Article Type: Research Article

RESEARCH ARTICLE

Short Title: Givnish et al.—Monocot plastid phylogenomics

Monocot plastid phylogenomics, timeline, net rates of species diversification, the power of multi-gene analyses, and a functional model for the origin of monocots

Thomas J. Givnish[1,16], Alejandro Zuluaga[2], Daniel Spalink[3], Marybel Soto Gomez[4], Vivienne K. Y. Lam[4], Jeffrey M. Saarela[5], Chodon Sass[6], William J. D. Iles[7], Danilo José Lima deSousa[8], James Leebens-Mack[9], J. Chris Pires[10], Wendy B. Zomlefer[9], Maria A. Gandolfo[11], Jerrold I. Davis[11], Dennis W. Stevenson[12], Claude dePamphilis[13], Chelsea D. Specht[11], Sean W. Graham[4], Craig F. Barrett[14], and Cécile Ané[1,15]

Manuscript received 14 May 2018; revision accepted 3 August 2018.

[1] Department of Botany, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA

[2] Departamento de Biologia, Universidad del Valle, Cali, Colombia

[3] Department of Ecosystem Science, Texas A&M University, College Station, Texas 77840, USA

[4] Department of Botany, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

[5] Canadian Museum of Nature, Ottawa, ON K1P 6P4, Canada

[6] The University and Jepson Herbarium, University of California–Berkeley, Berkeley, California 94720, USA

[7] Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, Michigan 48109, USA

[8] Departamento de Ciéncias Biológicas, Universidade Estadual de Feira de Santana, Feira de Santana, Bahia 44036-900, Brazil

[9] Department of Plant Biology, University of Georgia, Athens, Georgia 30602, USA
[10] Division of Biological Sciences, University of Missouri–Columbia, Columbia, Missouri 65211, USA
[11] School of Integrative Plant Sciences and L.H. Bailey Hortorium, Cornell University, Ithaca, New York 14853, USA
[12] New York Botanical Garden, New York, New York 10458, USA
[13] Department of Biology, Pennsylvania State University, University Park, Pennsylvania 16802, USA
[14] Department of Biology, West Virginia University, Morgantown, West Virginia 26506, USA
[15] Department of Statistics, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA
[16] Author for correspondence (e-mail: givnish@wisc.edu); ORCID iD 0000-0003-3166-4566
Citation: Givnish, T. J., A. Zuluaga, D. Spalink, M. S. Gomez, V. K. Y. Lam, J. M. Saarela, C. Sass, et al. 2018. Monocot plastid phylogenomics, timeline, net rates of species diversification, the power of multi-gene analyses, and a functional model for the origin of monocots. *American Journal of Botany* 105(11): XXXX.
DOI: XXXX

PREMISE OF THE STUDY: We present the first plastome phylogeny encompassing all 77 monocot families, estimate branch support, and infer monocot-wide divergence times and rates of species diversification.
METHODS: We conducted maximum likelihood analyses of phylogeny and BAMM studies of diversification rates based on 77 plastid genes across 545 monocots and 22 outgroups. We quantified how branch support and ascertainment vary with gene number, branch length, and branch depth.
KEY RESULTS: Phylogenomic analyses shift the placement of 16 families in relation to earlier studies based on four plastid genes, add seven families, date the divergence between monocots and eudicots+*Ceratophyllum* at 136 Mya, successfully place all mycoheterotrophic taxa examined, and support recognizing Taccaceae and Thismiaceae as separate families and Arecales and Dasypogonales as separate orders. Only 45% of interfamilial divergences occurred after the Cretaceous. Net species diversification underwent four large-scale accelerations in PACMAD-BOP Poaceae, Asparagales sister to Doryanthaceae, Orchidoideae-Epidendroideae, and Araceae

sister to Lemnoideae, each associated with specific ecological/morphological shifts. Branch ascertainment and support across monocots increase with gene number and branch length, and decrease with relative branch depth. Analysis of entire plastomes in Zingiberales quantifies the importance of non-coding regions in identifying and supporting short, deep branches. **CONCLUSIONS**: We provide the first resolved, well-supported monocot phylogeny and timeline spanning all families, and quantify the significant contribution of plastome-scale data to resolving short, deep branches. We outline a new functional model for the evolution of monocots and their diagnostic morphological traits from submersed aquatic ancestors, supported by convergent evolution of many of these traits in aquatic Hydatellaceae (Nymphaeales).

KEY WORDS: aquatic origin; chloroplast; divergence times; diversification; fossil calibration; molecular phylogeny; monocot syndrome; monocotyledons; mycoheterotrophy; Zingiberales.

Monocotyledons—with ~85,000 species (Lughadha et al., 2016) in 77 families and 11–12 orders (Angiosperm Phylogeny Group [APG], 2009, 2016; cf. Givnish et al., 1999)—are one of the most species-rich, ecologically dominant, and economically important of all land-plant lineages. Since they arose 136–140 Mya (Magallón et al., 2015; Smith and Brown, 2018), monocots have radiated into almost every terrestrial and aquatic habitat occupied by angiosperms, display extraordinary variation in morphology, comprise 21% of all angiosperm species, and directly and indirectly underpin most of the human diet and, thus, civilization. Understanding their relationships and patterns of morphological divergence, geographic diversification, and ecological radiation is therefore a critical challenge for biologists (Givnish et al., 2010).

Over the past 35 years, molecular systematics has greatly improved our understanding of monocot relationships. Analyses based on the sequences of one to few plastid genes (e.g., Chase et al., 1993, 2000, 2006; Duvall et al., 1993a, b; Givnish et al., 1999, 2005; Graham et al., 2006) overturned several conclusions regarding the placement of individual genera and families based on morphology, and led to a dramatic reclassification of the monocots at the familial and ordinal levels (e.g., APG, 2009). Yet these phylogenetic studies failed to resolve or strongly support relationships among several monocot orders and families. Even when these taxa were strongly supported—and often newly described or circumscribed to reflect that fact—relationships within them frequently were not. Uncertainty remains regarding the position of orders Arecales,

Commelinales, Dasypogonales (Givnish et al., 1999), Poales, and Zingiberales; of several families within Alismatales, Dioscoreales, Liliales, Pandanales, Poales, and Zingiberales; and of some tribes or subfamilies within the large families Poaceae, Orchidaceae, Araceae, and Asparagaceae (Chase et al., 2003, 2006; Graham et al., 2006; Cabrera et al., 2008; Grass Phylogeny Working Group II, 2012). The placement of non-photosynthetic, mycoheterotrophic lineages, which are especially well represented in monocots (Merckx et al., 2013), has also been a perennial challenge, given the relaxed selection on the presence (i.e., retention) and substitution rate of plastid genes (Lam et al., 2018).

To address these difficulties, Givnish et al. (2010) and several subsequent authors conducted phylogenomic analyses of plastome-scale data (sequences of ≥65 plastid genes) combined with a relatively dense taxon sampling across monocots or individual orders or families, to produce fully resolved, strongly supported monocot phylogenies unlikely to be skewed by sparse sampling or long-branch attraction (Steele et al., 2012; Barrett et al., 2013, 2014b, 2016; Burke et al., 2014, 2016a, b; Henriquez et al., 2014; Jones et al., 2014; Givnish et al., 2015, 2016a, b; Lam et al., 2015, 2018; Mennes et al., 2015; Wysocki et al., 2015, 2016; Duvall et al., 2016; Ross et al., 2016; Sass et al., 2016; Kim et al., 2017; Saarela et al., 2018).

Here we draw on the enormous amount of data generated by these studies and present the first plastome phylogeny to encompass all 77 monocot families (and most subfamilies) through a dense sampling of 77 plastid genes and 545 species. We calibrate this phylogeny against time using 13 fossils, identify the most likely sister group to the monocots, estimate stem ages of all monocot orders and families, compare branch lengths of mycoheterotrophic and autotrophic sister taxa, and calculate rates of net species diversification across 101 major lineages at the family or subfamily level. To evaluate the power of phylogenomic vs. classical phylogenetic studies based on a few genes, we subsampled data to determine how the probability of identifying branches in the 77-gene tree, and their bootstrap support, vary with the number of genes sampled, branch length, and relative branch depth. We expect the probability of accurate branch ascertainment and the average level of bootstrap support to increase with branch length, the number of coding regions sampled, and the inclusion of non-coding regions, and that ascertainment and support will both decrease with relative branch depth, based on simulations and analyses by Givnish and Sytsma (1997), Townsend and Leuenberger (2011), and Klopfstein et al. (2017). We also present an analysis of complete aligned plastomes—including both coding

and non-coding regions—to quantify the importance of both kinds of regions and reassess the phylogeny of Zingiberales, whose early branches have proven recalcitrant to satisfactory resolution (Kress et al., 2001; Barrett et al., 2014b; Sass et al., 2016).

<h1>MATERIALS AND METHODS

<h2>Sampling of taxa and genes

We compiled sequences of 77 protein-encoding plastid genes for 545 species of monocots and 22 angiosperm outgroups mostly from GenBank (Appendices S1–S3; see Supplemental Data with this article), drawing on research by our MonAToL (Monocot Assembling the Tree of Life) Team (Givnish et al., 2010, 2015, 2016a, b; Steele et al., 2012; Barrett et al., 2013, 2014b, 2016; Comer et al., 2015; Lam et al., 2015, 2018; Ross et al., 2016), by M. Duvall, L. Clark, and their collaborators (Burke et al., 2012, 2014, 2016a, b; Jones et al., 2014; Cotton et al., 2015; Saarela et al., 2015; Wysocki et al., 2015, 2016; Attigala et al., 2016; Duvall et al., 2016), and by a number of other investigators (Wu and Ge, 2012; Henriquez et al., 2014; Logacheva et al., 2014; Huang et al., 2016; Sass et al., 2016; Vieira et al., 2016; Kim et al., 2017).

Complete, circularized plastomes were generated de novo for five species of the order Commelinales. and draft plastomes were generated for 52 species of order Zingiberales (Appendix S3). We extracted total genomic DNA from silica-dried material using a modified CTAB protocol (Doyle and Doyle, 1987) or DNeasy Plant Mini Kits (Qiagen, Germantown, Maryland, USA). DNA quantity and quality were examined via spectrophotometry (Nanodrop; Thermo Fisher, Waltham, Massachusetts, USA) and nucleic acid staining (Qubit; Thermo Fisher). Libraries were prepared at Cold Spring Harbor Laboratories or at the Evolutionary Genetics Lab at the University of California–Berkeley and then pooled with others at equimolar ratios and subjected to paired-end sequencing (100 base pair [bp] reads) on an Illumina HiSeq 2000 (Illumina). For the 52 Zingiberales species, libraries were subject to plastid sequence enrichment on an Agilent 1M microarray (G3358A) prior to sequencing (Sass et al., 2016). Library preparation, sequencing, quality trimming, plastome assembly, and annotation were performed as described in Barrett et al. (2013, 2015) and Sass et al. (2016). For the *Anigozanthos* accession, we performed library preparation (Nextera kits; New England Biolabs, Ipswich, Massachusetts, USA) and Illumina sequencing at the Génome Québec Innovation Centre, with

125 bp paired-end reads on an Illumina HiSeq 2000. Genome assembly and annotation were conducted following Ross et al. (2016).

Monocot accessions were stratified across all families and most subfamilies of monocots, with the inclusion of a number of congeneric species to enhance the range of branch lengths represented. Outgroups follow Givnish et al. (2015). We included representatives of Anarthriaceae, Centrolepidaceae, Taccaceae, and Thismiaceae—as recognized by Briggs et al. (2014) and Merckx et al. (2010) but not APG (2016)—to test whether plastome-scale data support segregation of these families. In a small number of cases (Appendix S4: 21 species), we drew from the 17 plastid genes sequenced by Graham et al. (2006), Saarela (2006), and Saarela et al. (2008) to represent lineages lacking plastome-scale data or to increase taxon sampling for other lineages. We conducted analyses with and without *Thismia tentaculata*, given that the plastome of this species has by the far the fewest genes known among monocots (Lim et al., 2016) and is one of the most rapidly evolving (Lam et al., 2018); its presence might have a substantial effect on the ascertainment, length, and/or support of other branches in the monocot phylogeny (Lam et al., 2016).

To compare results from phylogenomic vs. phylogenetic data, we analyzed a second data set of sequences for four plastid genes (*atpB*, *matK*, *ndhF*, *rbcL*) from Chase et al. (2006) for 125 monocots stratified across families and for 16 angiosperm outgroups. We chose this data set as a benchmark because it had the most thorough sampling of monocot families among multi-gene analyses prior to the phylogenomic era. We excluded *Trithuria* (Hydatellaceae) from the benchmark data because subsequent analyses (Saarela et al., 2007; Soltis et al., 2011) showed that the early sequences for this taxon were erroneous. To assess the informativeness of coding vs. non-coding regions for monocot plastid phylogeny, we compiled a third data set of complete aligned plastomes, including 52 species representing all eight families of Zingiberales (Sass et al., 2016) as well as four outgroup species from Commelinales and Poales (Appendix S5).

<h2>Alignment

We parsed and independently aligned coding and non-coding regions for the monocot and Zingiberales data sets using MAFFT version 7.309 (Katoh et al., 2013) as implemented in Geneious version 9.1.8 (http://www.geneious.com; Kearse et al., 2012) and then refined the alignments using MUSCLE (Edgar, 2004) as implemented in Geneious. The alignments were

manually checked using DNA translation to maintain the reading frame of the coding regions. The ends were trimmed when alignment showed highly divergent regions that could not be properly aligned. The genes *accD* and *ycf1* were excluded from the monocot data set because they were present in less than half of the samples and the genes were too difficult to align because of high variability. We concatenated individual alignments in a single matrix. Final alignment lengths are 83,478 bp for the monocot set without *Thismia* and 94,597 bp with *Thismia*, and 179,545 bp for Zingiberales. For the across-monocots alignments, the proportion of missing bases with *Thismia* included is 9.60%, and 9.45% with *Thismia* excluded; the corresponding proportion of alignment gap cells is 35.07% and 34.96%, respectively, reflecting the substantial divergence across monocots and angiosperm outgroups. Missing data accounted for only 0.13% of the Zingiberales matrix; gaps, 30.52%. These alignments are available on the Open Science Framework at https://osf.io/g4me6/. The distribution of missing data by loci and taxa is provided in a heat map in the OSF folders. The four-gene data included 7019 aligned bases; missing data accounted for 4.7% of the matrix; gaps, 19.5%.

<h2>Phylogenomic analyses

We estimated phylogenetic relationships with maximum likelihood (ML) using RAxML version 8.2.11 (Stamatakis, 2014). Partitions were determined using PartitionFinder 2 (Lanfear et al., 2017). A total of 54 partitions of the 77 genes were designated for the across-monocots analysis excluding *Thismia*; 48 used a GTR+Γ+I model, and the remaining six used a GTR+Γ model (Appendix S6). Because RAxML can accommodate only one DNA substitution model, we parameterized all 54 partitions using GTR+Γ+I. The same data partitions were used for consistency in the analysis including *Thismia*, but parameterized independently. We calculated the Robinson-Foulds (RF) distance between the trees with and without *Thismia*—the number of clades that they do not have in common—as a measure of tree similarity (Penny et al., 1982; Sheikh et al., 2013). Gaps were treated as missing data. We employed 500 rapid bootstrap replicates to quantify support for individual branches. *Amborella* was used as the operational outgroup for tree visualization, given its consensus position as sister to all other angiosperms (e.g., Drew et al., 2014).

We used similar approaches for identifying partitions and running analyses for the four-gene data set and the aligned full-plastome data set for Zingiberales. Each gene in the four-gene data

set was designated as an independent partition using a GTR+Γ+I model. We also analyzed the four-gene data set using maximum parsimony (MP), the approach originally employed by Chase et al. (2006); our analyses were run using PAUP* 4.0b10 (Swofford and Bell, 2017) treating gaps as missing data. In Zingiberales, we designated 250 loci by separating exons, introns, and spacers (Appendix S7). PartitionFinder 2 grouped these into 95 partitions: 68 were assigned to the GTR+Γ model, 25 to GTR+Γ+I, and 2 to GTR (Appendix S8). We therefore applied GTR+Γ to each of the 95 partitions, individually parameterized.

We tested the hypothesis that non-green mycoheterotrophic taxa have consistently higher plastid substitution rates than green, photosynthetic relatives using a two-tailed Wilcoxon paired-sample sign test. We used patristic GTR-distances from each mycoheterotrophic taxon and its sister photosynthetic taxon to their common ancestor on the partitioned RAxML tree, employing the "ape" package (Paradis et al., 2004) in R (R Core Team, 2018). We compared the average branch length for all mycoheterotrophs with that of all autotrophs in the corresponding sister clade, except for *Burmannia bicolor*, which we contrasted with the average branch lengths of *Dioscorea elephantipes* and *D. rotundata*, so that this comparison was computationally independent from that of the sister-pair *Thismia tentaculata* and *Tacca chantrieri*. We regressed the average distance from ancestor to mycoheterotroph(s) vs. that from ancestor to green sister group to test for a general lineage effect on branch lengths for both green and non-green taxa.

<h2>Dating and net rates of species diversification

We modeled divergence times on the full monocot data set using BEAST version 2.4.7 (Bouckaert et al., 2014). With the exception of Musaceae, we used the ML tree as a topological constraint in this analysis, thereby restricting the MCMC to estimate divergence times only. We constrained the position of Musaceae to accord with the Zingiberales phylogeny based on whole aligned plastomes (see below). For computation purposes and to minimize the effects of missing data, we used only the genes *atpB, psaA, psbD, rbcL*, and *rps4* in this analysis, rather than the full set of 77 genes (see Givnish et al., 2016a, b). We validated this procedure by comparing branch lengths in the RAxML tree based on these five loci vs. all 77 loci. A linear regression showed a very tight fit of branch lengths based on five loci to the total lengths ($y = 0.745x - 0.000117$; $r^2 = 0.947$, $P << 0.0001$; the intercept does not differ significantly from zero). We selected models of nucleotide substitution for each of the five genes using jModelTest version

2.1.4 (Darriba et al., 2012). These were identified as SYM+Γ+I for *atpB*, SYM for *psaA*, TVM+Γ+I for *psbD* and *rps4*, and TIM+Γ+I for *rbcL*. We used a Yule tree prior and an uncorrelated lognormal clock, and unlinked the clock and site models for each gene.

