogenetic studies of Asparagales based on four plastid DNA regions, pp. 360–371. *In* K. L. Wilson and D. A. Morrison, [eds.], Monocots: systematics and evolution. CSIRO Publishing, Collingwood, Victoria, Australia.

FELSENSTEIN, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27**: 401–410.

———. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783–791.

———. 2004. Inferring phylogenies. Sinauer Associates, Inc., Sunderland, Massachusetts, USA. 664 p.

———, AND H. KISHINO. 1993. Is there something wrong with the bootstrap on phylogenies? A reply to Hillis and Bull. *Syst. Biol.* **42**: 193–200.

GAUT, B. S., B. R. MORTON, B. C. MCCAIG, AND M. T. CLEGG. 1996. Synonymous rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate difference at the plastid gene *rbcL. Proc. Natl. Acad. Sci. U.S.A.* **93**: 10274–10279.

———, S. V. MUSE, W. D. CLARK, AND M. T. CLEGG. 1992. Relative rates of nucleotide substitution at the *rbc*L locus of monocotyledonous plants. *J. Molec. Evol.* **35**: 292–303.

GIVNISH, T. J., J. C. PIRES, S. W. GRAHAM, M. A. MCPHERSON, L. M. PRINCE, T. B. PATTERSON, H. S. RAI, E. H. ROALSON, T. M. EVANS, W. J. HAHN, K. C. MILLAM, A. W. MEEROW, M. MOLVRAY, P. J. KORES, H. E. O'BRIEN, J. C. HALL, W. J. KRESS, AND K. J. SYTSMA.

2005. Phylogeny of the monocots based on the highly informative cpDNA gene *ndh*F: evidence for widespread concerted convergence, pp. xx–xx. *In* J. T. Columbus, E. A. Friar, J. M. Porter, L. M. Prince, and M. G. Simpson [eds.], Monocots: comparative biology and evolution, 2 vols. Rancho Santa Ana Botanic Garden, Claremont, California, USA.

GOREMYKIN V. V., K. I. HIRSCH-ERNST, S. WÖLFL, AND F. H. HELLWIG. 2003. Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. *Molec. Biol. Evol.* **20**: 1499–1505.

———, ———, ———, AND ———. 2004. The chloroplast genome of *Nymphaea alba*: whole genome analyses and the problem of identifying the most basal angiosperm. *Molec. Biol. Evol.* **21**: 1445–1454.

GRAHAM, S. W., AND S. C. H. BARRETT. 2004. Phylogenetic reconstruction of the evolution of stylar polymorphisms in *Narcissus* (Amaryllidaceae). *Amer. J. Bot.* **91**: 108–122.

———, J. R. KOHN, B. R. MORTON, J. E. ECKENWALDER, AND S. C. H. BARRETT. 1998. Phylogenetic congruence and discordance among one morphological and three molecular data sets from Pontederiaceae. *Syst. Biol.* **47**: 545–567.

———, AND R. G. OLMSTEAD. 2000. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *Amer. J. Bot.* **87**: 1712–1730.

———, P. A. REEVES, A. C. E. BURNS, AND R. G. OLMSTEAD. 2000. Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. *Int. J. Pl. Sci.* **161**: S83–S96.

HENDY, M. D., AND D. PENNY. 1989. A framework for the quantitative study of evolutionary trees. *Syst. Zool.* **38**: 297–309.

HILLIS, D. M. 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Syst. Biol.* **47**: 3–8.

———, AND J. BULL. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* **42**: 182–192.

———, D. P. POLLOCK, J. A. MCGUIRE, AND D. J. ZWICKL. 2003. Is sparse taxon sampling a problem for phylogenetic inference? *Syst. Biol.* **52**: 124–126.

HILU, K. W., T. BORSCH, K. MÜLLER, D. E. SOLTIS, P. S. SOLTIS, V. SAVOLAINEN, M. W. CHASE, M. P. POWELL, L. A. ALICE, R. EVANS, H. SAUQUET, C. NEINHUIS, T. A. B. SLOTTA, J. G. ROHWER, C. S. CAMPBELL, AND L. W. CHATROU. 2003. Angiosperm phylogeny based on *mat*K sequence information. *Amer. J. Bot.* **90**: 1758–1776.

IMHOF, S. 1999. Subterranean structures and mycorrhiza of the achlorophyllous *Burmannia tenella* (Burmanniaceae). *Canad. J. Bot.* **77**: 637–643.

