



Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Short Communication

How well do we understand the overall backbone of cycad phylogeny? New insights from a large, multigene plastid data set

Jessie M. Zgurski^a, Hardeep S. Rai^b, Quentin M. Fai^a, David J. Bogler^c, Javier Francisco-Ortega^{d,e}, Sean W. Graham^{b,*}^a Department of Biological Sciences, CW405 Biological Sciences Centre, University of Alberta, Edmonton, AB, Canada T6G 2E9^b UBC Botanical Garden & Centre for Plant Research, Faculty of Land and Food Systems, & Department of Botany, 2357 Main Mall, University of British Columbia, Vancouver, BC, Canada V6T 1Z4^c Missouri Botanical Garden, P.O. Box 299, St. Louis, MO 63166-0299, USA^d Department of Biological Sciences, Florida International University, Miami, FL 33199, USA^e Center for Tropical Plant Conservation, Fairchild Tropical Botanic Garden, Coral Gables, FL 33156, USA

ARTICLE INFO

Article history:

Received 31 August 2007

Revised 24 January 2008

Accepted 3 March 2008

Available online xxx

1. Introduction

The cycads have a fossil record that extends back at least 250 Myr to the Permian age (Gao and Thomas, 1989), making them the oldest crown-clade of seed plants (Hermsen et al., 2006). They are the most species-rich group of gymnosperms after the conifers, with ~300 species in 10–12 extant genera (Hill et al., 2003). Despite their remarkable persistence and diversity, they are also among the most threatened of all plant groups (e.g., Donaldson et al., 2003). These features make them a prime candidate for comparative study by plant biologists (e.g., Brenner et al., 2003). However, their evolutionary history is still relatively poorly characterized, as their relationship to the other major seed-plant groups, both extant and extinct, continues to elude satisfactory resolution (e.g., Burleigh and Mathews, 2004, 2007a,b; Doyle, 2006; Hilton and Bateman, 2006), and our understanding of their internal phylogenetic relationships also lacks clarity. The latter question is the subject of this study.

Trees inferred in recent studies of higher-order cycad phylogeny (Hill et al., 2003; Chaw et al., 2005; Hermsen et al., 2006) have major points of disagreement concerning higher-order cycad relationships. We inferred a ‘greatest agreement subtree’ (see Swofford, 1991) using PAUP* (version 4.0b10; Swofford, 2002) from these studies to summarize common aspects of the main cycad backbone. This approach prunes off taxa that tend to float along the backbone in different studies; the resulting subtree is the largest consistent set of branching relationships given the least

amount of pruning. For this comparison we focused on trees from two recent molecular studies of cycad phylogeny: (i) the combined plastid (‘chloroplast’), combined nuclear and combined molecular analyses in Hill et al. (2003, see their Figs. 5, 7 and 9, respectively); (ii) separate analyses of the first two codons of the relatively rapidly evolving plastid *matK* locus, and the nuclear 5.8S/ITS2 region in Chaw et al. (2005, see their Figs. 7–8, respectively; note that we interpolated the positions of *Ginkgo* and *Cycas*, as in their Fig. 9). We also considered the morphological study of Hermsen et al. (2006), pruning extinct taxa from their single tree (their Fig. 18). We reduced these various trees to genus-level phylogenies, a reasonable simplification for studying higher-order cycad phylogeny, as the genera are monophyletic if *Chigua* is assumed to be part of *Zamia*, and *Dyerocycas* part of *Cycas* (see Hill et al., 2003; Chaw et al., 2005); the genus *Epicycas* mentioned in Hill et al. (2003) is a synonym of *Cycas* (Chen et al., 2004). A single greatest agreement subtree is inferred from these genus-level topologies. We present this topology and the various attachment points of pruned taxa (*Bowenia*, *Dioon* and *Stangeria*) in Fig. 1a. Trees inferred from recent analysis of nuclear ITS data and plastid data from the *trnL* intron by Bogler and Francisco-Ortega (2004, see their Figs. 2–3; not depicted here) are also consistent with this subtree, although their plastid tree is partly unresolved.

The low to moderate support typical of much of the backbone of cycad phylogeny in recent studies may be a function of the extremely slow rate of DNA substitution in cycads compared to most other seed plants (Rai et al., 2003; see also Hill et al., 2003), which yields a relatively low amount of DNA sequence variation for phylogenetic analysis per nucleotide examined. Limited variation can be addressed by increasing the amount of data

* Corresponding author. Fax: +1 604 822 6089.

E-mail address: swgraham@interchange.ubc.ca (S.W. Graham).