We used a total of 13 fossil and seven secondary calibration priors (Appendix S9; see Givnish et al., 2015; Iles et al., 2015). Fossil priors were placed with a lognormal distribution with a broad standard deviation (SD = 2), accommodating uncertainty in both fossil age and phylogenetic placement. We placed secondary priors with a uniform distribution on the root and *Illicium* stem node, following Sytsma et al. (2014) and Givnish et al. (2015, 2016a, b). We also placed priors on the crown nodes of the rosids (Sytsma et al., 2014), Cyperaceae (Spalink et al., 2016), Caryophyllales + asterids (Givnish et al., 2015), magnoliids (Givnish et al., 2015), and monocots (Magallón et al., 2015). We ran two independent chains of 100 million generations on CIPRES (CyberInfrastructure for Phylogenetic RESearch Science Gateway; Miller et al., 2010); assessed effective sampling sizes, stabilization, and convergence using Tracer version 1.5 (Drummond and Rambaut, 2009); and annotated the maximum clade credibility tree using TreeAnnotator version 2.4.6 (Bouckaert et al., 2014) after removing the first 25% of samples as burn-in.

We used BAMM (Bayesian analysis of macroevolutionary mixtures: Rabosky, 2014; Rabosky et al., 2014) to measure diversification rates across monocots and test for significant shifts in such rates. We pruned the BEAST chronogram to include only the monocots and assigned tips to the smallest taxonomic unit for which we had adequate coverage and for which richness values could be confidently assigned. *Thismia* was grafted onto the chronogram, diverging from *Tacca* 11.51 Mya based on the proportional position of the *Thismia-Tacca* divergence on the branch sister to Dioscoreaceae in the analysis including *Thismia* (see above). Species were assigned to individual families or (in some cases) subfamilies, tribes, or subtribes. Species richness values of terminal taxa were obtained from eMonocot (http://e-monocot.org). We set BAMM priors using the "setBAMMpriors" function in BAMMtools version 2.1.6, with a conservative expected number of shifts prior of 1 (Rabosky et al., 2014). We ran two chains of 200 million generations each and assessed effective sampling using the R package "coda" (Plummer et al., 2006). We assessed the distribution of species richness among APG IV families plus Taccaceae and Thismiaceae using linear regression against family rank based on richness, and examined whether young families are especially poor in species using a two-tailed *t*-test.

<h2>Effects of gene number, branch length, relative branch length, and inclusion of non-coding spacers on branch ascertainment and bootstrap support

We hypothesized that branch ascertainment (the probability of inferring a branch found in the total-evidence tree) and branch support (average bootstrap value) should increase with the number of loci sampled, branch length, and (for the Zingiberales analysis) inclusion of non-coding spacers and introns in addition to exons, and that ascertainment and support should decline with relative branch depth. Other things being equal, the greater the number of inferred substitutions—caused either by sampling more loci or by sampling the greater number undergoing shifts along longer branches—the greater should be the likelihood of branch ascertainment and associated bootstrap support (Givnish and Sytsma, 1997). Relatively greater branch depth should, other things being equal, decrease branch ascertainment and support, potentially reducing the support for deep branches as subsequent mutations overwrite the signal of early common ancestry, with that effect being greater for deeper branches (Givnish and Sytsma, 1997; Townsend, 2007; Klopfstein et al., 2010, 2017; Townsend and Leuenberger, 2011).

To test these ideas, branch length was taken from the RAxML tree using the full alignment. We excluded the 24 taxa represented by 16 or fewer genes, and 25 additional taxa to avoid overrepresentation of certain genera (Appendix S10). Relative branch depth was defined as the average distance between the branch midpoint to the branch's descendant taxa, divided by the distance from the root to these same taxa, and thus varied from 0 to 1.

For the across-monocots analysis, we randomly sampled 2, 5, 10, 20, or 50 loci (each consisting of a gene ± adjacent spacer) without replacement, as well as a run including all 77 loci in the monocot alignment. For the across-Zingiberales analysis, we considered a total of 90 loci (each consisting of an exon ± intron(s) and an adjacent spacer; Appendix S7) and randomly sampled 2, 5, 10, 20, 50, or 77 loci without replacement, as well as a run including all 90 loci. Each of the 79 protein-coding genes and four ribosomal RNAs formed the core of a locus, with intron(s) included if present. Five transfer RNAs (tRNAs) were long (≥800 bp) due to the presence of introns, and we considered each to form the core of a composite "locus." The remaining 23 tRNAs were short (≤93 bp), so we lumped them into two composite "loci" (data partitions): one composed of those in the large single-copy (LSC) region, and the other

composed of those from the small-single-copy region (SSC) and one copy of the inverted repeat (IR). These 90 loci spanned all coding regions and introns. In analyses to quantify the potential added support lent by spacers in Zingiberales, we expanded each locus to include a spacer next to the exon (inclusive of intron(s)) for the across-monocots analysis, and a spacer + intron(s) next to the exon for the across-Zingiberales analysis, except for four loci (*psbC, ndhC, atpB, ndhF*) that are directly adjacent to (or, for *psbD/psbC*, overlapping with) another coding region.

For the across-monocots and across-Zingiberales analyses, we randomly sampled loci 100 times for each number of loci and then analyzed each sampling with RAxML using the HKY+Γ model and 100 bootstrap replicates, for computational feasibility. For each branch in the best tree from the corresponding full alignment across monocots or Zingiberales, we calculated *branch ascertainment* as the proportion of times that branch was present in the best tree from each of the 100 subsampled alignments, and *branch support* as its average bootstrap support across those 100 alignments.

We used logistic regression to explain branch ascertainment and support using all predictors (number of loci as a categorical variable, whole-alignment branch length and relative depth, and—for the across-Zingiberales analysis only—inclusion of spacers), with three models. The first included main effects only, where branch ascertainment (or support) ζ was modeled as $\zeta = 1/(1+\exp(-(a + b \log n \text{ genes} + c \log \text{branch length} + d \log \text{relative branch depth} + e \text{ spacer presence})))$, where a, b, c, d, and e are coefficients to be estimated from the data (e pertains to the Zingiberales analysis only). We included two-way interactions in a second model, and all interactions in a third model; all models allowed for overdispersion. For each model, we estimated coefficients with standard errors and compared the model deviance (akin to residual variation) to the null deviance (akin to the total variation in the data). All statistical analyses were conducted in R (code available at https://osf.io/g4me6/).

<h1>RESULTS

<h2>Phylogenomic analyses

Plastome data excluding *Thismia* yield a fully resolved, strongly supported RAxML tree; 421 of 543 branches within the monocots have 100% bootstrap support, and average branch support is 96.6 ± 10.2% (Appendix S11). All 12 orders—including Dasypogonales (*sensu* Givnish et al., 1999; not recognized by APG 2016)—have 100% bootstrap support, as do all but two of the 51

APG IV families represented by two or more accessions (Fig. 1A). Corsiaceae, a family composed solely of mycoheterotrophs, has 99.2% bootstrap support based on plastid gene data, as does the autotrophic family Stemonaceae. All subfamilies represented by two or more accessions also have 100% bootstrap support (except Arundinoideae with 99.8% support), as evident in the large families Orchidaceae, Poaceae, Araceae, Asparagaceae, Arecaceae, and Asphodelaceae. Monocots as a whole have 100% bootstrap support. *Ceratophyllum* + eudicots are resolved as their sister group, with 89.8% bootstrap support; the *Ceratophyllum* + eudicots clade has 73.8% support. Among the commelinid monocots (100% support), Poales is resolved as sister to Commelinales + Zingiberales (95.2% support); Arecales is sister to Dasypogonales (74.2% support). Asparagales is resolved as sister to the commelinids with 100% bootstrap support. Liliales is sister to this large clade, with Dioscoreales + Pandanales, Petrosaviales, Alismatales, and Acorales sister to progressively more inclusive monocot clades, all with ≥99.8% bootstrap support (Fig. 1A and Appendix S11). *Trithuria* of Hydatellaceae—formerly placed in Poales—is resolved as a member of order Nymphaeales with 100% bootstrap support, confirming the results of Saarela et al. (2007). Relationships among subfamilies in the two largest monocot families—Orchidaceae and Poaceae—are identical to those resolved by Givnish et al. (2015) and the Grass Phylogeny Working Group II (2012), except that Aristidoideae is placed sister to Panicoideae within the PACMAD clade of grasses but with weak support (44%) (Appendix S11).

The weakest support values for branches above the family level within the monocots are those that unite (1) Tofieldiaceae to Araceae (bootstrap support, BS = 35.0%); (2) Musaceae to the ginger families Cannaceae, Marantaceae, Costaceae, and Zingiberaceae (50.0%); (3) Philydraceae to Haemodoraceae-Pontederiaceae (50.6%); (4) Doryanthaceae to Asparagaceae through Iridaceae (62.0%); (5) Typhaceae to all Poales except Bromeliaceae (62.6%); (6) Heliconiaceae to Lowiaceae-Strelitziaceae (73.4%); (7) Dasypogonales to Arecales (74.2%); (8) Lanariaceae to Hypoxidaceae (75.2%); and (9) Asteliaceae to Hypoxidaceae-Lanariaceae (77.0%). Of these, the moderate support for the last two clades may reflect the small number of loci (17) for those three families in our alignment. The remaining problematic branches are all exceedingly short in length, and often deep in the monocot tree; two involve the positions of Musaceae and Heliconiaceae in Zingiberales. All other interfamilial branches have support values of 100% or nearly so (Fig. 1A). Among subfamilies formerly recognized as families in

earlier APG schemes, only the position of (10) Aphyllanthoideae sister to Asparagoideae in Asparagaceae has relatively weak support (53.2%); relationships among the remaining subfamilies are strongly supported.

Compared to the MP analysis of the data of Chase et al. (2006) (Fig. 1C), our RAxML phylogenomic analysis shifts the positions of one order (Dasypogonales) and 16 monocot families or their equivalents (subfamilies of Amaryllidaceae, Asparagaceae, and Asphodelaceae regarded as families in earlier APG classifications). Our analysis also adds seven families not represented in that earlier data set: Corsiaceae in Liliales; Triuridaceae in Pandanales; and Aponogetonaceae, Maundiaceae, Posidoniaceae, Ruppiaceae, and Scheuchzeriaceae in Alismatales. Phylogenomic analysis resolves the two among-family polytomies in the MP strict consensus tree, changes the positions of families in eight of the nine orders with multiple families relative to the MP analysis of four-gene data, and substantially increases bootstrap support for several higher-order relationships within monocots (Fig. 1A, C). The contrast of the phylogenomic tree with the ML analysis of the four-gene data set (Fig. 1B) is somewhat less marked, shifting the position of Dasypogonales and eight monocot families, but embedding the sole representative of the former Centrolepidaceae in a paraphyletic Restionaceae. Support values in the ML analysis of the benchmark four-gene data are also somewhat higher than those in the MP analysis.

Addition of *Thismia* to the across-monocots data set results in a tree of nearly identical topology, with *Thismia* placed sister to *Tacca* with 61.0% bootstrap support and only two clades differing between the trees with and without *Thismia*. These results support the recognition of segregate Taccaceae and mycoheterotrophic Thismiaceae, as sister to each other and jointly sister to Dioscoreaceae (Appendix S12). *Thismia* has undergone unparalleled loss of plastid genes and sits on a long branch; its inclusion alters the position of 12 taxa and lowers support values for several (mostly short) branches throughout the tree, although the greatest impact is on support of nearby branches in Dioscoreales and Pandanales (Fig. 1 and Appendices S11, S12). Given the distortions that *Thismia* might impose on the dating analysis, we base that initially on the tree excluding it (Fig. 1 and Appendix S1).

The Zingiberales data set yields a basal split based on an analysis of complete aligned plastomes, with the four ginger families ((Costaceae, Zingiberaceae), (Cannaceae, Marantaceae)) forming a clade sister to the four banana families (((Musaceae, (Heliconiaceae, (Lowiaceae,

Strelitziaceae))) (Fig. 2). Relationships among families or family lineages all have 100% bootstrap support, except for the placement of Heliconiaceae sister to Lowiaceae-Strelitziaceae (94.2%), and that of Musaceae sister to that trio (83.8%). All but three other branches in the Zingiberales tree have 100% bootstrap support, with the only weaker value (71.8%) being for the very short branch at the root of *Riedelia* + Zingiberaceae tribe Alpinieae, a result that would make the tribe Riedelieae (including *Siamanthus*) paraphyletic based on the sequences represented.

Mycoheterotrophic taxa had longer branches (more substitutions/site) than their photosynthetic sister taxa in all paired comparisons. This pattern was statistically significant, either including *Thismia* (Wilcoxon $V = 36$, $P < 0.008$) or excluding it ($V = 28$, $P < 0.016$). The average ratio of mycoheterotrophic to photosynthetic branch lengths was 6.9 ± 4.1 excluding *Thismia* and *Tacca*; the ratio for *Thismia* vs. *Tacca* was 363.7. Mycoheterotroph branch length did not, however, increase significantly with the branch length of its photosynthetic sister group across lineages, including *Thismia* ($r = -0.327$, $P > 0.42$) or excluding it ($r = 0.444$, $P > 0.27$). With the partial exception of that last finding, comparisons based on β (substitutions/site/My) yield nearly identical results. Excluding *Thismia*-*Tacca*, $\beta = 5.3 \times 10^{-4} \pm 4.2 \times 10^{-4}$ for green taxa, while $\beta = 3.6 \times 10^{-3} \pm 4.0 \times 10^{-3}$ for sister mycoheterotrophic taxa. For *Thismia*, $\beta = 1.5 \times 10^{-1}$, which is 277× the mean rate for autotrophs and 353 standard deviations above the green mean, and 40.6× the mycoheterotrophic mean excluding *Thismia* and 36.5 standard deviations above the non-green mean. With *Thismia* excluded, there is a highly significant relationship of β for mycoheterophs vs. paired autotrophs, $y = 8.3x - 0.0008$ ($r^2 = 0.77$, $P < 0.01$ for two-tailed *t*-test, df = 5). But with *Thismia* included, or excluding both outliers (*Thismia* and *Rhizanthella*), that relationship essentially disappears ($r = 0.05$ and -0.27, respectively).

<h2>Dating and net rates of species diversification

Monocotyledons appear to have diverged from the common ancestor of *Ceratophyllum* and the eudicots 136.1 Mya, in the early Cretaceous, and extant monocot lineages began differentiating 132.4 Mya (Fig. 3 and Appendix S3). By 114 Mya, all 12 orders recognized here had diverged from each other; by the end of the Cretaceous (65 Mya), 45 (58%) of the 77 APG IV families had diverged from each other (44/79 = 56% if Taccaceae and Thismiaceae are recognized). Our estimates place the stem and crown nodes of Alismatales at 130 and 124 Mya; Pandanales, at

124 and 93 Mya; Dioscoreales, at 124 and 119 Mya; Liliales, at 126 and 111 Mya; Asparagales, at 125 and 116 Mya; Dasypogonales, at 119 and 34 Mya; Arecales, at 119 and 85 Mya; Zingiberales at 114 and 83 Mya; Commelinales, at 114 and 110 Mya; and Poales at 124 and 120 Mya. The most recently divergent families were Taccaceae and Thismiaceae at 11.5 Mya, and Lapageriaceae and Philesiaceae at 20.0 Mya; the most ancient divergence between families (and orders) was that between Arecaceae and Dasypogonaceae at 119 Mya. Families that diverged after the Cretaceous are concentrated in Zingiberales, Asparagales, Liliales, and submersed members of Alismatales (i.e., excluding Araceae and Tofieldiaceae) (Fig. 3). Families of Zingiberales had the youngest average stem age (43.4 My) and those of Poales the oldest average stem age (105.6 My). Monocot family ages are bimodally distributed, with peaks at 50–60 Mya and 110–120 Mya, and a trough at 70–90 Mya (Fig. 4). Monocot family sizes show a close approach to a negative exponential (log-series) distribution (Fig. 5). The 12 youngest families based on stem age have a significantly smaller number of species than the remaining 67 families (95 ± 163 vs. 1055 ± 3388, $P < 0.026$ for two-tailed t-test with unequal variances, $t = -2.28$, df = 67.65). Eight of these 12 families have distributions limited to one or a few regions (e.g., Lanariaceae, Philesiaceae, Ripogonaceae, Strelitziaceae); Hypoxidaceae and Marantaceae are notable exceptions to this trend.

The rate-shift configuration with the highest posterior probability supports four significant, large-scale increases in net diversification rate (Fig. 3 and Appendix S14; for the set of credible shifts, see Appendix S15). These include accelerations in the stem groups of (1) the PACMAD-BOP clade of Poaceae, (2) Asparagales sister to Doryanthaceae, (3) Orchidoideae-Epidendroideae of Orchidaceae, and (4) Araceae sister to Lemnoideae. Net rates of diversification varied sixfold, from 0.029 to 0.18 sp sp^{-1} My^{-1}, with the highest rates in orchids. Significant accelerations in diversification began between 75 Mya (in orchids) and 54 Mya (in aroids). Speciation rates (λ) varied tenfold, from 0.027 to 0.26 sp sp^{-1} My^{-1}, with aroids having the highest rates. Extinction rates (μ) ranged from 0.002 to 0.13 sp sp^{-1} My^{-1}, with the highest rates again being in aroids (Appendix S14). Note that there was a slight general slowdown in diversification after the initial rapid divergence of orders early in monocot evolution (Fig. 3).