KIM, K.-J., AND R. K. JANSEN. 1995. *ndh*F sequence evolution and the major clades in the sunflower family. *Proc. Natl. Acad. Sci. U.S.A.* **92**: 10379–10383.

KIMURA, M. 1968. Evolutionary rate at the molecular level. *Nature* **217**: 624–626.

———. 1983. The neutral theory of molecular evolution. Cambridge University Press, New York, USA. 353 p.

KOLACZKOWSKI, B., AND J. W. THORNTON. 2004. Performance of maximum parsimony and likelihood phylogenetic when evolution is heterogeneous. *Nature* **431**: 980–984.

LEMMON, A. R., AND E. C. MORIARTY. 2004. The importance of proper model assumption in Bayesian phylogenetics. *Syst. Biol.* **53**: 265–277.

MADDISON, W. 1997. Gene trees in species trees. *Syst. Biol.* **46**: 523–536.

MATHEWS, S., AND M. J. DONOGHUE. 1999. The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science* **286**: 947–950.

MCPHERSON, M. A., M. F. FAY, M. W. CHASE, AND S. W. GRAHAM.

2004. Parallel loss of a slowly evolving intron from two closely related families in Asparagales. *Syst. Bot.* **29**: 296–307.

MEEROW, A. W., M. F. FAY, C. L GUY, Q.-B. LI, F. Q. ZAMAN, AND M. W. CHASE. 1999. Systematics of Amaryllidaceae based on cladistic analysis of plastid sequence data. *Amer. J. Bot.* **86**: 1325–1345.

NAYLOR, G. P., AND W. M. BROWN. 1998. Amphioxus mitochondrial DNA, chordate phylogeny, and the limits of inferences based on comparisons of sequences. *Syst. Biol.* **47**: 61–76.

OHTA, T. 1992. The nearly neutral theory of molecular evolution. *Ann. Rev. Ecol. Syst.* **23**: 263–286.

OLMSTEAD, R. G., AND J. A. SWEERE. 1994. Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. *Syst. Biol.* **43**: 467–481.

PARKINSON, C. L., K. L. ADAMS, AND J. D. PALMER. 1999. Multigene analyses identify the three earliest lineages of extant flowering plants. *Curr. Biol.* **9**: 1485–1488.

PENNY, D., L. R. FOULDS, AND M. D. HENDY. 1982. Testing the theory of evolution by comparing phylogenetic trees constructed from five different protein sequences. *Nature* **297**: 197–200.

PIRES, J. C., I. J. MAUREIRA, T. J. GIVNISH, K. J. SYTSMA, O. SEBERG, G. PETERSEN, J. I DAVIS, D. W. STEVENSON, P. J. RUDALL, M. F. FAY, AND M. W. CHASE. 2005. Phylogeny, genome size, and chromosome evolution of Asparagales, pp. xx–xx. *In* J. T. Columbus, E. A. Friar, J. M. Porter, L. M. Prince, and M. G. Simpson [eds.], Monocots: comparative biology and evolution, 2 vols. Rancho Santa Ana Botanic Garden, Claremont, California, USA.

POE, S., AND D. L. SWOFFORD. 1999. Taxon sampling revisited. *Nature* **398**: 299–300.

POLLOCK, D. D., D. J. ZWICKL, J. A. MCGUIRE, AND D. M. HILLIS. 2002. Increased taxon sampling is advantageous for phylogenetic inference. *Syst. Biol.* **51**: 664–671.

QIU, Y.-L., J. LEE, F. BERNASCONI-QUADRONI, D. E. SOLTIS, P. S. SOLTIS, M. ZANIS, E. A. ZIMMER, Z. CHEN, V. SAVOLAINEN, AND M. W. CHASE. 1999. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* **402**: 404–407.

RAI, H. S., H. E. O'BRIEN, P. A. REEVES, R. G. OLMSTEAD, AND S. W. GRAHAM. 2003. Inference of higher-order relationships in the cycads from a large chloroplast data set. *Molec. Phylogen. Evol.* **29**: 350–359.

RANNALA, B., J. P. HUELSENBECK, Z. YANG, AND R. NIELSEN. 1998. Taxon sampling and the accuracy of large phylogenies. *Syst. Biol.* **47**: 702–710.

RUDALL, P. J., K. L. STOBART, W.-P. HONG, C. A. FURNESS, G. C. KITE, AND M. W. CHASE. 2000. Consider the lilies: systematics of Liliales, pp. 347–357. *In* K. L. Wilson and D. A. Morrison, [eds.], Monocots: systematics and evolution. CSIRO Publishing, Collingwood, Victoria, Australia.

SAARELA, J. M., P. M. PETERSON, R. M. KEANE, J. CAYOUETTE, AND S. W. GRAHAM. 2005. Molecular phylogenetics of *Bromus* (Poaceae: Pooideae) based on chloroplast and nuclear DNA sequence data, pp. xx–xx. *In* J. T. Columbus, E. A. Friar, J. M. Porter, L. M. Prince, and M. G. Simpson [eds.], Monocots: comparative biology and evolution, 2 vols. Rancho Santa Ana Botanic Garden, Claremont, California, USA.

SANDERSON, M. J., M. F. WOJCIECHOWSKI, J.-M. HU, T. SHER KHAN, AND S. G. BRADY. 2000. Error, bias and long-branch attraction in data for two chloroplast genes in seed plants. *Molec. Biol. Evol.* **17**: 783–797.

SAVOLAINEN, V., M. W. CHASE, N. SALAMIN, D. E. SOLTIS, P. S. SOLTIS, A. J. LÓPEZ, O. FÉDRIGO, AND G. J. P. NAYLOR. 2002. Phylogeny reconstruction and functional constraints in organellar genomes: plastid *atp*B and *rbc*L sequences versus animal mitochondrion. *Syst. Biol.* **51**: 638–647.

SAS INSTITUTE. 2000. JMP vers. 4.0.1. Statistical Discovery Software. Cary, North Carolina, USA.

SHARP, P. M. 1991. Determinants of DNA sequence divergence be-

tween *Escherichia coli* and *Salmonella typhimurium*: codon usage, map position, and concerted evolution. *J. Molec. Evol.* **33**: 23–33.

SHAW, J., E. B. LICKEY, J. T BECK, S. B. FARMER, W. LIU, J. MILLER, K. C. SIRIPUN, C. T. WINDER, E. E. SCHILLING, AND R. L. SMALL. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Amer. J. Bot.* **92**: 142–166.

SOLTIS D. E., AND P. S. SOLTIS. 2004. *Amborella* not a "basal angiosperm"? Not so fast. *Amer. J. Bot.* **91**: 997–1001.

————, ————, M. W. CHASE, M. E. MORT, D. C. ALBACH, M. ZANIS, V. SAVOLAINEN, W. H. HAHN, S. B. HOOT, M. F. FAY, M. AXTELL, S. M. SWENSEN, L. M. PRINCE, W. J. KRESS, K. C. NIXON, AND J. S. FARRIS. 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbc*L, and *atp*B sequences. *Bot. J. Linn. Soc.* **133**: 381–461.

————, ————, M. E. MORT, M. W. CHASE, V. SAVOLAINEN, S. B. HOOT, AND C. M. MORTON. 1998. Inferring complex phylogenies using parsimony: an empirical approach using three large DNA data sets for angiosperms. *Syst. Biol.* **47**: 32–42.

SOLTIS, P. S., D. E. SOLTIS, AND M. W. CHASE. 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* **402**: 402–403.

STEANE, D. A., R. W. SCOTLAND, D. J. MABBERLEY, AND R. G. OLMSTEAD. 1999. Molecular systematics of *Clerodendrum* (Lamiaceae): ITS sequences and total evidence. *Amer. J. Bot.* **86**: 98–107.

STEFANOVIC, S., D. W. RICE, AND J. D. PALMER. 2004. Long branch attraction, taxon sampling, and the earliest angiosperms: *Amborella* or monocots? *B. M. C. Evol. Biol.* **4**:35.

STEVENSON, D. W., J. I DAVIS, J. V. FREUDENSTEIN, C. R. HARDY, M.

P. SIMMONS, AND C. D. SPECHT. 2000. A phylogenetic analysis of monocotyledons based on morphological and molecular character sets, with comments on the placement of *Acorus* and Hydatellaceae, pp. 17–24. *In* K. L. Wilson and D. A. Morrison, [eds.], Monocots: systematics and evolution. CSIRO Publishing, Collingwood, Victoria, Australia.