<h2>Effects of gene number, branch length, relative branch length, and inclusion of non-coding spacers on branch ascertainment and bootstrap support

Across the pruned 77-gene monocot tree used for resampling analyses, both branch length and relative depth are lognormally distributed; log relative depth is positively correlated with log length in highly significant fashion, explaining ~25% of the variance in the former ($r = 0.50$, $P < 2.2 \times 10^{-16}$; Appendix S16). Very shallow branches at or near the tips of the tree tend to be slightly shorter than might be expected by chance. This pattern mostly reflects our definition of branch depth, which used the midpoint of each branch. If a long branch runs from a deep diversification event near the root to a recently diverged clade very close to the tips, its midpoint must necessarily be relatively deep. This pattern may also reflect the inclusion of several congeners or otherwise closely related species in the sample, which often reside on short, shallow branches.

Across monocots, branch ascertainment and bootstrap support both increased logistically with log branch length and log number of genes sampled, and decreased logistically with log branch depth, in highly significant fashion ($P < 10^{-16}$) for all three predictors (Table 1 and Fig. 6). In general, ascertainment and support were lowest on the shortest and/or deepest branches, and low values of ascertainment and support were increasingly confined to the shortest, deepest branches as gene number increased (Fig. 7). Main effects explained 66.6% of the deviance in ascertainment, vs. 67.0% in a model including all two-way interactions, and 67.1% when the three-way interaction was added. For bootstrap support, the main-effects model explained 76.5% of deviance, vs. 77.0% when two-way interactions were included, and 77.1% when the three-way interaction was added.

At a relative branch depth of 0.05 (i.e., near the root), and branch lengths (substitutions/site) between 5×10^{-5} and 3×10^{-3}, the probability ζ of branch ascertainment increased sharply with the number of genes n sampled, with ζ rising from 0.15 to 0.85 as n increased from 2 to 77 (Fig. 6). A similar increase in branch support with the number of genes sampled occurred over a slightly wider range of branch lengths. Very short branches had low ascertainment and support that increased only slightly with gene number; long branches had high ascertainment and support regardless of gene number. At a branch length of 0.001, ascertainment and support decreased at an ever-increasing rate with the logarithm of branch depth (Fig. 6). The maximum effect of increasing the number of genes sampled occurred between relative depths of 0.05 to 1.0. When multiple variables are considered, the benefit of increased numbers of genes sampled—in terms of the *absolute* increase in percent ascertainment and support—was greatest for intermediate

branch length × depth combinations that occupy a SW-NE swath in the space of possible combinations, and greatest for short but shallow branches (Appendix S17).

Across Zingiberales, in the analysis of whole aligned plastomes, branch ascertainment and bootstrap support also increased logistically with log branch length and log number of genes sampled, and decreased logistically with log branch depth, in highly significant fashion in each case (Table 1 and Appendix S18). The main-effects models again explained virtually all the deviance accounted for by models, including two- and three-way interactions among predictors—82% vs. 87% for ascertainment, and 84% vs. 89% for bootstrap support. Including non-coding regions significantly increased both ascertainment and support, with the benefit greatest for "intermediate" branch length × depth combinations that occupy a SW-NE swath in the space of possible combinations, an effect that was greatest in the shortest, shallowest branches for genes only, and for the deepest branches for genes and spacers (Appendix S19).

<h1>DISCUSSION

Maximum likelihood analysis of 77 plastid genes simultaneously clarifies for the first time the evolutionary position of all monocot families, places even non-green mycoheterotrophic taxa with strong support, and supports recognition of Dasypogonales, Taccaceae, and Thismiaceae. Analysis of whole aligned plastomes—including intergenic spacers—further increases support for several relationships within Zingiberales and unveils a basal split in the order between the banana and ginger families. Phylogenomic data provide 100% bootstrap support for 49 of 51 families with two or more accessions, and 99.2% for Corsiaceae and Stemonaceae. Resampling studies demonstrate the substantially greater power of plastome-scale data vs. many fewer plastid genes, showing that ascertainment of, and support for, individual branches increase with the number of genes analyzed and branch length, and decrease with relative branch depth, as predicted. For Zingiberales, the same patterns hold, and inclusion of aligned spacers further increases ascertainment and support. Finally, dating of the plastome phylogeny based on the ages of both fossils and secondary calibration points provides a new timeline for monocot evolution, and identifies four significant accelerations of net species diversification. Key points regarding these findings are detailed below.

<h2>Phylogenetic relationships

This study is the first to analyze phylogenetic relationships among all 77 monocot families using plastome-scale sequences. The resulting (presumably) maternal tree is fully resolved and strongly supports almost all higher-order relationships for the first time, providing an important backbone and timeline for future studies of monocot evolution. Our analyses of 77 plastid genes add seven families—including the wholly mycoheterophic Corsiaceae and Triuridaceae—and shift the positions of one order and 16 families in relation to those based on sequences of four plastid genes compiled by Chase et al. (2006). Plastome-scale data and ML analysis substantially increase bootstrap support for several branches across the monocot tree, and resolve two polytomies in the strict consensus phylogeny emerging from MP analysis of the benchmark data (Fig. 1A, B, C). Comparison of the ML and MP trees based on the benchmark data suggests that generally higher levels of support for the 77-gene tree mainly reflect the scope of the data, not methodology (Fig. 1B, C). Our phylogenomic analyses—coming a decade after benchmark study of Chase et al. (2006) and drawing on efforts by our team and several other labs—use 11.9× more aligned sequence data per taxon and a 4.4-fold denser sampling of taxa, for a >50-fold increase in the total amount of data analyzed.

<h3>Commelinids

Plastid phylogenomics confirms the relationships among the five orders of commelinid monocots first demonstrated with strong support by Givnish et al. (2010) and Barrett et al. (2013, 2016): the woody (i.e., strongly lignified) orders Arecales and Dasypogonales are sister to each other (74.2% BS) and jointly sister (95.2% BS) to the herbaceous orders Poales and Commelinales-Zingiberales (100% BS for both nodes in the Poales-Commelinales-Zingiberales clade, PCZ). Barrett et al. (2016) obtained 81% bootstrap support for Arecales-Dasypogonales and 92% for PCZ for a nearly identical sampling within the orders using partitioned, codon-based ML analysis; Givnish et al. (2010) obtained 86% and 93% support for these same nodes using a less extensive taxon sampling and unpartitioned ML analysis.

APG (2016) lumped Arecaceae and Dasypogonaceae into a single order, based on a preference for multiple families in every order. We believe that Arecales and Dasypogonales should instead be recognized as separate orders, because they are individually distinctive, share few if any morphological synapomorphies other than woody habit (Givnish et al., 2010), and diverged further back in time (>125 Mya) than any other pair of monocot families or orders (Fig.

3). Rudall and Conran (2012) cite similarities of Dasypogonaceae to Bromeliaceae, Cyperaceae, Rapateaceae, and Thurniaceae, as well as Arecaceae and several zingiberalean families in silica bodies; to Eriocaulaceae, Rapateaceae, and Typhaceae in inflorescence structure; and to Rapateaceae in ovule anatomy and diversity of nectaries. Thus, there is no clear set of phenotypic characters uniting Dasypogonaceae and Arecaceae within the commelinids beyond their woody habit, and analyses based on both morphology and molecules have long struggled to identify the closest extant relatives of both families. However, each of these families individually is highly distinctive morphologically and easy to diagnose. There is no dictum that orders cannot comprise single families; indeed, APG IV recognized several single-family orders (e.g., Amborellales, Buxales, Chloranthales). At lower levels, across monocots and the angiosperms as a whole, there are large numbers of monogeneric families and monospecific genera. Recognition of orders should not be based on rigid avoidance of single-family units, or merely on evolutionary relationships (i.e., cladograms), but should also reflect phenotypic divergence, morphological diagnosability, and age of divergence between different clades. We propose that it is better to recognize Dasypogonales as its own order rather than as part of an extraordinarily anomalous, expanded order of "palms," and that the next update of the APG schema should be emended accordingly.

Within Poales, the only interfamilial relationships with substantially less than 100% support involve Typhaceae and the grade of xyrid families (Eriocaulaceae, Mayacaceae, Xyridaceae). Typhaceae lies on a very short, very deep (>130 Mya) branch in an order with extreme variation in rates of molecular evolution, and *Mayaca* and Eriocaulaceae-Xyridaceae also appear to lie on short, deep branches. It might be possible to increase support for both nodes via analyses of whole-plastome sequences, as used to good effect in Zingiberales (see Figs. 1 and 2). Yet even that approach succeeded in placing Musaceae sister to Heliconiaceae-Lowiaceae-Strelitziaceae with "only" 83.8% support, and the problematic branches in Poales are both deeper and shorter. It would be helpful to sequence plastomes for *Mayaca baumii* (the only species of Mayacaceae found in Africa) and for several more genera of Eriocaulaceae. The alternative placement of Mayacaceae as sister to cyperids, with both sister to graminids, restiids, and remaining xyrids by Bouchenak-Khelladi et al. (2014) in a Bayesian analysis of *rbcL* and *ndhF* sequences for 545 species of Poales is not well supported, and rests on far less sequence data per taxon than our analysis. Our simulations indicate that analyses based on two genes have a substantially lower

chance of correct ascertainment than those based on plastome-scale data. For example, nine of 15 interfamily relationships in Poales collapsed, compared to our analyses, even when four loci (plastid *rbcL*, *matK*, *ndhF*; nrDNA ITS) were analyzed by Tang et al. (2017).

Our plastome data are consistent with retention of Anarthriaceae, Centrolepidaceae, and Restionaceae. However, our current sampling is inadequate to determine whether Centrolepidaceae is embedded within a paraphyletic Restionaceae, as inferred from ML but not MP analyses of plastid *trnL-F*, *trnK*, and *rbcL* with dense taxon sampling for both families (Briggs et al., 2014). The APG (2016) submerged both Anarthiaceae and Centrolepidaceae into Restionaceae due to this uncertainty, yet there is no doubt that Anarthriaceae (3 genera, 11 species) is morphologically distinct from the other two entities and strongly supported as being sister to both on the basis of molecular data. We thus recommend that Anarthriaceae be resurrected in the next version of the APG schema.

Within Poaceae, plastome data resolve the same relationships among the 12 grass subfamilies as those obtained by the Grass Phylogeny Working Group II (2012) based on plastid *rbcL*, *ndhF*, and *matK*, except that Aristidoideae is weakly supported (44% BS) as sister to Panicoideae here, rather than all other families in the PACMAD clade (Appendix S11). However, GPWG II found weak (65%) ML bootstrap support for their proposed position of Panicoideae as sister to the remaining four PACMAD subfamilies. Saarela et al. (2018) recovered varying topologies among Aristidoideae, Panicoideae, and the rest of the PACMAD clade, depending on the plastome partition analyzed (coding regions, non-coding regions, whole plastomes) and whether gapped alignment sites were included or excluded. Based on a partitioned ML analysis of complete plastomes excluding gapped sites, they found Panicoideae to be sister to all other subfamilies, with Aristidoideae sister to the four remaining PACMAD subfamilies. Bootstrap support for this position of Aristidoideae was still weak (58%), although it climbed to 93% in analyses based purely on non-coding regions. When gapped sites were excluded from analysis, however, they found Aristidoideae to be sister to the rest of the PACMAD clade. In one analysis (coding regions with gapped sites stripped), Saarela et al. (2018) recovered the same topology recovered here: Aristidoideae Panicoideae sister to the rest of the PACMAD clade, also with weak support. Teisher et al. (2017) similarly addressed this uncertainty in the position of the aristidoids, using whole-plastome analyses to show that it is dependent on gap coding. Including no gaps or half the gaps placed Aristidoideae sister to the remainder of PACMAD, while including all gaps

placed Panicoideae in that position. Given that regions with alignment gaps are more likely to be ambiguously aligned, placing aristidoids sister to the rest of the PACMAD clade might be an artifact of ambiguous alignments or failure of ML to correctly treat gaps. Set against this, of course, is the greater amount of data included in the gapped alignments.

Within Zingiberales, ML analysis of whole aligned plastomes vs. 77 plastid genes shifted the position of Musaceae, substantially increased support for the placement of Musaceae and Heliconiaceae, and identified the four "banana" families (defined by broad leaves that are easily torn between the secondary veins) as a clade sister to the four ginger families (defined by one or one-half fertile anther and four or five highly modified staminodia) (Fig. 2). Our topology differs from that based on morphology (Kress, 1990) or on morphology and four genes (plastid *atpB* and *rbcL*, 18S and 26S nrDNA: Kress et al., 2001); from an alternative based on three plastid regions and three of nrDNA (Johansen, 2005); and from seven different tree topologies based on 83 plastid genes analyzed using ML without partition, or gene- or codon-based partitions (Barrett et al., 2014b). Across these analyses, support is strong for each ginger family and all the relationships among them, but the positions of the other well-supported clades (Heliconiaceae, Lowiaceae-Strelitziaceae, Musaceae) are inconsistent and have little support. For 52 taxa in Zingiberales, Sass et al. (2016) obtained complete aligned plastomes (nearly identical to those used in this paper) plus 308 nuclear genes, analyzed the sequences using gene-partitioned ML analysis, and found 100% bootstrap support for Heliconiaceae being sister to Lowiaceae-Strelitziaceae, all three being sister to the ginger families, and Musaceae being sister to all of these. Support for the latter, however, disappeared in their ASTRAL coalescent tree. Our partitioned ML analysis of these data produces the same phylogeny, but with Musaceae sister to Heliconiaceae + Lowiaceae-Strelitziaceae. The problematic nodes involving Musaceae and Heliconiaceae are very short and moderately deep (57.0 and 48.2 Mya, respectively) (Fig. 3 and Appendices S11, S13). Further studies should test whether the apparent conflict between the plastome tree and the tree based on both plastid and nuclear data reflects an ancient reticulation involving Musaceae.

<h3>*Asparagales*

Plastome-level analyses shift the position of seven families (including subfamilies formerly recognized as families in earlier APG versions) relative to the benchmark data set, and add the

two remaining subfamilies of Orchidaceae (Fig. 1). The only higher-level relationships within Asparagales now not strongly supported are those of Aphyllanthoideae to Agavoideae, and of Doryanthaceae to the clade subtended by Iridaceae and Agavaceae (Fig. 1). These branches are poorly supported, exceedingly short, and moderately deep (47.6 and 70.0 Mya, respectively) (Fig. 3 and Appendices S11, S13). *Aphyllanthes* has proven to be a "rogue taxon" in previous analyses (e.g., Graham et al., 2006), with its position unstable and its inclusion tending to reduce support values of several nearby branches. Replacing the 17 genes representing some astelid families (the clade subtended by Lanariaceae and Boryaceae in Asparagaless) with complete plastome gene sets should increase the support for several branches in that portion of the tree.

<h3>Liliales

Our data shift the position of two families relative to the benchmark: Smilacaceae sister to (Liliaceae, (Philesiaceae, Ripogonaceae)) rather than Liliaceae alone, and Melanthiaceae sister to the preceding four families rather than the clade subtended by Colchicaceae and Petermanniaceae (Fig. 1). The wholly mycoheterotrophic family Corsiaceae was added and is sister to Campynemataceae, in accord with analyses of 82 plastid genes by Mennes et al. (2015) and 77 plastid genes in Lam et al. (2018). Relationships within the order are identical to those obtained by Givnish et al. (2016b) using the same data but a reduced set of non-Liliales as outgroups. Only the sister relationship of Liliaceae to Philesiaceae-Ripogonaceae remains somewhat uncertain and requires further analyses based whole aligned plastomes.

<h3>Dioscoreales

Our findings excluding *Thismia* are consistent with those of Chase et al. (2006), but better supported (Fig. 1 and Appendix S11). When mycoheterotrophic *Thismia* is included, it is resolved as sister to *Tacca* with 61.0% bootstrap support (Appendix S12), making Dioscoreaceae s.l. (*sensu* APG IV, including *Tacca*) and Burmanniaceae s.l. paraphyletic. Almost surely the low support for the relationship of *Tacca* to *Thismia*—and the drop in support of nearby branches—reflects the unparalleled loss of plastid genes in *Thismia* and rapid rate of evolution of the surviving genes. These results provide some support for recognizing Taccaceae and Thismiaceae, and answer the call for further data testing this proposition (APG, 2016). Our results parallel those of a one-to-three gene analysis by Lam et al. (2016) and a plastid

phylogenomic analysis by Lam et al. (2018). The latter called for recognition of Thismiaceae as distinct from Burmanniaceae, and also recognized Taccaceae, consistent with the findings of Merckx et al. (2006, 2009) and Lam et al. (2016). We did, however, independently conduct the challenging alignment of *Thismia tentaculata* using our standard approaches (see above) and excluded *accD* whereas they included it. Inclusion of *accD* may account for the higher support (84–87% BS) for *Thismia*-Taccaceae in Lam et al. (2018), either because *acc*D is an information-rich gene or because difficulty in its alignment may have introduced artifacts under ML analysis with gaps included. Lam et al. (2016) sequenced three plastid genes (*accD*, *clpP*, *matK*, with one to three genes recovered per taxon) for 19 additional mycoheterotrophic species in five genera of Burmanniaceae in a monocot-wide survey of autotrophs and mycoheterotrophs, providing weak (<50%) bootstrap support for the monophyly of Burmanniaceae s.s. and each interfamilial relationship in (Burmanniaceae, (Dioscoreaceae s.s., (Taccaceae, Thismiaceae))). In addition, Merckx and Smets (2014) found that *Afrothismia* was sister to *Tacca* plus other Thismiaceae, based on ML and Bayesian analyses of sequences of nrDNA 18S and mitochondrial *atp1*. Thus, three families (Taccaceae, Thismiaceae, and an undescribed one including *Afrothismia*) may ultimately need to be segregated from Dioscoreaceae s.l., which would swell the number of monocot families to 80.