SUZUKI, Y., G. V. GLAZKO, AND M. NEI. 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proc. Natl. Acad. Sci. U.S.A.* **99**: 16138–16143.

SWOFFORD, D. L. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods), vers. 4.0b10. Sinauer Associates, Inc., Sunderland, Massachusetts, USA.

————, P. J. WADELL, J. P. HUELSENBECK, P. G. FOSTER, P. O. LEWIS, AND J. S. ROGERS. 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. *Syst. Biol.* **50**: 525–539.

TAMURA, M. N., J. YAMASHITA, S. FUSE, AND M. HARAGUSCHI. 2004. Molecular phylogeny of the monocotyledons inferred from combined analysis of plastid *mat*K and *rbc*L gene sequences. *J. Pl. Res.* **117**: 109–120.

WILSON, M. A., B. GAUT, AND M. T. CLEGG. 1990. Chloroplast DNA evolves slowly in the palm family (Arecaceae). *Molec. Biol. Evol.* **7**: 303–314.

WOLFE, K. H., AND P. M. SHARP. 1993. Mammalian gene evolution: nucleotide sequence divergence between mouse and rat. *J. Molec. Evol.* **37**:441–456.

ZANIS, M. J., D. E. SOLTIS, P. S. SOLTIS, S. MATHEWS, AND M. J. DONOGHUE. 2002. The root of angiosperms revisited. *Proc. Natl. Acad. Sci. U.S.A.* **99**: 6848–6853.

ZWICKL, D. J., AND D. M. HILLIS. 2002. Increased taxon sampling greatly reduces phylogenetic error. *Syst. Biol.* **51**: 588–598.

VOLUME 22, NUMBER 1   Deep Monocot Phylogeny   17

Appendix 1. GenBank accession numbers and vouchers for exemplar monocot taxa. Alternative source details (e.g., living collection accession numbers) are given in place of voucher information in several cases. GenBank numbers of sequences generated by workers outside the Graham laboratory are underlined; several of these were generated from the same or closely related species (the latter are noted; see relevant GenBank accessions for further details). The three grasses (Poaceae) included here (*Oryza sativa* L., *Triticum aestivum* L. and *Zea mays* L.) are from whole plastid genomes deposited on GenBank (accessions NC_001320, NC_002762 and NC_001666.2, respectively). Further details on the regions used are presented in Graham and Olmstead (2000) and McPherson et al. (submitted).