<h3>Pandanales

Our sampling adds mycoheterophic Triuridaceae to the benchmark data; *Sciaphila* is strongly supported (91.4% BS) as sister to Pandanaceae-Cyclanthaceae (Fig. 1 and Appendix S11), consistent with Lam et al. (2015) and a plastome-scale analysis by Lam et al. (2018). All other relationships are consistent with the analyses of the four-gene benchmark data.

<h3>Petrosaviales

The relationship of autotrophic *Japonolirion* to mycoheterotrophic *Petrosavia*, and their position as sister to all monocots except Alismatales and Acorales, has long been recognized and strongly supported, based initially on sequences of one plastid gene (Fuse and Tamura, 2000) or two plastid genes and one nrDNA gene (Chase et al., 2000; Soltis et al., 2000). Plastome-scale data support these relationships as well; see Fig. 1A and the plastome-scale analyses of Barrett et al. (2013), Logacheva et al. (2014), and Lam et al. (2018).

<h3>Alismatales

Plastome-scale data, partitioned ML analysis, and inclusion of exemplars for all families and genera of Alismatales s.s. (aquatic clade of Chase et al., 2006: Alismatales *sensu* APG IV excluding Araceae and Tofieldiaceae) provided 100% bootstrap for all interfamilial relationships within this group, consistent with the findings of Ross et al. (2016) based on the same sequences and their ML analysis involving codon × gene partitioning. Both represent major advances over the benchmark data, in terms of taxonomic inclusion and bootstrap support. Curiously, however, the plastome data yield much weaker (36%) support for Tofieldiaceae as sister to Alismatales s.s. than the benchmark data (Fig. 1). This may partly reflect the absence of five families of Alismatales s.s. from the benchmark data—especially Aponogetonaceae and Scheuchzeriaceae, which diverge from short branches at the two deepest nodes in the tepaloid clade (see Ross et al., 2016)—and partly the effects of analytical methodology and sporadic loss of the *ndhF* gene in both the petaloid and tepaloid clades of Alismatales s.s. and Tofieldiaceae. Ross et al. (2016) found that MP analysis (with or without *ndhF* included) and unpartitioned and most partitioned ML analyses resolved Araceae as sister to a core clade comprising all Alismatales except Tofieldiaceae, albeit with limited bootstrap support (54–76%). They found Tofieldiaceae to be sister to all other Alismatales only using the more powerful codon × gene partitioning if *ndhF* genes were included, but with only 48% bootstrap support. Computational limitations made it impossible for us to conduct a codon × gene partitioned ML analysis across a much broader and more numerous set of monocots. However, stronger support for a more stable arrangement of the three major clades in Alismatales (relationships among Araceae, Tofieldiaceae, and a core clade of Alismatales) might emerge if such analyses become more feasible.

<h2>Placement of mycoheterotrophic taxa

Chase et al. (2006) had difficulty in placing several mycoheterotrophic taxa (i.e., *Arachnitis* of Corsiaceae; *Burmannia* of Burmanniaceae; *Sciaphila* of Triuridaceae; *Thismia* of Thismiaceae), and all but *Burmannia* were excluded from the data set circulated afterward. Few details of the analyses including all these taxa were given. *Arachnitis* was placed sister to *Lilium* with weak support; *Burmannia* was sister to Dioscoreaceae with unspecified support; *Sciaphila* was placed in an uncertain position within Pandanales with unspecified support; and the placement of

Thismia was left unmentioned. Several of the mycoheterotrophic taxa were missing some or all of the plastid genes, and Chase et al. (2006) stated that there were problems with placing these taxa with mitochondrial and nuclear genes as well (several of these sequences may involve contaminants; Lam et al., 2016). Several authors have found nrDNA 18S to evolve several times faster in mycoheterotrophs than in related autotrophs (Nickrent and Starr, 1994; Nickrent et al., 1998; Merckx et al., 2006; for parasitic plants, see also Lemaire et al., 2011; Bromham et al., 2013). Rate heterogeneity associated with mycoheterotrophy is less marked in mitochondrial genes (Nickrent et al., 2002, 2004; Barkman et al., 2007), but can show quite substantial acceleration associated with sporadic transfers of mtDNA genes to the nucleus (Petersen et al., 2006).

Using plastid markers to infer the placement of heterotrophic plants has been viewed as problematic, especially given the likelihood of long branch attraction distorting phylogenetic inference—particularly under MP—as a result of greatly increased rates of substitution in many plastid genes due to relaxed selection, or outright loss of others (Merckx et al., 2009). Despite these concerns, several recent studies have demonstrated the feasibility of using plastome data to place mycoheterotrophic taxa congruent with outcomes based on mitochondrial and/or nuclear genes (Logacheva et al., 2014; Lam et al., 2015, 2018; Barrett et al., 2014a; Mennes et al., 2015). Givnish et al. (2015) used plastome sequences to place several mycoheterotrophic orchids in tribes of Orchidaceae to which they had previously been assigned based on morphology. Lam et al. (2016) succeeded in placing mycoheterophs in seven monocot families using the sequences of just three plastid genes.

Our success in placing mycoheterotrophs within a far broader sampling of photosynthetic monocots is notable, given the sevenfold higher rate of nucleotide substitutions we found in most mycoheterotrophs vs. their closest green relatives (see also Lam et al., 2016). Our ability to add to this list by placing *Thismia* sister to *Tacca* using plastome data—in agreement with earlier studies based on mitochondrial and nuclear sequences (Merckx et al., 2006, 2010) and the comparable analysis of Lam et al. (2018)—is remarkable, given that *Thismia* has lost all but 12 plastid genes (Lim et al., 2016) and these evolve 277× faster than autotrophs sister to other mycoheterophs, and 40.6× faster than other mycoheterotrophic lineages (see Results and Appendix S12). In terms of absolute branch length compared with autotrophic sister groups, *Thismia* is exceeded only by *Epigonium* (Orchidaceae: Epidendroideae: Gastrodieae) (Lam et al.,

2018). Despite this substantial rate elevation and gene loss, these mycoheterotrophic taxa can be placed with confidence using the intact genes they retain, demonstrated here and even more comprehensively in Lam et al. (2018).

<h2>Monocot timeline and rates of net species diversification

Our dating analysis places the divergence of monocots from the *Ceratophyllum*-eudicot lineage 136.1 Mya, in the Valanginian age of the Early Cretaceous, and the earliest divergence of extant monocot lineages from each other at 132.4 Mya, less than 4 My later. All monocot orders had diverged from each other within another 19 My, by 118 Mya in the Aptian age. This timeline is backed by 13 fossil and seven secondary calibrations. These findings should provide a vital resource for future studies of monocot evolution.

The strongly bimodal distribution of family stem ages—with a sharp trough between 70 and 90 Mya, combined with a trailing peak 60 to 50 Mya (Fig. 4)—raises the possibility that several newly arisen families vanished during the Cretaceous mass extinction 65 Mya, followed by a rebound in family origination shortly thereafter. Newly arisen families might, other things being equal, be less diverse and geographically more restricted than older families, and thus more susceptible to extinction by a general deterioration in environmental conditions. In fact, the dozen youngest monocot families (stem age 11.5–42 My, including Taccaceae and Thismiaceae) have only one-tenth as many species as older families, and 10 of 12 have relatively narrow geographic distributions (see Results). Mass extinction of families may have triggered the apparent origin of several families soon afterward, based on the opening up of resources and regions, or on creating morphological “gaps” through thinning of clades, leading to present-day recognition of lineages that might have been viewed simply as part of larger lineages in the absence of extinction of related forms (e.g., see Darwin, 1859; Tamura, 1998). Across all angiosperm families, Magallón et al. (2015) found a slight (but unnoted) depression in the numbers of families arising 60 to 70 Mya based on their BEAST analyses, but the pattern is much weaker than shown here across monocots.

Our dates for divergences at the ordinal level agree well with those of Magallón et al. (2015). A linear regression of their stem ages of monocot orders against our estimates yields a tight fit: $y = 2.47x - 192.9$ (Appendix S20a: $r^2 = 0.80$, $P < 0.001$, df = 8 to account for two pairs of sister orders); the relationship is even stronger if the single outlier (Asparagales) is excluded ($r^2 =$

0.95). We used seven secondary calibrations (six external to monocots) based on Magallón et al. (2015), but our 13 monocot fossil dates were independent of those used in that paper, pointing to remarkable consistency in the ages obtained by both analyses. By contrast, there is essentially no correlation of our ordinal stem ages and those reported by Tang et al. (2017) (Appendix S20b: $r = 0.145$, $P > 0.65$). However, that study—involving representatives of two-thirds of all monocot genera but based on only three plastid genes and nrDNA ITS—produced a RAxML tree with several polytomies and poorly supported nodes, which may limit its utility for dating and diversification analyses.

Based on our analyses, net species diversification underwent four significant, large-scale accelerations, in (1) the PACMAD-BOP clade of Poaceae, (2) Asparagales sister to Doryanthaceae (i.e., Asparagaceae, Amaryllidaceae, Asphodelaceae, Iridaceae, Xeronemataceae), (3) Orchidoideae-Epidendroideae of Orchidaceae, and (4) Araceae sister to Lemnoideae. These accelerations are associated with a number of factors. In the PACMAD-BOP clade, accelerations in net species diversification have been shown to be correlated with the repeated evolution of C4 photosynthesis in warm, seasonally dry habitats (Bouchenak-Khalladi et al., 2014; Spriggs et al., 2014) and high speciation rates of certain C3 lineages in cool grasslands and tropical forests (Spriggs et al., 2014). High ecological dominance of PACMAD-BOP grasses in open habitats is promoted by (1) moderate to tall stature, coupled with rhizomatous spread; (2) morphological adaptations (narrow erect leaves, heavy root allocation) to dry, sunny conditions; (3) a positive feedback among grasses, fire, and nitrogen, driven by the low N content of C_4 grasses (an inherent feature of their CO_2 concentrating mechanism) and resulting low rates of decomposition and high flammability, the volatilization of N during fires, and the low N requirement of C_4 grasses; and (4) the positive feedback between grasses and mammalian grazers, driven by the attractiveness of grasses to many grazers, their resistance to grazing damage conferred by basal meristems, and collateral damage to other plants caused by grazers (Givnish et al., 2010). Linder et al. (2017) used these hypotheses to argue that PACMAD-BOP grasses exhibit a “Viking syndrome,” coupling high rates of dispersal with traits that alter the environment to their advantage by promoting fire and mammalian grazing. The ecological advantage and relatively high speciation rates of bamboos remain unstudied. We suggest that the shrubby, multi-stemmed habit of woody bamboos is favored in frequently disturbed habitats (e.g., tropical montane forests on steep slopes) where they would have an

advantage in filling recent gaps by having more meristems active and more potential points for stem regeneration (Givnish, 1984; Götmark et al., 2016). Occurrence on steep tropical and subtropical mountains cut by heavy rainfall or tectonic activity could promote extensive speciation via the numerous barriers to dispersal created by a dissected, topographically complex terrain (Givnish et al., 2011, 2015, 2016).

High rates of net species diversification in the sister clade to Doryanthaceae within Asparagales appear to be associated, at a broad level, with the evolution of bulbous, cormous, or xeromorphic plants adapted to seasonally dry to arid habitats in the temperate zone. Shrublands and grasslands in winter rainfall regions of South Africa, the Mediterranean basin, California, and Australia are especially rich in bulbous or cormous members of Scilloideae, Brodeioideae, Lomandroideae, and certain Agavoideae (e.g., *Camassia*, *Hesperolirion*) of Asparagaceae, Amaryllidaceae, and Iridaceae. Large storage organs allow rapid leaf expansion and photosynthesis during almost all of the favorable season, at the expense of reduced growth in less seasonal habitats, and inability to persist in arid areas without predictable rainfall. Fire and grazing may further favor geophytes in seasonal habitats by reducing coverage of dominant grasses or woody plants (e.g., see Fragman and Schmida, 1997; Noy-Meir and Oron, 2001; Marques et al., 2017). CAM photosynthesis or thick, tough C3 leaves has evolved independently in Agavoideae, Lomandroideae, Nolinoideae, and Asphodelaceae. A unifying feature of these drought-adapted lineages is the dominance of dry capsules enclosing gravity-, wind-, or ant-dispersed seeds, in line with the general tendency for possession of fleshy fruits to increase toward rainier habitats, especially on nutrient-rich soils (Givnish, 2010). Limited distances of seed dispersal may be an overlooked driver of high rates of plant species diversification in semiarid to arid communities.

Acceleration of diversification in Orchidoideae and Epidendroideae of Orchidaceae is significantly coupled with the rise of distinct pollen packets (pollinia) in their common ancestor, and with the subsequent origin of epiphytism in the upper epidendroids sister to Tropideae-Nervilieae (Givnish et al., 2015). The evolution of pollinia should accelerate speciation by permitting precise placement of pollen and permitting specialization of different pollinators (e.g., bees vs. moths) or pollinator parts (e.g., proboscis or eyes) (Dressler, 1973; Sheviak and Bowles, 1986; Johnson et al., 1998; Van der Niet et al., 2014). Pollinia may also increase the importance of genetic drift, perhaps in alternation with strong selection on sexual characteristics (Tremblay

et al., 2005). In addition, pollinia—combined with numerous, tiny seeds—may allow very small numbers of variants in a population to produce large numbers of offspring, promoting speciation by fixing phenotypic differences from small numerical bases, and by preventing local demographic collapse and extinction (Givnish et al., 2015).

Epiphytism is a key innovation that allows invasion of a new adaptive zone (the branches and boles of trees and shrubs) largely unoccupied by other vascular plants, and should accelerate speciation and permit large numbers of species to coexist via partitioning of within-tree and across-habitat gradients. Epiphytism is also associated with high humidity and rainfall, and thus often with tropical montane habitats, which can provide numerous barriers to gene flow (e.g., high ridges, deep valleys) and thus foster local genetic differentiation and, ultimately, speciation (Givnish et al., 2015). Epiphytism and life in extensive tropical cordilleras is also significantly coupled to accelerated speciation in Bromeliaceae; diversification rates are especially high in the two epiphytic subfamilies Bromelioideae and Tillandsioideae (Givnish et al., 2011). Speciation rates in terrestrial Neotropical lobelioids are higher in the Andes than elsewhere, and higher in habitats >1900 m elevation than at lower elevations (Lagomarsino et al., 2016), supporting topographic dissection in montane habitats as an accelerator of speciation.

Most species of Araceae sister to Lemnoideae are epiphytes, hemiepiphytes, vines, or herbs of tropical rainforests and cloud forests and bear fleshy berries. Tropical understory plants with fleshy fruits have been hypothesized to speciate at high rates due to their dependence on relatively sedentary understory birds, leading to local differentiation and subsequent speciation (Givnish et al., 1995, 2009; Givnish, 2010). Theim et al. (2014) found that, as predicted, the spatial scale of gene flow estimated from population genetic structure was quite low (ca. 10–100 m) in four understory, fleshy-fruited species of the large eudicot genus *Psychotria* (Rubiaceae). In a survey of Neotropical understory lineages with fleshy fruits, Smith (2001) showed that 11 of 14 clades had more species than their dry-fruited sister clades. Ten of the 12 largest genera in the Hawaiian flora are fleshy-fruited plants of forest understories (Givnish, 1998, 2010); seven of the 11 largest Hawaiian clades are bird-dispersed denizens of wet-forest understories (Price and Wagner, 2004). The high rates of diversification in Araceae sister to Lemnoideae may thus reflect limited dispersal of fleshy fruits in tropical forest understories, as well as the topographic dissection of mountainous terrain occupied by epiphytic and hemiepiphytic taxa. High speciation rates in Araceae have not previously been reported, but the stem ages estimated for several large

genera by Nauheimer et al. (2012) and for *Philodendron* by Loss-Oliveira et al. (2016) are consistent with this hypothesis. Our estimates of the stem ages for several large genera—31 Mya for *Anthurium* (950 species), 29 Mya for *Philodendron* (482 species), 9.7 Mya for *Rhaphidophora* (105 species), and 3.1 Mya for *Alocasia* (78 species; all species counts from Boyce and Croat, 2018)—suggest that a more detailed analysis of Araceae might uncover several extraordinarily rapid diversifications nested within the higher aroids.

Three of the four broad accelerations of diversification we detected in monocots accord with one or more previous studies (Givnish et al., 2015; Spriggs et al., 2015; Magallón et al., 2018). In general, however, there is scale dependence in the resolution of such accelerations. For example, Givnish et al. (2015) found two additional accelerations of diversification nested within the orchidoid-epidendroid clade, associated with the evolution of epiphytism and, subsequently, the rise of deceit-based fly pollination. Spriggs et al. (2015) found an additional six accelerations within the PACMAD-BOP clade, five associated with C_4 lineages and one with the cold-adapted, C3 pooid grasses. Givnish et al. (2011) found two accelerations within Bromeliaceae not detected in the present study. In an across-angiosperms analysis, Magallón et al. (2018) detected five significant accelerations of diversification in monocots, corresponding to all Poaceae, all Orchidaceae, Cyperaceae and allies, Arecales + Commelinales + Zingiberales, and Asparagales sister to Doryanthaceae but including Tecophilaeaceae. That is, broader analyses—in this case and the others mentioned—often failed to detect important accelerations of net diversification at finer phylogenetic scales. Tang et al. (2017) is an exception to this rule but also encompasses the most comprehensive sampling of monocot genera. The general decline in numbers of diversification accelerations with the taxonomic breadth of studies may reflect a lower intensity of taxon sampling or fossil calibrations, or greater background-foreground contrasts in broader studies that impair ability to detect some accelerations. The accelerations identified by Magallón et al. (2018) but not by us may reflect their inability to obtain correct tree resolution with far more limited sequence data (three plastid genes, 18S, 26S nrDNA): Arecales are not sister to Commelinales-Zingiberales, and Tecophilaeaceae are not part of the Asparagales clade sister to Doryanthaceae (Fig. 1).