Taxon (Voucher [herbarium])	*atp*B	*ndh*F	*psb*B, T, N, & *psb*H
Acorales			
Acorus calamus L. (Denver Botanic Gardens, Colorado; *Olmstead 97-149* [DNA])	<u>AJ235381</u>	AY007647	AF123843
Alismatales			
Butomus umbellatus L. (*Chase 6414* [K])	AY147593	AF546997	AY147499
Sagittaria latifolia Willd. (*Barrett s.n.* [TRT])	AF239788	AY007657	AY007469
Scheuchzeria palustris L. (*Waterway & Graham 97-60* [ALTA])	AY147594	AF547007	AY147500
Spathiphyllum wallisii Regel (*Chase 201* [NCU])	<u>AJ235606</u>	AY007658	AY007471
Tofieldia glutinosa (Michx.) Pers. (*Stefanovic s.n.* [UBC])	AY147595	AF547023[1]	AY147501
Arecales			
Roystonea princeps (Becc.) Burret (*Santiago J-4* [UPR])	AY147608	AY147772	AY147515
Asparagales			
Agapanthus africanus (L.) Hoffmanns (*McPherson 020420-1* [ALTA])	AY465542	AY465647	AY465568
Alania endlicheri Kunth (*Vitt 27706* [ALTA])	AY147612	AY147773	AY147519
Allium textile A. Nelson & J. F. MacBr. (*McPherson 990704-79* [ALTA]) and *A. senescens* Pall.	AY147628 —	AF547000 —	AY147536 —
Amerorchis rotundifolia (Banks ex Pursh) Hultén (*McPherson 010610-1* [ALTA])	AY147623	AY147783	AY147531
Aphyllanthes monspeliensis L. (*Graham 00-04-08* [ALTA])	AY147629	AY147787	AY147537
Asparagus officinalis L. (*McPherson 010819-2* [ALTA])	AY147630	AY147788	AY147538
Asphodelus albus Mill. (*Harder 1-000430* [ALTA]) and *A. aestivus* Brot.	AY147613 —	AY147774 —	AY147520 —
Astelia alpina Banks & Sol. ex R. Br. (*Chase 1103* [K])	AY147614	AY147775	AY147521
Blandfordia punicea (Labill.) Sweet (*Chase 519* [K]) and *B. nobilis* Sm.	AY147615 —	AY147776 —	AY147522 —
Chlorophytum comosum (Thunb.) Jacques (*McPherson 000321-1* [ALTA]) and *Anthericum ramosum* L.	AY147631 —	AY147789 —	AY147539 —
Coelogyne cristata Lindl. (*McPherson 010921-1* [ALTA]) and *C. macdonaldii* F. Muell. & Kraenzl.	AY147616 —	AY147777 —	AY147523 —
Curculigo capitulata (Lour.) Kuntze (*Chase 205* [NCU]) and *Hypoxis villosa* L. f.	AY147617 —	AY147778 —	AY147524 —
Cyanastrum cordifolium Oliv. (*Graham & Barrett 2* [TRT]) and *Kabuyea hostifolia* (Engl.) Brummit	<u>AF168902</u> —	U79228 —	AY147525 —
Cypripedium passerinum Richardson (*McPherson 010722-6* [ALTA]) and *C. irapeanum* La Llave & Lex.	AY147618 —	AY147779 —	AY147526 —
Doryanthes palmeri W. Hill ex Benth. (*Chase 2837* [K])	AY465543	AY465648	AY465569
Hemerocallis littorea Makino (*Chase 3833* [K])	AY147619	AY147780	AY147527
Iris missouriensis Nutt. (*McPherson 000707-5a-7* [ALTA]) and *I. unguicularis* Poir.	AY147620 —	AF547003 —	AY147528 —
Ixiolirion tataricum (Pall.) Herb. (*Chase 489* [K])	AY147621	AY147781	AY147529
Lanaria lanata (L.) Druce (*Goldblatt 9410* [MO])	AY147622	AY147782	AY147530
Lomandra longifolia Labill. (*Vitt 27411* [ALTA]) and *Thysanotus spiniger* Brittan	AY147632 —	AF547004 —	AY147540 —
Maianthemum racemosum (L.) Link (*McPherson 990704-97* [ALTA]) and *M. bifolium* (L.) F.W. Schmidt	AY147633 —	AF547005 —	AY147541 —
Muilla maritima (Torr.) S. Watson (*Pires 98-028* [WIS])	AY147634	AY147790	AY147542
Muscari comosum (L.) Mill. (*Harder 000419-1* [ALTA])	AY147635	AF547006	AY147543
Narcissus elegans (Haw.) Spach (*Barrett 1434* [TRT])	AY147636	U79216	AY147544
Phormium tenax J.R. Forst. & G. Forst. (*McPherson 000612-3* [ALTA])	AY147624	AY147784	AY147532
Sisyrinchium montanum Greene (*McPherson 990704-71* [ALTA]) and *S. micranthum* Cav.	AY147625 —	AF547008 —	AY147533 —
Xanthorrhoea resinosa Pers. (*Chase 192* [NCU])	AY147626	AY147785	AY147534
Xeronema callistemon W. R. B. Oliv. (*Chase 653* [K])	AY147627	AY147786	AY147535
Yucca glauca Nutt. (Provided by D. Hurlburt, University of Alberta; field collection at Onefour, Alberta; *Graham 00121* [DNA]) and *Agave celsii* Hook.	AY147637 —	AF547014 —	AY147545 —

18   Graham et al.   ALISO

Appendix 1.   Extended.

	Gene or region (continued)						Length (unaligned)	
psbD & psbC	psbE, F, L & psbJ	rbcL	rpl2	3'-rps12, rps7, ndhB, trnL (CAA)	atpB–rbcL spacer region	trnL (UAA)–trnF (GAA)	All nucleotides	Protein-coding nucleotides
AF123813	AF123828	D28865	AF123785	AF123771	—	—	13.8 kb	11.0 kb
AY147638	AY147546	AY149345	AY147685	AY147450	—	—	14.8 kb	11.2 kb
AF239789	AY007484	L08767	AY007497	AF238074	—	—	14.1 kb	11.2 kb
AY147639	AY147547	U03728	AY147686	AY147451	—	—	14.8 kb	11.2 kb
AF239794	AY007487	AJ235807	AY007500	AF238077	—	—	14.2 kb	11.4 kb
AY147640	AY147548	AY149346	AY147687	AY147452	—	—	14.9 kb	11.3 kb
AY147654	AY147562	AY149357	AY147701	AY147466	AY147735	—	16.4 kb	11.5 kb
AY465672	AY465595	AY465699	AY465724	AY465623	AY699127	AY699224	16.6 kb	11.3 kb
AY147658	AY147566	Y14982	AY147705	AY147471	AY147737	Fay et al. (2000)[2]	16.6 kb	11.5 kb
AY147675	AY147583	AY149372	AY147721	AY147489	AY147753	—	17.0 kb	11.3 kb
—	—	—	—	—	—	Fay et al. (2000)[2]		
AY147670	AY147578	AY149368	AY147716	AY147484	AY147748	—	16.0 kb	11.2 kb
AY147676	AY147584	AY149373	AY147722	AY147490	AY147754	Fay et al. (2000)[2]	17.2 kb	11.3 kb
AY147677	AY147585	AY149374	AY147723	AY147491	AY147755	Fay et al. (2000)[2]	17.2 kb	11.3 kb
AY147659	AY147567	AY149360	AY147706	AY147472	AY147738	—	16.8 kb	11.3 kb
—	—	—	—	—	—	AJ290257/AJ290291		
AY147660	AY147568	Z77261	AY147707	AY147473	AY147739	Fay et al. (2000)[2]	16.7 kb	10.7 kb
AY147661	AY147569	Z73694	AY147708	AY147474	AY147740	—	17.3 kb	11.5 kb
—	—	—	—	—	—	AJ232441/AJ232564		
AY147678	AY147586	AY149375	AY147724	AY147492	AY147756	—	16.8 kb	11.3 kb
—	—	—	—	—	—	AJ232445/AJ232568		
AY147662	AY147570	AY149361	AY147709	AY147475	AY147741	—	17.9 kb	11.3 kb
—	—	—	—	—	—	AF463396/AF463381		
AY147663	AY147571	AY149362	AY147710	AY147476	AY147742	—	16.7 kb	11.3 kb
—	—	—	—	—	—	X74579		
AY147664	AY147572	U41572	AY147711	AY147477	AY147743	—	17.1 kb	11.3 kb
—	—	—	—	—	—	AJ290312/AJ290278		
AY147665	AY147573	AY149363	AY147712	AY147478/ AY147479	—	—	14.7 kb	10.6 kb
—	—	—	—	—	—	Fay et al. (2000)[2]		
AY465673	AY465596	AY465700	AY465725	AY465624	AY699128	AY699160/AY699161	16.3 kb	11.3 kb
AY147666	AY147574	AY149364	AY147713	AY147480	AY147744	AJ290308/AJ290274	16.1 kb	11.3 kb
AY147667	AY147575	AY149365	AY147714	AY147481	AY147745	—	16.8 kb	11.3 kb
—	—	—	—	—	—	AJ409609		
AY147668	AY147576	AY149366	—	AY147482	AY147746	AJ290280/AJ290314	15.0 kb	9.8 kb
AY147669	AY147577	AY149367	AY147715	AY147483	AY147747	—	16.2 kb	11.3 kb
AY147679	AY147587	L05039	AY147725	AY147493	AY147757	—	17.2 kb	11.4 kb
—	—	—	—	—	—	Fay et al. (2000)[2]		
AY147680	AY147588	AY149376	AY147726	AY147494	AY147758	—	17.2 kb	11.3 kb
—	—	—	—	—	—	AJ441175		
AY147681	AY147589	AY149377	AY147727	AY147495	AY147759	AF117019/AF117047	17.2 kb	11.3 kb
AY147682	AY147590	AY149378	AY147728	AY147496	AY147760	AJ232546/AJ232669	17.1 kb	11.3 kb
AY147683	AY147591	AY149379	AY147729	AY147497	AY147761	AY357142	17.3 kb	11.3 kb
AY147671	AY147579	Z69232	AY147717	AY147485	AY147749	AY828229	17.2 kb	11.5 kb
AY147672	AY147580	AY149369	AY147718	AY147486	AY147750	—	16.7 kb	11.3 kb
—	—	—	—	—	—	AJ409603		
AY147673	AY147581	AY149370	AY147719	AY147487	AY147751	AJ290271 (in part) and Fay et al. (2000)[2]	17.3 kb	11.3 kb
AY147674	AY147582	AY149371	AY147720	AY147488	AY147752	Fay et al. (2000)[2]	17.2 kb	11.3 kb
AY147684	AY147592	AY149380	AY147730	AY147498	AY147762	—	17.3 kb	11.3 kb
—	—	—	—	—	—	AF508509		

Appendix 1.   Continued.