<h2>Resampling studies

As predicted, branch ascertainment and bootstrap support increase with number of genes sampled, inclusion of non-coding regions, and branch length; and decrease with relative branch depth (Figs. 6 and 7; Appendix S17). The increases in ascertainment and support with the number of genes sampled across monocots are substantial (Fig. 7) and provide compelling evidence of the power of phylogenomic vs. phylogenetic approaches, given the actual level of variation and homoplasy seen in monocots. Such increases in ascertainment and support with the number of genes sampled are least for long shallow branches (which even modest amounts of data can correctly resolve and strongly support) and short deep branches (which even large amounts of data may fail to resolve with certainty). These findings, however, were obtained with a relatively dense sampling of taxa across major monocot lineages and must be seen in that context. Given our high density of taxon sampling, and extensive sampling of plastid gene sequences, other studies that sample far fewer loci across the same or greater taxon density (e.g., Bouchenak-Khelladi et al. [2014] for Poales) may produce less reliable results. The approach taken by Saarela et al. (2018) for inferring relationships in Poaceae from plastid data is exemplary: they used whole aligned plastomes sampled across a dense sampling of taxa (250 species—2.5× that used here) and analyzed using codon × gene partitioned ML. Such an approach is a step beyond the analysis conducted here, albeit for a much narrower taxonomic group, and should be replicated when possible.

Inclusion of non-coding regions in our Zingiberales-focused analysis shifted the position of Heliconiaceae and Musaceae in relation to our partitioned ML analysis of 77 plastid genes, earlier analyses based on a few genes with (or without) a few morphological characters for a small number of taxa (Kress et al., 2001), or analysis of whole aligned plastomes (Barrett et al., 2014b). Sass et al. (2016) produced 100% bootstrap support for all interfamily relationships in Zingiberales using MP or codon × gene partitioned ML applied to 68 plastid genes and 308 single-copy nuclear genes. Support for the position of Musaceae sister to all other families, however, disappeared in their coalescent tree. This conflict among genes in placing Musaceae, and the conflict between the strongly supported positions for Musaceae in Sass et al. (2016) vs. the present study, points to a likely incongruence in the taxonomic signal between the plastome and nuclear genomes, and perhaps a deep reticulation event close to the origin of the order Zingiberales. A meticulous analysis of incongruence between the plastid and nuclear data sets for Zingiberales should now be conducted to test this idea.

<h1>CODA

Plastomes offer a quantum leap over individual plastid genes in the amount of sequence data and provide the basis for far more powerful analyses of plant phylogeny and evolution. However, plastome sequences putatively reflect only maternal lines of ancestry and ultimately must be compared with sequences from the biparentally inherited nuclear genome to resolve challenges posed by hybridization, introgression, incomplete lineage sorting, and horizontal gene transfer (e.g., Linder and Rieseberg, 2004; Willyard et al., 2009; Sessa et al., 2012; García et al., 2014; Davis and Xi, 2015; Vargas et al., 2017). Sequences of hundreds to thousands of nuclear loci can now be obtained via transcriptomes, targeted enrichment, genome skimming, and whole genome sequencing (Grover et al., 2012; Lemmon et al., 2012; Wickett et al., 2014; Zimmer and Wen, 2015; Cardillo et al., 2017; Leveille-Bouret et al., 2018; Morris et al., 2018). Such data are required to screen for reticulation events deep in phylogenies and will provide additional bases for reconstructing phylogenetic trees and networks. However, the far greater information content of the nuclear vs. plastid genome must be weighed against the greater uncertainty in identifying homologous nuclear loci (Springer and Gatesy, 2018)—or detecting paralogy, including cryptic paralogy, in which different gene copies have been lost in different taxa—given how extensive gene duplications and losses are in the nuclear genome.

Finally, based on our plastome phylogeny and some functional insights, we advance a new hypothesis that the ancestral monocots were submersed aquatic plants, or amphibious aquatic plants that produced submersed and emergent foliage at different ages or at different depths. Several authors have previously suggested that monocots may have had an aquatic origin (Henslow, 1893, 1911; Hallier, 1905; Arber, 1925; Cronquist, 1968, 1981; Takhtajan, 1969, 1991; Stebbins, 1974; Duvall et al., 1993b; see important review by Les and Schneider, 1995). These early proposals, however, were highly problematic. All but the last were based on evidence and arguments that we would not now find compelling—for example, that many monocots are aquatic plants, that they arose via Lamarckian or wholly nonadaptive evolution, that character states reflect ancestry rather than ecology, that some present-day monocot lineages are ancestral to others, and that the functional significance of traits can somehow be inferred on the basis of phenotypic similarity alone, in the absence of any phylogenetic or functional analyses.

Chase (2004) noted that the aquatic tendencies of Acorales and Alismatales, sequentially sister to all remaining monocots, suggested an aquatic monocot ancestor. But this brief suggestion did not include critical details regarding the phylogenetic positions and specific kinds of aquatic plants within Alismatales or monocot sister groups. Aquatic plants are rooted on substrates that are flooded during all or much of the growing season, including *emergent species* (which hold their leaves above the water's surface), *floating-leaved species*, and submersed species (Sculthorpe, 1967). Many aquatic plants, regardless of habit, have submersed seedlings (Du et al., 2016). Our phylogeny implies an aquatic plant as ancestral to the monocots, given (1) the submersed habit of *Ceratophyllum*, one of the pair of lineages sister to the monocots, and the upland habitat implied as ancestral to the other, given the distribution of ecologies in the basal grade of the eudicots; (2) the emergent habit of *Acorus*; (3) the emergent, floating, and submersed habits of the former Najadales (Alismatales minus Araceae and Tofieldiaceae), with a predominance of emergent and floating species among the lineages sequentially sister to other members of both the tepaloid and sepaloid clades, given the deep nodes of Aponogetonaceae characterized by the emergent and floating habit (for phylogeny, see Chen et al., 2015) and the emergent habit of Scheuchzeriaceae, Juncaginaceae, and Maundiaceae, and the emergent/submersed amphibious habit of genera of Alismataceae, then emergent habit of *Butomus*; (4) the emergent/wetland habit of Tofieldiaceae; and (5) the floating habit of Orontioideae, then Lemnoideae, successively sister to all members of Araceae (see Fig. 1 and Appendix S11).

Plant life underwater should be shaped by two fundamental physical constraints: (1) that CO_2 and O_2 diffuse 10,000× more slowly in water than in air, and (2) that water is far denser than air and incompressible. The first constraint favors the evolution of very narrow, deeply divided, or fenestrate leaves in submersed plants, to reduce leaf boundary-layer thickness and increase photosynthesis (Givnish, 1979). We posit that, of these options, narrow linear leaves characterized the aquatic lineage ancestral to the monocots. Buoyancy of leaves underwater should remove selection for branched venation to provide mechanical support and favor parallel venation (Givnish, 1979). In addition, the lack of transpiration underwater and the very low rates of photosynthesis and consequent return through the phloem should favor one or few highly reduced veins in linear leaves. Tensile forces pulling in various directions on leaves underwater should select for clasping leaf bases; in contrast, above the water line, resisting flexure under

gravity should favor the coalescence of support tissue into thick midribs and petioles (Givnish, 1995).

Underwater leaves with clasping bases (a monocot hallmark; see Cronquist, 1981; Dahlgren et al., 1985) and single veins—characteristics of our proposed aquatic monocot ancestor—have evolved convergently in the submerged plants of Hydatellaceae of Nymphaeales, another early-divergent, non-monocot angiosperm lineage (Sokoloff et al., 2009). The many-faceted resemblance of these plants to several submersed monocots (e.g., *Najas*, *Potamogeton*) is so great that, in the absence of critical molecular data, they were long thought to be monocots (Saarela et al., 2007). These plants also strongly resemble several aquatic non-monocots (e.g., *Littorella* [eudicot Plantaginaceae] and *Isoetes* [lycopod Isoetaceae]) with an isoetid habit, marked by a compact rosette of narrow, stiff leaves (Boston, 1986; Keeley, 1999). Selection for broader leaves (or leaf-like organs) from such ancestral forms in other contexts (e.g., above water in amphibious plants late in life, in shallow water, or during dry periods) might then simply be accompanied by a multiplication of parallel veins. However, in especially broad leaves with thin cross-sections—favored in forest understories—mechanical loading should favor the evolution of branched venation. Monocots have re-evolved such venation ≥20 times, strongly associated (as predicted on functional grounds) with the invasion of shady habitats (Givnish et al., 2005).

Slow diffusion underwater leads to anoxic substrates. This should favor the evolution of the herbaceous habit, because woodiness and regular secondary thickening would regularly cut off aerenchyma that could carry oxygen from the leaves to the roots. Wave action should also favor the loss of woody stems and the development of slender, herbaceous stems that can bend without breaking (Givnish, 1995). Loss of secondary thickening would free the xylem and phloem from being concentrated in a circumferential cambium, and possible damage to stem surfaces by herbivores, debris, or other plants might favor dispersal of vascular tissue throughout the stem cross-section—another monocot synapomorphy. Anoxic substrates should select against persistent, deep-delving roots and favor shallow primary roots; secondary root thickening might cut off tissues from oxygen provided by aerenchyma. Constant movement of the substrate by dense, incompressible water could bury or excavate seedlings, selecting for vegetative spread via sympodial growth and adventitious rooting. Each of these monocot hallmarks has evolved, as

expected on functional grounds, in Hydatellaceae (for morphological and anatomical descriptions, see Sokoloff et al., 2008, 2009).

Thus, life as a submersed aquatic could favor the evolution of a "monocot syndrome" comprising several characteristic traits: narrow leaves, parallel venation, clasping leaf bases, loss of secondary thickening, scattered vascular bundles, sympodial growth, shallow primary roots, and adventitious rooting. Production of broader aerial leaves in later growth phases of an amphibious species, or in relatives invading shallower water, could yield foliage with multiple parallel veins, and possibly production of unifacial leaves through retention of central aerenchyma. Some of these traits may well have resulted in the downstream evolution of other traits and ecologies characteristic of monocots. For example, mycoheterotrophy—a trait that monocots seem especially prone to evolving (Imhoff, 2010; Merckx et al., 2013)—could have arisen ≥35 times in monocots due to their production of primary roots only, given that myco-heterotrophy appears to arise with the herbaceous habit and voluminous primary root parenchyma, usually obliterated by secondary root growth in non-monocot angiosperms (Imhoff, 2010; Lam et al., 2018). Bulbs—an organ nearly restricted to monocots, and key as an adaptation to permit photosynthesis during almost all of the short period of favorable conditions in seasonal habitats—are a natural development from clasping leaf bases, involving a shortening of internodes and increased carbon storage in those now overlapping leaf bases. Production of leaves with parallel veins naturally can lead to the development of basal meristems that continuously unspool new lengths of leaf tissue, serving as an adaptation to grazing or fire in graminoids.

Finally, what about the origin of the single cotyledon that characterizes monocots? We hesitate to argue that the single cotyledon has strong selective value, given that the trait has been rigorously conserved across monocots despite their invasion of an extremely wide range of environments. But single cotyledons can permit the formation of a cotyledonary tube, and Stebbins (1974) argued that such a tube could be adaptive in dry habitats by pushing the seed deep into the soil. We believe that such "planting" behavior might be even more important in seedling establishment in shallow water, where seeds on the substrate surface could easily be washed away. Cotyledonary tubes that push seeds down deep into the substrate—or anchor them, by themselves pushing into the substrate—might well be adaptive. Again, we point to the highly reduced cotyledon of aquatic Hydatellaceae as evidence for this hypothesis; apparently, this

structure pushes deep into the substrate after germination (Sokoloff et al., 2008). Hydatellaceae may have one bilobate cotyledon or two fused cotyledons, but functionally it appears to have just one, and this is supportive of our hypothesis. The same is true for Nymphaeaceae, its immediate sister and another aquatic group with submersed seedlings. Whatever the value of single cotyledons, narrow leaves, parallel venation, loss of secondary thickening, scattered bundles, sympodial growth, and adventitious production of primary roots may have been, with a stable, dated, highly inclusive phylogeny for monocots, we now have a phylogeny, timeline, and conceptual framework with which to pursue the early evolution of the monocots.

<h1>ACKNOWLEDGEMENTS

This research was supported by National Science Foundation grants DEB 0830036 to T.J.G., C.A., and S.W.G.; DEB 0830020 to J.I.D. and M.A.G.; DEB 0830009 to J.L.-M. and W.B.Z.; DEB 0829849 to J.C.P.; DEB 0829762 to D.W.S.; and DEB 1257701 to C.D.S.; and by NSERC (Natural Sciences and Engineering Research Council) Discovery grants to S.W.G. We thank P. Rudall, P. Stevens, B. Tomlinson, and especially A. Litt for helpful discussions regarding the origin of bulbs. K. Bi at the Computational Genomics Resource Lab at UC Berkeley assisted in cleaning raw sequence data in preparation for plastome assembly. S. Friedrich produced the beautiful figures. We are grateful to B. Briggs and an anonymous individual for their meticulous reviews.

<h1>DATA ACCESSIBILITY

Sequence alignments, heat map for missing sequence data, analyses, and code are available on the Open Science Framework at https://osf.io/g4me6/. Newly generated sequences and vouchers have been lodged in GenBank; accession numbers are displayed in Appendix S3.

<h1>LITERATURE CITED

Angiosperm Phylogeny Group (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* 161: 105–121.

Angiosperm Phylogeny Group (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* 181: 1–20.

Arber, A. (1925). Monocotyledons: A morphological study. Cambridge University Press, Cambridge, UK.

Attigala, L., W. P. Wysocki, M. R. Duvall, and L. G. Clark (2016). Phylogenetic estimation and morphological evolution of Arundinarieae (Bambusoideae: Poaceae) based on plastome phylogenomic analysis. *Molecular Phylogenetics and Evolution* 101: 111–121.

Bacon, C. D., W. J. Baker, and M. P. Simmons (2012). Miocene dispersal drives island radiations in the palm tribe Trachycarpeae (Arecaceae). *Systematic Biology* 61: 426–442.

Barkman, T. J., J. R. McNeal, S. H. Lim, G. Coat, H. B. Croom, N. D. Young, and C. W. dePamphilis (2007). Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. *BMC Evolutionary Biology* 7: 248.

Barrett, C. F., J. Comer, J. Leebens-Mack, J. Li, D. R. Mayfield-Jones, J. R. Medina, L. Perez, et al. (2016). Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. *New Phytologist* 209: 855–870.

Barrett, C. F., J. I. Davis, J. Leebens-Mack, J. G. Conran, and D. W. Stevenson (2013). Plastid genomes and deep support among the commelinid monocot angiosperms. *Cladistics* 29: 65–87.

Barrett, C. F., J. V. Freudenstein, J. Li, D. R. Mayfield-Jones, L. Perez, J. C. Pires, and C. Santos (2014a). Investigating the path of plastid genome degradation in an early-transitional clade of mycoheterotrophic orchids, and implications for heterotrophic angiosperms. *Molecular Biology and Evolution* 31: 3095–3112.

Barrett, C. F., C. D. Specht, J. Leebens-Mack, D. W. Stevenson, W. B. Zomlefer, and J. I. Davis (2014b). Resolving ancient radiations: Can complete plastid gene sets elucidate deep relationships among the tropical gingers (Zingiberales Griseb.)? *Annals of Botany* 113: 119–133.

Bouchenak-Khelladi, Y., A. M. Muasya, and H. P. Linder (2014). A revised evolutionary history of Poales: orIgins and diversification. *Botanical Journal of the Linnean Society* 175: 4–16.

Bouckaert, R. R., J. Heled, D. Kuehnert, T. G. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, et al. (2014). BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 10: e1003537.

Boston, H. L. (1986). A discussion of the adaptation for carbon acquisition in relation to the growth strategy of aquatic isoetids. *Aquatic Botany* 26: 259–270.

Boyce, P. C., and T. B. Croat (2018). The Überlist of Araceae, totals for published and estimated number of species in aroid genera. http://www.aroid.org/genera/180211uberlist.pdf.

Briggs, B. G., A. D. Marchant, and A. J. Perkins (2014). Phylogeny of the restiid clade (Poales) and implications for the classification of Anarthriaceae, Centrolepidaceae and Australian Restionaceae. *Taxon* 63: 24–46.

Bromham, L., P. F. Cowman, and R. Lanfear (2013). Parasitic plants have increased rates of molecular evolution across all three genomes. *BMC Evolutionary Biology* 13: 126.

Burke, S. V., L. G. Clark, J. K. Triplett, C. P. Grennan, and M. R. Duvall (2014). Biogeography and phylogenomics of New World Bambusoideae (Poaceae), revisited. *American Journal of Botany* 101: 886–891.

Burke, S. V., C. P. Grennan, and M. R. Duvall (2012). Plastome sequences of two New World bamboos—*Arundinaria gigantea* and *Cryptochloa strictiflora* (Poaceae)—extend phylogenomic understanding of Bambusoideae. *American Journal of Botany* 99: 1951–1961.

Burke, S. V., C. S. Lin, W. P. Wysocki, L. G. Clark, and M. R. Duvall (2016a). Phylogenomics and plastome evolution of tropical forest grasses (*Leptapsis, Streptochaeta*: Poaceae). *Frontiers in Plant Science* 7: 1993.

Burke, S. V., W. P. Wysocki, F. O. Zuloaga, J. M. Craine, J. C. Pires, P. P. Edger, D. Mayfield-Jones, et al. (2016b). Evolutionary relationships in panicoid grasses based on plastome phylogenomics (Panicoideae; Poaceae). *BMC Plant Biology* 16: 140.

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

Cardillo, M., P. H. Weston, Z. K. M. Reynolds, P. M. Olde, A. R. Mast, E. M. Lemmon, A. R. Lemmon, and L. Bromham (2017). The phylogeny and biogeography of *Hakea* (Proteaceae) reveals the role of biome shifts in a continental plant radiation. *Evolution* 71: 1928–1943.

Chase, M. W., K. M. Cameron, R. L. Barrett, and J. V. Freudenstein (2003). DNA data and Orchidaceae systematics: A new phylogenetic classification. *In* K. W. Dixon, S. P. Kell, R. L. Barrett, and P. J. Cribb [eds.], Orchid conservation, 69–89. Natural History Publications, Kota Kinabalu, Malaysia.

Chase, M. W., M. F. Fay, D. S. Devey, O. Maurin, N. Rønsted, T. J. Davies, Y. Pillon, et al. (2006). Multigene analyses of monocot relationships: A summary. *Aliso* 22: 63–75.

Chase, M. W., D. E. Soltis, R. G. Olmstead, D. Morgan, D. H. Les, B. D. Mishler, M. R. Duvall, et al. (1993). Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbcL*. *Annals of the Missouri Botanical Garden* 80: 528–580.

Chase, M. W., P. S. Soltis, P. J. Rudall, M. F. Fay, W. H. Hahn, S. Sullivan, J. Joseph, et al. (2000). Higher-level systematics of the monocotyledons: An assessment of current knowledge and a new classification. *In* K. L. Wilson and D. A. Morrison [eds.], Monocots: systematics and evolution, 3–16. CSIRO, Collingwood, Australia.

Chen, L. Y., G. W. Grimm, Q. F. Wang, and S. S. Renner (2015). A phylogeny and biogeographic analysis for the Cape-pondweed family Aponogetonaceae (Alismatales). *Molecular Phylogenetics and Evolution* 82: 111–117.

Comer, J. R., W. B. Zomlefer, C. F. Barrett, J. I. Davis, D. W. Stevenson, K. Heyduk, and J. H. Leebens-Mack (2015). Resolving relationships within the palm subfamily Arecoideae (Arecaceae) using plastid sequences derived from next-generation sequencing. *American Journal of Botany* 102: 888–899.

Cotton, J. L., W. P. Wysocki, L. G. Clark, S. A. Kelchner, J. C. Pires, P. P. Edger, D. Mayfield-Jones, and M. R. Duvall (2015). Resolving deep relationships of PACMAD grasses: A phylogenomic approach. *BMC Plant Biology* 15: 178.

Darriba, D., G. L. Taboada, R. Doallo, and D. Posada (2012). jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods* 9: 772–772.

Davis, C. C., and Z. Xi (2015). Horizontal gene transfer in parasitic plants. *Current Opinion in Plant Biology* 26: 14–19.

Dressler, R. L. (1973). Phylogeny and classification of the orchid family. Cambridge University Press, Cambridge, UK.

Drew, B. T., B. R. Ruhfel, S. A. Smith, M. J. Moore, B. G. Briggs, M. A. Gitzendanner, P. S. Soltis, and D. E. Soltis (2014). Another look at the root of the angiosperms reveals a familiar tale. *Systematic Biology* 63: 368–382.

Duvall, M. R., M. T. Clegg, M. W. Chase, W. D. Clark, W. J. Kress, H. G. Hills, L. E. Eguiarte, et al. (1993a). Phylogenetic hypotheses for the monocotyledons constructed from *rbcL* sequence data. *Annals of the Missouri Botanical Garden* 80: 607–619.

Duvall, M. R., A. E. Fisher, J. T. Columbus, A. L. Ingraham, W. P. Wysocki, S. V. Burke, L. G. Clark, and S. A. Kelchner (2016). Phylogenomics and plastome evolution of the chloridoid grasses (Chloridoideae: Poaceae). *International Journal of Plant Science* 177: 235–246.

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

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797.

Fragman, O., and A. Schmida (1997). Diversity and adaptation of wild geophytes along an aridity gradient in Israel. *Acta Horticulturae* 430: 795–802.

García, M. A., M. Costea, M. Kuzmina, and S. Stefanovic (2014). Phylogeny, character evolution, and biogeography of *Cuscuta* (dodders; Convolvulaceae) inferred from coding plastid and nuclear sequences. *American Journal of Botany* 101: 670–690.

Givnish, T. J. (1979). On the adaptive significance of leaf form. *In* O. T. Solbrig, S. Jain, G. B. Johnson, and P. H. Raven [eds.], Topics in plant population biology, 375–407. Columbia University Press, New York, New York, USA.

Givnish, T. J. (1984). Leaf and canopy adaptations in tropical forests. *In* E. Medina, H. A. Mooney, and C. Vásquez-Yánes [eds.], Physiological ecology of plants of the wet tropics, 51–84. Dr. Junk, The Hague.

Givnish, T. J. (1995). Plant stems: Biomechanical adaptations for energy capture and influence on species distributions. *In* B. L. Gartner [ed.], Plant stems: Physiology and functional morphology, 3–49. Chapman and Hall, New York, New York, USA.

Givnish, T. J. (1998). Adaptive radiation of plants on oceanic islands: Classical patterns, molecular data, new insights. *In* P. Grant [ed.], Evolution on islands, 281–304. Oxford University Press, New York, New York, USA.

Givnish, T. J. (2010). Ecology of plant speciation. *Taxon* 59: 1326–1366.

Givnish, T. J., M. Ames, J. R. McNeal, P. R. Steele, C. W. dePamphilis, S. W. Graham, J. C. Pires, et al. (2010). Assembling the tree of the monocotyledons: Plastome sequence phylogeny and evolution of Poales. *Annals of the Missouri Botanical Garden* 97: 584–616.

Givnish, T. J., M. H. J. Barfuss, B. Van Ee, R. Riina, K. Schulte, R. Horres, P. A. Gonsiska, et al. (2011). Phylogeny, adaptive radiation, and historical biogeography in Bromeliaceae: Insights from an 8-locus plastid phylogeny. *American Journal of Botany* 98: 872–895.

Givnish, T. J., T. M. Evans, J. C. Pires, and K. J. Sytsma (1999). Polyphyly and convergent evolution in Commelinales and Commelinidae: Evidence from *rbcL* sequence data. *Molecular Phylogenetics and Evolution* 12: 360–385.

Givnish, T. J., K. C. Millam, T.T. Theim, A. R. Mast, T. B. Patterson, A. L. Hipp, J. M. Henss, et al. (2009). Origin, adaptive radiation, and diversification of the Hawaiian lobeliads (Asterales: Campanulaceae). *Proceedings of the Royal Society of London B* 276: 407–416.

Givnish, T. J., J. C. Pires, S. W. Graham, M. A. McPherson, T. B. Patterson, H. S. Rai, T. M. Evans, et al. (2005). Repeated evolution of net venation and fleshy fruits among monocots in shady habitats confirms a priori predictions: Evidence from an *ndhF* phylogeny. *Proceedings of the Royal Society B* 272: 1481–1490.

Givnish, T. J., D. Spalink, M. Ames, S. P. Lyon, S. J. Hunter, A. Zuluaga, M. A. Clements, et al. (2015). Orchid phylogenomics and multiple drivers of extraordinary diversification. *Proceedings of the Royal Society of London B* 282: 20151553.

Givnish, T. J., D. Spalink, M. Ames, S. P. Lyon, S. J. Hunter, A. Zuluaga, M. A. Clements, et al. (2016a). Orchid historical biogeography, diversification, Antarctica, and the paradox of orchid dispersal. *Journal of Biogeography* 43: 1905–1916.

Givnish, T. J., and K. J. Sytsma (1997). Consistency, characters, and the likelihood of correct phylogenetic inference. *Molecular Phylogenetics and Evolution* 7: 320–330.

Givnish, T. J., K. J. Sytsma, J. F. Smith, and W. S.Hahn (1995). Molecular evolution, adaptive radiation, and geographic speciation in *Cyanea* (Campanulaceae, Lobelioideae). *In* W. L. Wagner and V. Funk [eds.], Hawaiian biogeography: Evolution on a hot spot archipelago, 288–337. Smithsonian Institution Press, Washington, D.C., USA.

Givnish, T. J., A. Zuluaga, V. K. Y. Lam, M. S. Gomez, W. J. D. Iles, D. Spalink, J. R. Moeller, et al. (2016b). Plastome phylogeny and historical biogeography of the monocot order Liliales: Out of Australia and through Antarctica. *Cladistics* 32: 581–605.

Götmark, F., E. Götmark, and A. M. Jensen (2016). Why be a shrub? A basic model and hypotheses for the adaptive values of a common growth form. *Frontiers in Plant Science* 7: 1095.

Graham, S. W., J. M. Zgurski, M. A. McPherson, D. M. Cherniawsky, J. M. Saarela, E. F. C. Horne, S. Y. Smith, et al. (2006). Robust inference of monocot deep phylogeny using an expanded multigene plastid data set. *Aliso* 22: 3–21.

Grass Phylogeny Working Group II (2012). New grass phylogeny resolves deep evolutionary relationships and discovers C4 origins. *New Phytologist* 193: 304–312.

Grover, C. E., A. Salmon, and J. F. Wendel (2012). Targeted sequence capture as a powerful tool for evolutionary analysis. *American Journal of Botany* 99: 312–319.

Hallier, H. (1905). Provisional scheme of the natural (phylogenetic) system of flowering plants. *New Phytologist* 4: 151–162.

Henriquez, C. L., T. Arias, J. C. Pires, T. B. Croat, and B. A. Schaal (2014). Phylogenomics of the plant family Araceae. *Molecular Phylogenetics and Evolution* 75: 91–102.

Henslow, G. (1893). A theoretical origin of endogens from exogens through self-adaptation to an aquatic habit. *Botanical Journal of the Linnean Society* 29: 485–528.

Henslow, G. (1911). The origin of monocotyledons from dicotyledons through self-adaptation to a moist or aquatic habit. *Annals of Botany* 26: 717–744.

Huang, Y. L., X. J. Li, Z. Y. Yang, C. J. Yang, J. B. Yang, and Y. H. Ji (2016). Analysis of complete chloroplast genome sequences improves phylogenetic resolution in *Paris* (Melanthiaceae). *Frontiers in Plant Science* 7: 1797.

Iles, W. J. D, S. Y. Smith, M. A. Gandolfo, and S. W. Graham (2015). Monocot fossils suitable for molecular dating analyses. *Botanical Journal of the Linnean Society* 178: 364–374.

Imhoff, S. (2010). Are monocots particularly suited to develop mycoheterotrophy? *In* O. Seberg, G. Petersen, A. Barfod, and J. I. Davis [eds.], Diversity, phylogeny, and evolution in the monocotyledons, 11–23. Aarhus University Press, Aarhus, Denmark.

Johansen, L. B. (2005). Phylogeny of *Orchidantha* (Lowiaceae) and the Zingiberales based on six DNA regions. *Systematic Botany* 30: 106–117.

Johnson, S. D., H. P. Linder, and K. E. Steiner (1998). Phylogeny and radiation of pollination systems in *Disa* (Orchidaceae). *American Journal of Botany* 85: 402–411.

Jones, S. S., S. V. Burke, and M. R. Duvall (2014). Phylogenomics, molecular evolution, and estimated ages of lineages from the deep phylogeny of Poaceae. *Plant Systematics and Evolution* 300: 1421–1436.

Katoh, S. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S. et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647–1649.

Keeley, J. E. (1999). Photosynthetic pathway diversity in a seasonal pool community. *Functional Ecology* 13: 106–118.

Kim, J. H., S. I. Lee, B. R. Kim, I. Y. Choi, P. Ryser, and N. S. Kim (2017). Chloroplast genomes of *Lilium lancifoliu*, *L. amabile*, *L. callosum*, and *L. philadelphicum*: Molecular characterization and their use in phylogenetic analysis in the genus *Lilium* and other allied genera in the order Liliales. *PLoS ONE* 12: e0186788.

Klopfstein, S., C. Kropf, and D. Quicke (2010). An evaluation of phylogenetic informativeness profiles and the molecular phylogeny of Diplazontinae (Hymenoptera, Ichneumonidae). *Systematic Biology* 59: 226–241.

Klopfstein, S., T. Massingham, and N. Goldman (2017). More on the best evolutionary rate for phylogenetic analysis. *Systematic Biology* 66: 769–785.

Kress, W. J. (1990). The phylogeny and classification of the Zingiberales. *Annals of the Missouri Botanical Garden* 77: 698–721.

Kress, W. J., L. M. Prince, W. J. Hahn, and E. A. Zimmer (2001). Unraveling the evolutionary radiation of the families of the Zingiberales using morphological and molecular evidence. *Systematic Biology* 50: 926–944.

Lagomarsino, L. P., F. L. Condamine, A. Antonelli, A. Mulch, and C. C. Davis (2016). The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). *New Phytologist* 210: 1430–1442.

Lam, V. K. Y., H. Darby, V. S. F. T. Merckx, G. Lim, T. Yukawa, K. N. Neubig, J. R. Abbott, et al. (2018). Phylogenomic inference *in extremis*: A case study with mycoheterotroph plastomes. *American Journal of Botany* 105: 480–494.

Lam, V. K. Y., M. S. Gomez, and S. W. Graham (2015). The highly reduced plastome of mycoheterotrophic *Sciaphila* (Triuridaceae) is collinear with its green relatives and is under strong purifying selection. *Genome Biology and Evolution* 7: 2220–2236.

Lam, V. K. Y., V. S. F. T. Merckx, and S. W. Graham (2016). A few-gene plastid phylogenetic framework for mycoheterotrophic monocots. *American Journal of Botany* 103: 692–708.

Lanfear, R., P. B. Frandsen, A. M. Wright, T. Senfeld, and B. Calcott (2017). PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution* 34: 772–773.

Lemaire, B., S. Huysmans, E. Smets, and V. Merckx (2011). Rate accelerations in nuclear 18S rDNA of mycoheterotrophic and parasitic angiosperms. *Journal of Plant Research* 124: 561–576.

Lemmon, A. R., S. A. Emme, and E. M. Lemmon (2012). Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology* 61: 727–744.

Les, D. H., and E. L. Schneider (1995). The Nymphaeales, Alismatidae, and the theory of an aquatic monocotyledon origin. *In* P. J. Rudall, P. J. Cribb, D. F. Cutler, and C. J. Humphries [eds.], Monocotyledons: Systematics and evolution, 23–42. Royal Botanic Gardens, Kew, UK.

Leveille-Bouret, E., J. R. Starr, B. A. Ford, E. M. Lemmon, and A. R. Lemmon (2018). Resolving rapid radiations within angiosperm families using anchored phylogenomics. *Systematic Biology* 67: 94–112.

Lim, G. S., C. F. Barrett, C. C. Pang, and J. I. Davis (2016). Drastic reduction of plastome size in mycoheterotrophic *Thismia tentaculata* relative to that of its autotrophic relative *Tacca chantrieri*. *American Journal of Botany* 103: 1129–1137.

Linder, H. P., C. E. R. Lehmann, S. Archibald, C. P. Osborne, and D. M. Richardson (2017). Global grass (Poaceae) success underpinned by traits facilitating colonization, persistence and habitat transformation. *Biological Reviews* 93: 1125–1144.

Linder, H. P., and L. H. Rieseberg (2004). Reconstructing patterns of reticulate evolution in plans. *American Journal of Botany* 91: 1700–1708.

Logacheva, M. D., M. I. Schelkunov, M. S. Nuraliev, T. H. Samigullin, and A. A. Penin (2014). The plastid genome of mycoheterotrophic monocot *Petrosavia stellaris* exhibits both gene losses and multiple rearrangements. *Genome Biology and Evolution* 6: 238–246.

Loss-Oliveira, L., C. Sakuragui, M. de Lourdes Soares, and C. G. Schrago (2016). Evolution of *Philodendron* (Araceae) species in Neotropical biomes. *PeerJ* 4: e1744.

Lughadha, E. N., R. Govaerts, I. Belyaeva, N. Black, H. Lindon, R. Allkin, R. E. Magill, and N. Nicolson (2016). Counting counts: revised estimates of numbers of accepted species of flowering plants, seed plants, vascular plants and land plants with a review of other recent estimates. *Phytotaxa* 272: 82–88.

Magallón, S., S. Gómez-Acevedo, L. L. Sánchez-Reyes, and T. Hernández-Hernández (2015). A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist* 207: 437–453.

Magallón, S., L. L. Sánchez-Reyes, and S. L. Gómez-Acevodo (2018). Thirty clues to the exceptional diversification of flowering plants. *bioRxiv preprint*, doi: http://dx.doi.org/10.1101/279620.

Marques, I., J. F. Aguilar, M. A. Martins-Louçao, F. Moharrek, and G. N. Feliner (2017). A three-genome five-gene comprehensive phylogeny of the bulbous genus *Narcissus* (Amaryllidaceae) challenges current classifications and reveals multiple hybridization events. *Taxon* 66: 832–854.

Mennes, C. B., V. K. Y. Lam, P. J. Rudall, S. P. Lyon, S. W. Graham, E. F. Smets, and V. S. F. T. Merckx (2015). Ancient Gondwana break-up explains the distribution of the mycoheterotrophic family Corsiaceae (Liliales). *Journal of Biogeography* 42: 1123–1136.

Merckx, V. S. F. T., F. T. Bakker, S. Huysmans, and E. F. Smets (2009). Bias and conflict in phylogenetic inference of myco-heterotrophic plants: A case study in Thismiaceae. *Cladistics* 25: 64–77.

Merckx, V. S. F. T., S. Huysmans, and E. F. Smets (2010). Cretaceous origins of mycoheterotrophic lineages in Dioscoreales. *In* O. Seberg, G. Petersen, A. S. Barfod, and J. I. Davis [eds.], Diversity, phylogeny and evolution in the monocotyledons, 39–53. Aarhus University Press, Aarhus, Denmark.

Merckx, V. [S. F. T.], P. Schols, H. Maas-van der Kamer, P. Maas, S. Huysmans, and E. Smets (2006). Phylogeny and evolution of Burmanniaceae (Dioscoreales) based on nuclear and mitochondrial data. *American Journal of Botany* 93: 1684–1698.