Taxon (Voucher [herbarium])	Gene or region		
	*atp*B	*ndh*F	*psb*B, T, N, & *psb*H
Commelinales			
Hydrothrix gardneri Hook f. (*Barrett 1414* [TRT])	AY147604	U41606	AY147511
Philydrum lanuginosum Banks & Sol. ex Gaertn. (*Graham & Barrett 1* [TRT])	AY147607	U41622	AY147514
Xiphidium caeruleum Aubl. (Provided by M. Simpson, California State University at San Diego; *Graham 5.7.94* [DNA])	AY147611	AF547013	AY147518
Dioscoreales			
Burmannia capitata (Walter ex J.F. Gmel.) Mart. (*Neyland 958* [MCN])	AY147596	—	AY147502
Dioscorea bulbifera L. (E.P.O. Biology, University of Colorado, Boulder; *Olmstead 97-151* [DNA])	AF187059	AY007652	AF123849
Liliales			
Alstroemeria aurea Graham (*Sheahan s.n.* [K])	AY465546	AY465651	AY465572
Anticlea elegans (Pursh) Rydb. (*McPherson 990704-68* [ALTA])	AY147600	AY147765	AY147506
Lilium superbum L. (*Chase 112* [NCU])	AY116649	AY007655	AY007465
Petermannia cirrosa F. Muell. (*Frederiksen et al. s.n.* [C])	AY465558	AY465662	AY465585/ AY465743
Smilax rotundifolia L. (*Uhl 92-07* [BH])	AY465554	AY465659	AY465581/ AY465746
Trillium grandiflorum (Michx.) Salisb. (*Eades s.n.* [ALTA])	AY465556	AY465661	AY465583
Wurmbea pygmaea (Endl.) Benth. (unpubl. data.) (*Case 77* [PERTH])	AY465557	AF547012	AY465585
Pandanales			
Carludovica drudei Mast. (Bailey Hortorium 73:574 [BH])	AY465545	AY465650	AY465571
Pandanus copelandii Merr. (*Sherman & Bynum 303* [MO])	AY465544	AY465649	AY465570
Stemona tuberosa Lour. (*Rothwell & Stockey 46* [ALTA])	AY147599	AF547009	AY147505
Talbotia elegans Balf. (*Rothwell & Stockey 48* [ALTA])	AY147609	AF547011	AY147516
Poales			
Ananas comosus (L.) Merr. (*Rai 1003* [ALTA])	AY147601	AY147766	AY147507
Cyperus papyrus L. (*Alan Yen 174* [private collection])	AY465534	AY465642	AY465560
Ecdeiocolea monostachya F. Muell. (*Hopper 8531* [KPBG])	AY465535	AY438617	AY465561
Elegia fenestrata Pillans [New York Botanical Garden accession 1697/95]	AY465536	AY547016	AY465562
Flagellaria indica L. (Bailey Hortorium 77:394 [BH])	AY465537	AY465643	AY465563
Mayaca fluviatilis Aubl. (*Abbott 10526* [FLAS])	AY147605	AY147770	AY147512
Sparganium eurycarpum Engelm. (*Hansen s.n.* [BH])	AY465539	AY465645	AY465565
Typha angustifolia L. (*Graham 1040* [ALTA])	—	U79230	AY147517
Typha latifolia L. (*McPherson 010819-3* [ALTA])	AY147610	—	—
and *Xyris jupicai* Rich. (*Goldman 1766* [BH])	AY465541	AF547017	AY465567
Zingiberales			
Ensete ventricosum (Welw.) Cheesman (*Kress 96-5372* [US])	<u>AF168910</u>	AY147769	AY147510
Strelitzia reginae Aiton (*O'Brien s.n.* [ALTA])	AY465540	AY465646	AY465566
Uncertain Ordinal Placement			
Dasypogonaceae			
Dasypogon hookeri J.R. Drumm. (*Chase 430* [NCU] & *Conran et al. 917* [PERTH])	AY147603	AY147768	AY147509
Kingia australis R. Br. (*Conran et al. 922* [PERTH])	AY465538	AY465644	AY465564
Petrosaviaceae			
Japonolirion osense Nakai (*Chase 3000* [K])	AY147598	AY147764	AY147504

[1] We generated an *ndh*F sequence here using different source material: *Tofieldia glutinosa* (*Morton & Venn 9282* [ALTA]).

[2] Sequences presented in Fay et al. (2000) but not yet available on GenBank.

20   Graham et al.   ALISO

Appendix 1.   Extended. Continued.

				3'-rps12, rps7, ndhB, trnL (CAA)	atpB–rbcL spacer region	trnL (UAA)–trnF (GAA)	Length (unaligned)	
psbD & psbC	psbE, F, L & psbJ	rbcL	rpl2				All nucleotides	Protein-coding nucleotides
AY147650	AY147558	U41582	AY147697	AY147462	—	—	14.8 kb	11.2 kb
AY147653	AY147561	U41596	AY147700	AY147465	AY147734	—	16.3 kb	11.3 kb
AY147657	AY147565	AY149359	AY147704	AY147470	—	—	14.8 kb	11.1 kb
AY147641	AY147549	AY149347	AY147688	AY147453	—	—	12.7 kb	9.9 kb
AF123819	AF123834	D28327	AF123791	AF123777	—	—	13.9 kb	11.0 kb
AY465676	AY465599	AY465703	AY465728	AY465627	AY699131	AY699225	16.8 kb	11.4 kb
AY147645	AY147553	AY149351	AY147692	AY147457	AY699130	AY699168/AY699169	16.3 kb	11.2 kb
AF239783	AY007480	L12682	AY007493	AF238070	AY699129	Fay et al. (2000)[2]	15.6 kb	11.4 kb
AY465689	AY465612	AY465714	AY465741	AY465640	AY699144	—	13.2 kb	9.4 kb
AY465685	AY465608	AY465710	AY465737	AY465636	AY699140	AY699170/AY699171	16.6 kb	11.2 kb
AY465687	AY465610	AY465712	AY465739	AY465638	AY699142	AY699166/AY699167	16.5 kb	11.3 kb
AY465688	AY465611	AY465713	AY465740	AY465639	AY699143	A.L. Case, Duke U. (unpubl. data)	16.6 kb	11.3 kb
AY465675	AY465598	AY465702	AY465727	AY465626	—	—	15.5 kb	11.3 kb
AY465674	AY465597	AY465701	AY465726	AY465625	—	—	15.0 kb	11.4 kb
AY147644	AY147552	AY149350	AY147691	AY147456	—	—	15.4 kb	11.3 kb
AY147655	AY147563	AY149358	AY147702	AY147467	—	—	14.8 kb	11.1 kb
AY147646	AY147554	L19977	AY147693	AY147458	AY147731	—	16.5 kb	11.5 kb
AY465664	AY465587	Y12966	AY465717	AY465615	—	—	14.8 kb	11.2 kb
AY465665	AY465588	AY465692	AY465718	AY465616	—	—	15.4 kb	11.1 kb
AY465666	AY465589	AY465693	AY465719	AY465617	—	—	15.4 kb	11.4 kb
AY465667	AY465590	AY465694	AY465720	AY465618	—	—	15.4 kb	11.2 kb
AY147651	AY147559	AY149355	AY147698	AY147463	—	—	13.8 kb	10.5 kb
AY465669	AY465592	AY465696	AY465721	AY465620	—	—	15.6 kb	11.4 kb
AY147656	AY147564	—	—	AY147468	—	—		
—	—	M91634	AY147703	AY147469	AY147736	—	16.4 kb	11.4 kb
AY465671	AY465594	AY465698	AY465723	AY465622	—	—	15.3 kb	11.2 kb
AY147649	AY147557	AY149354	AY147696	AY147461	AY147733	—	16.4 kb	11.3 kb
AY465670	AY465593	AY465697	AY465722	AY465621	—	—	13.9 kb	11.1 kb
AY147648	AY147556	AY149353	AY147695	AY147460	AY147732	—	16.3 kb	11.3 kb
AY465668	AY465591	AY465695	—	AY465619	—	—	13.9 kb	10.4 kb
AY147643	AY147551	AY149349	AY147690	AY147455	—	—	14.1 kb	11.1 kb