Merckx, V. S. F. T., and E. F. Smets (2014). *Thismia americana*, the 101st anniversary of a botanical mystery. *International Journal of Plant Sciences* 175: S165–S175.

Miller, M. A., W. Pfeiffer, and T. Schwartz (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *In* Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA, 1–8.

Morris, J. L., M. N. Puttick, J. W. Clark, D. Edwards, P. Kenrick, S. Pressel, C. H. Wellman, et al. (2018). The timescale of early plant evolution. *Proceedings of the National Academy of Sciences USA* 115: E2274–E2283.

Nauheimer, L., D. Metzler, and S. S. Renner (2012). Global history of the ancient monocot family Araceae inferred with models accounting for past continental positions and previous ranges based on fossils. *New Phytologist* 195: 938–950.

Nickrent, D. L., A. Blarer, Y. L. Qiu, D. E. Soltis, P. S. Soltis, and M. Zanis (2002). Molecular data place Hydnoraceae with Aristolochiaceae. *American Journal of Botany* 89: 1809–1817.

Nickrent, D. L., A. Blarer, Y. L. Qiu, R. Vidal-Russell, and F. E. Anderson (2004). Phylogenetic inference in Rafflesiales: The influence of rate heterogeneity and horizontal gene transfer. *BMC Evolutionary Biology* 4: 40.

Nickrent, D. L., R. J. Duff, A. E. Colwell, A. D. Wolfe, N. D. Young, K. E. Steiner, and C. W. dePamphilis (1998). Molecular phylogenetic and evolutionary studies of parasitic plants. *In* D. E. Soltis, P. S. Soltis, and J. J. Doyle [eds.], Molecular systematics of plants II: DNA sequencing, 211–241. Kluwer Academic Press, Boston, Massachusetts, USA.

Nickrent, D. L., and E. M. Starr (1994). High rates of nucleotide substitution in nuclear small subunit (18S) rDNA from holoparasitic flowering plants. *Journal of Molecular Evolution* 39: 62–70.

Noy-Meir, I., and T. Oron (2001). Effects of grazing on geophytes in Mediterranean vegetation. *Journal of Vegetation Science* 12: 749–760.

Paradis, E., J. Claude, and K. Strimmer (2004). APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20: 289–290.

Plummer, M., N. Best, K. Cowles, and K. Vines (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News* 6: 7–11.

Prasad, V., C. A. E. Strömberg, A. D. Leaché, B. Samant, R. Patnaik, L. Tang, D. M. Mohabey, et al. (2011). Late Cretaceous origin of the rice tribe provides evidence for early diversification in Poaceae. *Nature Communications* 2: 480.

Price, J. P., and W. L. Wagner (2004). Speciation in Hawaiian angiosperm lineages: Cause, consequence, and mode. *Evolution* 58: 2185–2200.

Rabosky, D. L. (2014). Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS ONE* 9: e89543.

Rabosky, D. L., M. C. Grundler, C. J. Anderson, P. O. Title, J. J. Shi, J. W. Brown, H. Huang, and J. G. Larson (2014). BAMMtools: An R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution* 5: 701–707.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ross, T. G., C. F. Barrett, M. S. Gomez, V. K. Y. Lam, C. L. Henriquez, D. H. Les, J. I. Davis, et al. (2016). Plastid phylogenomics and molecular evolution of Alismatales. *Cladistics* 32: 160–178.

Saarela, J. M. (2006). Molecular systematic studies in commelinid monocots. Ph.D. dissertation, Department of Botany, University of British Columbia.

Saarela, J. M., S. V. Burke, W. P. Wysocki, M. D. Barrett, L. G. Clark, J. M. Craine, P. M. Peterson, et al. (2018). A 250 plastome phylogeny of the grass family (Poaceae): Topological support for different data partitions. *PeerJ* 6: e4299.

Saarela, J. M., H. S. Rai, J. A. Doyle, P. K. Endress, S. Mathews, A. D. Marchant, B. G. Briggs, and S. W. Graham (2007). Hydatellaceae identified as a new branch near the base of the angiosperms. *Nature* 446: 312–315.

Saarela, J. M., W. P. Wysocki, C. F. Barrett, R. J. Soreng, J. I. Davis, L. G. Clark, S. A. Kelchner, et al. (2015). Plastid phylogenomics of the cool-season grass subfamily: Clarification of relationships among early-diverging tribes. *AoB Plants* 7: plv046.

Sass, C., W. J. D. Iles, C. F. Barrett, S. Y. Smith, and C. D. Specht (2016). Revisiting the Zingiberales: Using multiplexed exon capture to resolve ancient and recent phylogenetic splits in a charismatic plant lineage. *PeerJ* 4: e1584.

Sculthorpe, C. D. (1967). The biology of aquatic vascular plants. Edward Arnold, London, UK.

Sessa, E. B., E. A. Zimmer, and T. J. Givnish (2012). Unraveling reticulate evolution in North American *Dryopteris* (Dryopteridaceae). *BMC Biology* 12: 104.

Sheikh, S. I., T. Kahveci, S. Ranka, and J. G. Burleigh (2013). Stability analysis of phylogenetic trees. *Bioinformatics* 29: 166–174.

Sheviak, C. J., and M. L. Bowles (1986). The prairie fringed orchids—a pollinator-isolated species pair. *Rhodora* 88: 267–290.

Smith, J. F. (2001). High species diversity in fleshy-fruited tropical understory plants. *American Naturalist* 157: 646–653.

Smith, S. A., and J. W. Brown (2018). Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany* 105: 1–13.

Sokoloff, D. D., M. V. Remizowa, B. G. Briggs, and P. J. Rudallm (2009). Shoot architecture and branching pattern in perennial Hydatellaceae (Nymphaeales). *International Journal of Plant Science* 170: 869–884.

Sokoloff, D. D., M. V. Remizowa, T. D. MacFarlane, R. E. Tuckett, M. M. Ramsay, A. S. Beer, S. R. Yadav, and P. J. Rudall (2008). Seedling diversity in Hydatellaceae: Implications for the evolution of angiosperm cotyledons. *Annals of Botany* 101: 153–164.

Soltis, D. E., S. A. Smith, N. Cellinese, K. J. Wurdack, D. C. Tank, S. F. Brockington, N. F. Refulio-Rodriguez, et al. (2011). Angiosperm phylogeny: 17 genes, 640 taxa. *American Journal of Botany* 98: 704–730.

Soltis, D. E., P. S. Soltis, M. W. Chase, M. E. Mort, D. C. Albach, M. Zanis, V. Savolainen, et al. (2000). Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences. *Botanical Journal of the Linnean Society* 133: 381–461.

Spalink, D., B. T. Drew, M. C. Pace, J. G. Zaborsky, J. R. Starr, K. M. Cameron, T. J. Givnish, et al. (2016). Biogeography of the cosmopolitan sedges (Cyperaceae) and the area-richness correlation in plants. *Journal of Biogeography* 43: 1893–1904.

Spriggs, E. L., P. A. Christin, and E. J. Edwards (2014). C4 photosynthesis promoted species diversification during the Miocene grassland expansion. *PLoS ONE* 9: e97722.

Spriggs, E. L., W. L. Clement, P. W. Sweeney, S. Madriñán, E. J. Edwards, and M. J. Donoghue (2015). Temperate radiations and dying embers of a tropical past: The diversification of *Viburnum*. *New Phytologist* 207: 340–354.

Springer, M. S., and J. Gatesy (2018). On the importance of homology in the age of phylogenomics. *Systematics and Biodiversity* 16: 210–228.

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.

Stebbins, G. L. (1974). Flowering plants: Evolution above the species level. Harvard University Press, Cambridge, Massachusetts, USA.

Steele, R., K. L. Hertweck, D. Mayfield, M. R. McKain, J. Leebens-Mack, and J. C. Pires (2012). Quality and quantity of data recovered rom massively parallel sequencing: Examples in Asparagales and Poaceae. *American Journal of Botany* 99: 330–348.

Swofford, D. L., and C. D. Bell (2017). PAUP* manual. Available at http://phylosolutions.com/paup-documentation/paupmanual.pdf.

Sytsma, K. J., D. Spalink, and B. Berger (2014). Calibrated chronograms, fossils, outgroup relationships, and root priors: Re-examining the historical biogeography of Geraniales. *Biological Journal of the Linnean Society* 113: 29–49.

Takhtajan, A. L. (1969). Flowering plants: Origin and dispersal. Oliver and Boyd, Edinburgh, Scotland.

Takhtajan, A. L. (1991). Evolutionary trends in flowering plants. Columbia University Press, New York, New York, USA.

Tamura, M. N. (1998). Nartheciaceae. *In* K. Kubitzki [ed.], The families and genera of vascular plants, vol. 3, 381–392. Springer, Berlin, Germany.

Tang, C. Q., C. D. L. Orme, L. Bunnefeld, F. A. Jones, S. Powell, M. W. Chase, T. G. Barraclough, and V. Savolainen (2017). Global monocot diversification: Geography explains variation in species richness better than environment or biology. *Botanical Journal of the Linnean Society* 183: 1–15.

Teisher, J. K., M. R. McKain, B. A. Schaal, and E. A. Kellogg (2017). Polyphyly of Arundinoideae (Poaceae) and evolution of the twisted geniculate lemma awn. *Annals of Botany* 120: 725–738.

Theim, T. J., R. Y. Shirk, and T. J. Givnish (2014). Spatial genetic structure in four understory *Psychotria* species and implications for tropical forest diversity. *American Journal of Botany* 101: 1189–1199.

Townsend, J. (2007). Profiling phylogenetic informativeness. *Systematic Biology* 56: 222.

Townsend, J., and C. Leuenberger (2011). Taxon sampling and optimal rates of evolution for phylogenetic inference. *Systematic Biology* 60: 358–365.

Tremblay, R. L., J. D. Ackerman, J. K. Zimmerman, and R. N. Calvo (2005). Variation in sexual reproduction in orchids and its evolutionary consequences: A spasmodic journey to diversification. *Biological Journal of the Linnean Society* 84: 1–54.

Van der Niet, T., R. Peakall, and S. D. Johnson (2014). Pollinator-driven ecological speciation in plants: New evidence and future perspectives. *Annals of Botany* 113: 199–211.

Vargas, O. M., E. M. Ortiz, and B. B. Simpson (2017). Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae: *Diplostephium*). *New Phytologist* 214: 1736–1750.

Vieira, L. N., K. G. dos Anos, H. Faoro, H. P. de Freitas Fraga, T. M. Greco, F. de Oliveira Pedrosa, E. M. de Souza, et al. (2016). Phylogenetic inference and SSR characterization of tropical bamboos tribe Bambuseae (Poaceae: Bambusoideae) based on complete plastid genome sequences. Current Genetics 62: 443–453.

Wickett, N. J., S. Mirarab, N. Nguyen, T. Warnow, E. Carpenter, N. Matasci, S. Ayyampalayam, et al. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences USA* 111: 4859–4868.

Willyard, A., R. Cronn, and A. Liston (2009). Reticulate evolution and incomplete lineage sorting among the ponderosa pines. *Molecular Phylogenetics and Evolution* 52: 498–511.

Wu, Z. Q., and S. Ge (2012). The phylogeny of the BEP clade in grasses revisited: Evidence from the whole-genome sequences of chloroplasts. *Molecular Phylogenetics and Evolution* 62: 573–578.

Wysocki, W. P., S. V. Burke, W. D. Swingley, and M. R. Duvall (2016). The first complete plastid genome from Joinvilleaceae (*J. ascendens*; Poales) shows unique and unpredicted rearrangements. *PLoS ONE* 11: e0163218.

Wysocki, W. P., L. G. Clark, L. Attigala, E. Ruiz-Sanchez, and M. R. Duvall (2015). Evolution of the bamboos (Bambusoideae; Poaceae): A full plastome phylogenomic analysis. *BMC Evolutionary Biology* 15: 50.

Zimmer, E. A., and J. Wen (2015). Using nuclear gene data for plant phylogenetics: Progress and prospects II. Next-gen approaches. *Journal of Systematics and Evolution* 53: S371–S379.

FIGURE 1. Summary of relationships among monocot families and selected subfamilies based on (A) maximum likelihood (ML) analysis of 77 plastid genes, (B) ML analysis of sequences of four plastid genes from the benchmark study of Chase et al. (2006), and (C) MP analysis of benchmark data (one of 12 shortest trees; arrowheads indicate branches that collapse in the strict consensus). Bootstrap support values for internal branches are color-coded as shown in the legend; note increased support in moving from C to B to A. Orders are indicated by magenta brackets, except for Dasypogonales, whose position shifts among analyses; commelinids are highlighted by the large gray boxes. Subfamilies of Agavaceae, Amaryllidaceae, and Asphodelaceae recognized as separate families in previous versions of APG are joined by black brackets, as are the subfamilies of Orchidaceae. New positions of families or subfamilies in the ML plastome analysis vs. the MP benchmark study are highlighted by blue dots; families added in the plastome analysis are highlighted by red dots. Hollow dots at the end of branches indicate the presence of one or more mycoheterotrophic species. Gray cloud indicates paraphyly. Number of taxa per family or subfamily in the present plastome study is indicated after the taxon names in A.

FIGURE 2. Maximum likelihood phylogeny of Zingiberales based on complete aligned plastomes. Bootstrap support values for individual branches are 100% unless otherwise noted.

FIGURE 3. Monocot chronogram/diversigram. Ages of divergence of taxa at the subfamily, family, and ordinal levels of monocots and angiosperm outgroups are shown by branch depth. Significant accelerations of diversification are identified by red dots; estimated rates of net species diversification (sp sp^{-1} My^{-1}) from BAMM are color-coded as indicated. Area of bubbles is proportional to the number of species in terminal taxa. The Cretaceous-Tertiary boundary is

indicated by the dashed line. See Appendix S13 for ages and 95% confidence intervals for all nodes within and among families.

FIGURE 4. (A) Histogram of stem ages for APG IV families plus Taccaceae and Thismiaceae. Note the bimodal distribution of stem ages, with a trough just prior to the Cretaceous-Tertiary boundary. (B) Family ages and 95% confidence intervals, plotted from the youngest to the oldest families.

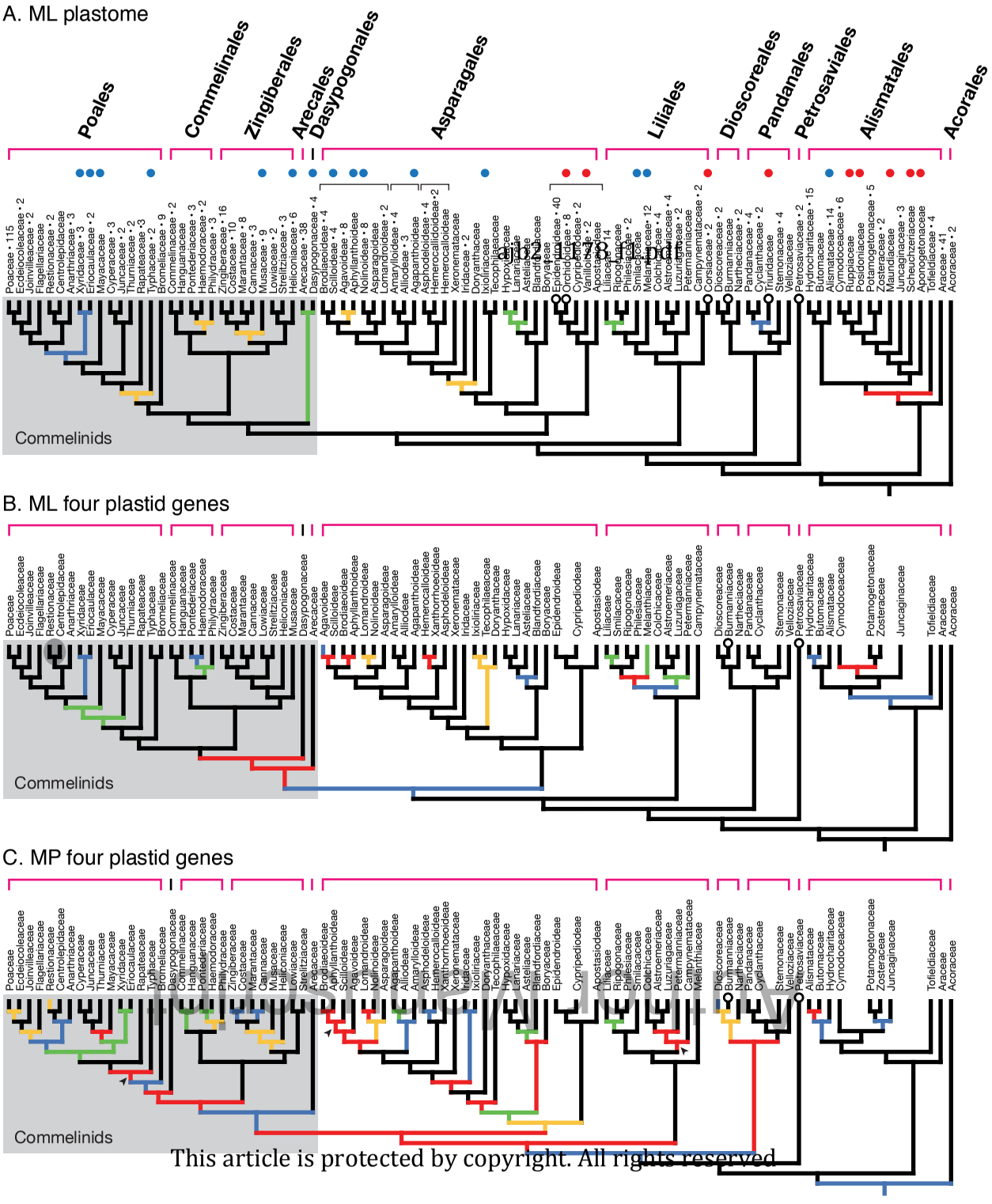
FIGURE 5. Distribution of species richness across families, showing a strong approach to a log-series distribution. The line and equation represent the regression of species number per family against family rank; each family has, on average, 10.3% fewer species than the next larger family on its left. Note that three families—Orchidaceae, Poaceae, and Cyperaceae—have substantially more species than expected from the log-series model.

FIGURE 6. Branch ascertainment (top) and average branch support (bottom) across monocots as predicted from the estimated logistic regression model with branch length, branch depth, and number of genes as predictors (all three log-transformed), and with their two-way interaction effects. (Left) Predictions as a function of branch length and gene number for a fixed branch depth of 0.05. (Right) Predictions as a function of branch depth and branch number for a fixed branch length of 0.001.

FIGURE 7. Branch ascertainment (left) and average branch support (right) across monocots, as a function of branch length (horizontal axis) and relative branch depth (vertical axis), for increasing numbers of genes (top to bottom). Each point represents one branch, with color to indicate support (from red for low support to blue for high support). Curves represent the combination of branch lengths and depths at which ascertainment or support is predicted to be 50% (yellow), 70% (light green), or 90% (blue-green), with predictions based on the logistic regression model with two-way interactions.

TABLE 1. Estimated main effects of branch length, branch depth, and number of genes (log-transformed) on branch ascertainment and bootstrap support.

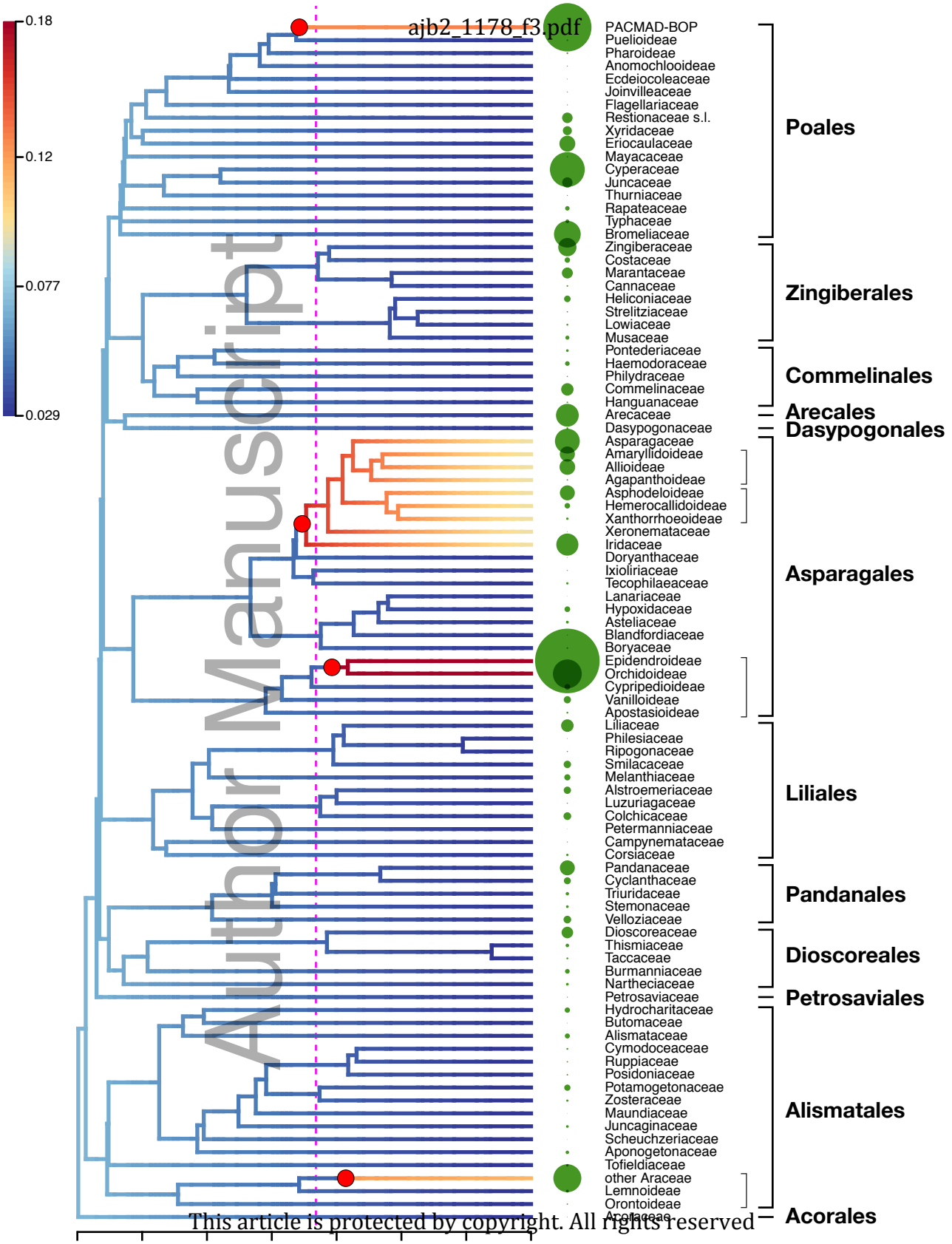
Clade	Response		Estimate	SE	*t*	*P*
Monocots	Branch ascertainment	Intercept (*a*)	6.19	0.184	33.6	$<10^{-16}$
		Branch length (*b*)	3.03	0.081	37.4	$<10^{-16}$
		Branch depth (*c*)	−1.65	0.079	−21.0	$<10^{-16}$
		Number of genes (*d*)	2.30	0.084	27.4	$<10^{-16}$
	Bootstrap support	Intercept (*a*)	5.42	0.182	29.8	$<10^{-16}$
		Branch length (*b*)	2.98	0.084	35.5	$<10^{-16}$
		Branch depth (*c*)	−1.67	0.082	−20.5	$<10^{-16}$
		Number of genes (*d*)	2.42	0.079	30.8	$<10^{-16}$
Zingiberales	Branch ascertainment	Intercept (*a*)	8.12	0.285	28.5	$<10^{-16}$
		Branch length (*b*)	3.89	0.121	32.2	$<10^{-16}$
		Branch depth (*c*)	−2.06	0.095	−21.7	$<10^{-16}$
		Number of genes (*d*)	2.44	0.087	28.0	$<10^{-16}$
		Spacer inclusion (*e*)	0.59	0.079	7.52	1.9×10^{-13}
	Bootstrap support	Intercept (*a*)	7.58	0.231	32.9	$<10^{-16}$
		Branch length (*b*)	3.84	0.101	38.0	$<10^{-16}$
		Branch depth (*c*)	−2.07	0.080	−25.7	$<10^{-16}$
		Number of genes (*d*)	2.36	0.068	34.9	$<10^{-16}$

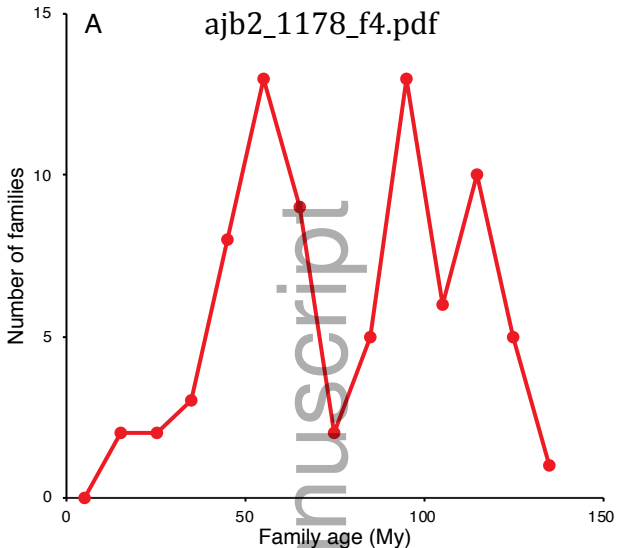


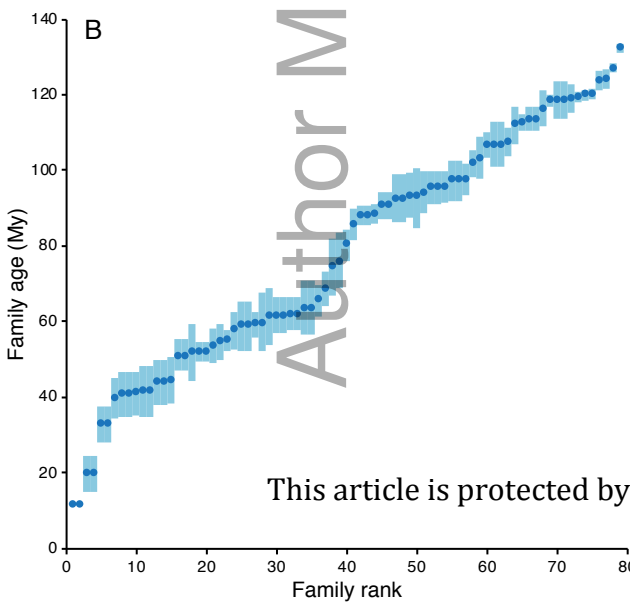
Zingiberales

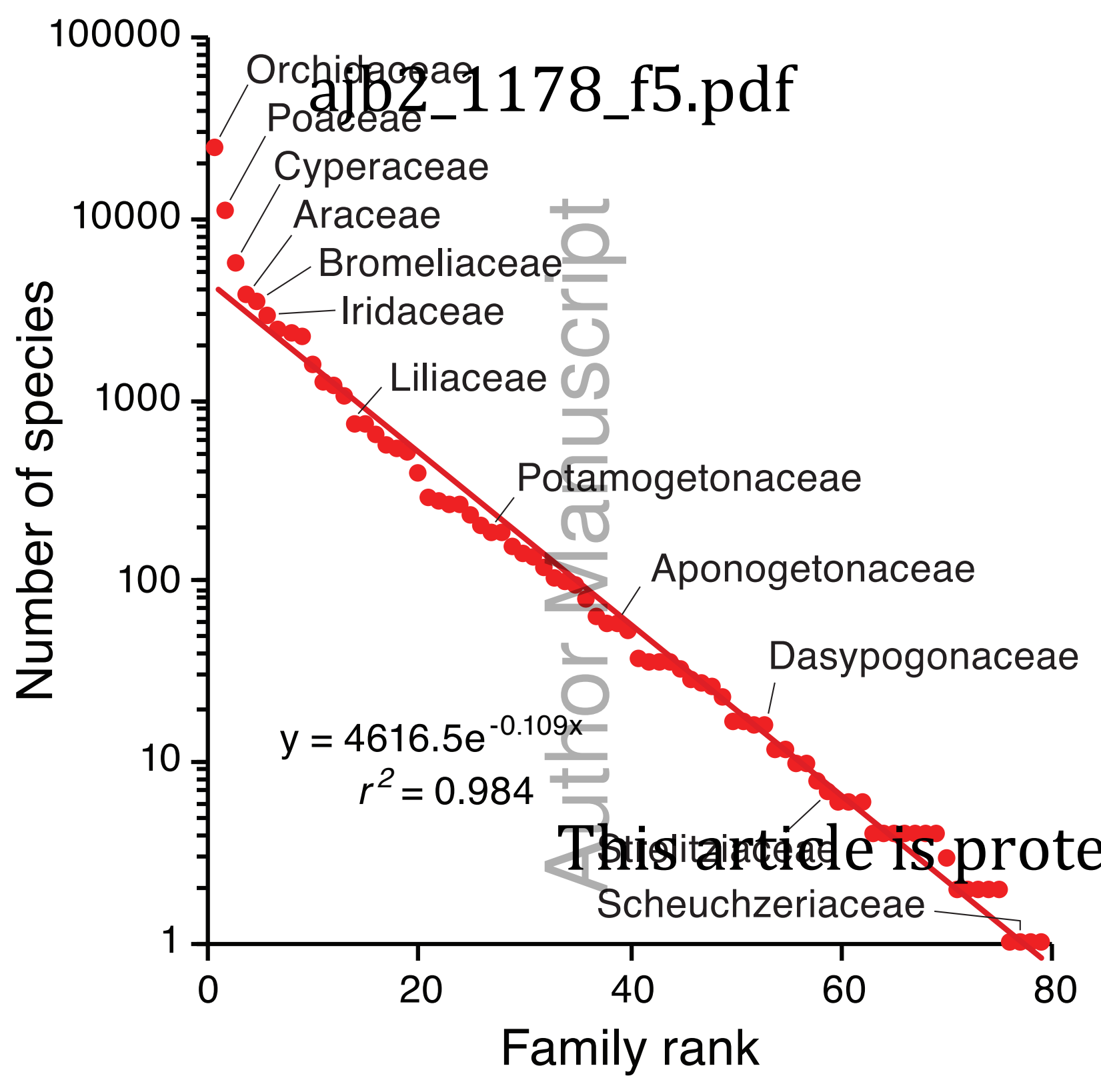
Etlingera elatior
Alpinia purpurata
Alpinia zerumbet
Aframomum angustifolium
71.8
Renealmia alpinia
Elettariopsis stenosiphon
Riedelia sp
Siamanthus siliquosus
Zingiberaceae
99.6
Zingiber officinale
Zingiber spectabile
Scaphochlamys sp
Hedychium coronarium
Curcuma longa
Curcuma roscoeana
Globba winitii
Siphonochilus kirkii
Costus pictus
Costus pulverulentus
Costus osae
Costus dubius
Costus gabonensis
Costaceae
Cheilocostus speciosus
Tapeinochilos ananassae
Dimerocostus strobilaceus
Monocostus uniflorus
Chamaecostus acaulis
96
Ischnosiphon heleniae
Calathea roseopicta
Donax canniformis
Marantaceae
Stromanthe stromanthoides
Halopegia azurea
Marantochloa leucantha
Thaumatococcus daniellii
Canna indica
Canna jaegeriana
Cannaceae
Canna iridiflora
Strelitzia reginae
Strelitziaceae
Strelitzia caudata
Ravenala madagascariensis
Orchidantha fimbriata
Lowiaceae
94.2
Orchidantha maxillarioides
Heliconia collinsiana
Heliconia nutans
Heliconiaceae
Heliconia sp. HNT 10823
Heliconia acuminata
83.8
Musa ornata
Musa basjoo
Musa sp. FTBG 2007-0825A
Musaceae
Musa coccinea
Ensete ventricosum
Ensete superbum
Musella lasiocarpa
Xiphidium caeruleum
Anigozanthos flavidus
Hanguana malayana
Typha latifolia

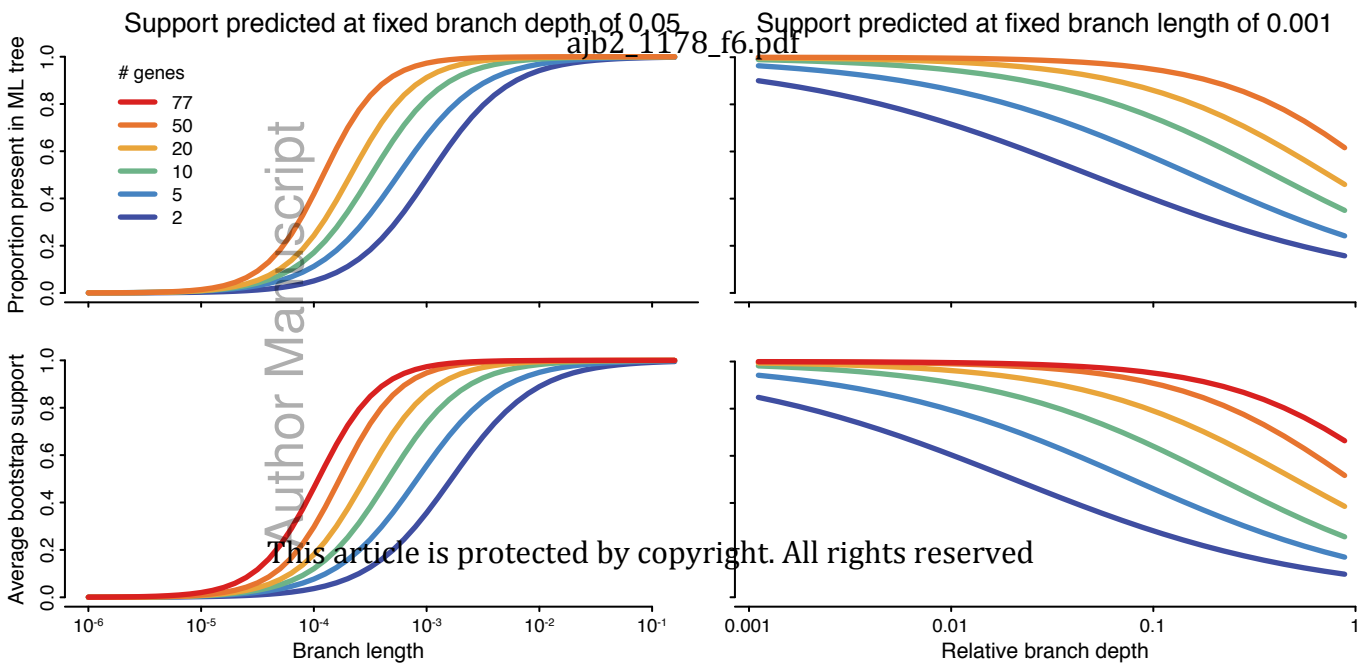
0.05










Support predicted at fixed branch depth of 0.05
Support predicted at fixed branch length of 0.001
ajb2_1178_f6.pdf
genes
77
50
20
10
5
2
Proportion present in ML tree
Average bootstrap support
0.0
0.2
0.4
0.6
0.8
1.0
Author Manuscript
This article is protected by copyright. All rights reserved
10-6
10-5
10-4
10-3
10-2
10-1
Branch length
0.001
0.01
0.1
1
Relative branch depth