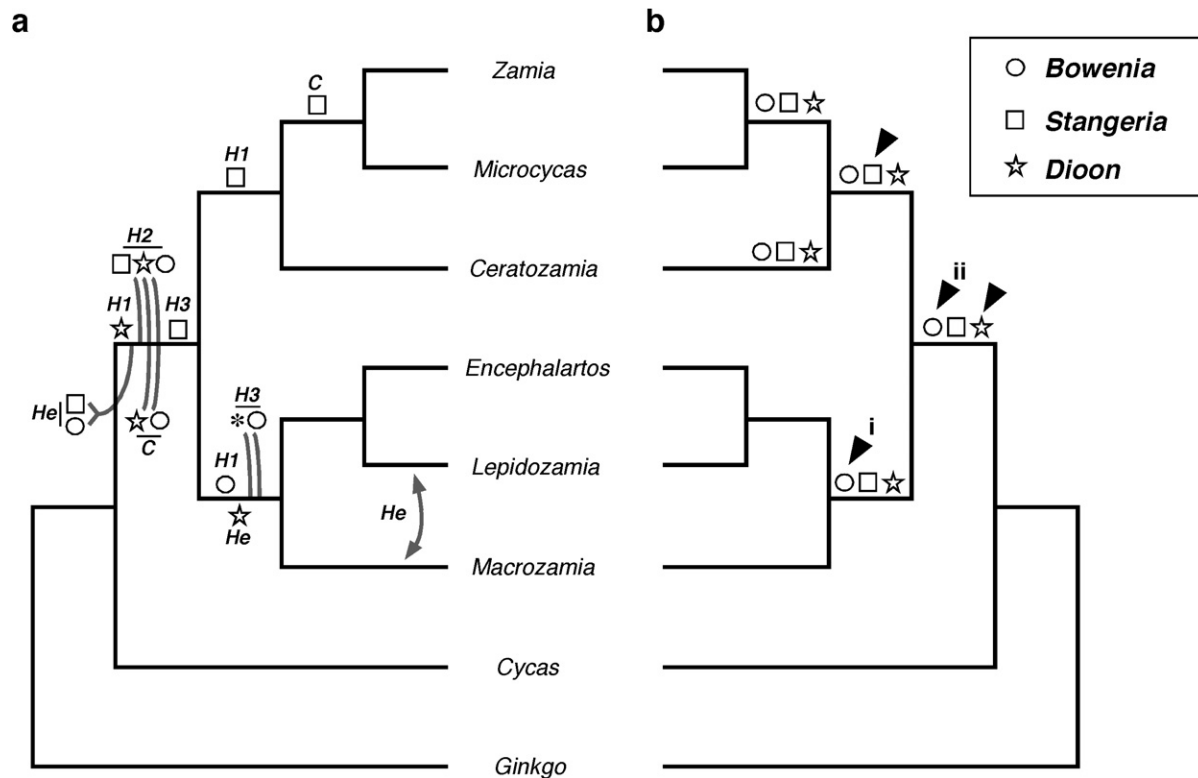


Fig. 1. (a) The single greatest agreement subtree for genus-level topologies from recent molecular studies (Hill et al., 2003; Chaw et al., 2005); outgroup = *Ginkgo biloba*. Symbols corresponding to *Bowenia*, *Dioon* and *Stangeria* (see caption) indicate the points at which they were pruned. The 'H1', 'H2', 'H3' labels refer (respectively) to analyses of combined plastid data, combined nuclear data and all data combined in Hill et al.; the 'C' labels refer to (identical) plastid and nuclear backbones in Chaw et al. (2005); see text. Apart from a contrasting arrangement of *Encephalartos*, *Lepidozamia* and *Macrozamia* (double-headed arrow) the agreement subtree is congruent with the morphological study of Hermsen et al. (2006; 'He' labels indicate pruning points). For subtree branches in which multiple curved lines emerge from the same subtree branch, the relative branching order corresponds to that seen in the original study. (b) Summary of Shimodaira–Hasegawa tests of divergent placements for each of the three pruned genera on the greatest agreement topology, for a multigene plastid data set consisting of 17 loci and associated noncoding regions. Branches without labels are rejected as possible positions ($P < 0.01$), labeled branches are not. Single arrowheads indicate the optimal ML placements of pruned genera on the subtree ("i" and "ii" are alternative optimal positions for *Bowenia* according to the GTR + Γ and GTR + Γ + I DNA substitution models, respectively, see text).

collected per taxon. This was the approach we took in a previous study of the backbone of cycad phylogeny, in which we surveyed a large subset of the plastid genome using an exemplar-based taxon sampling (Rai et al., 2003). However, that study lacked several key taxa along the backbone, and so we expand our original sample here by adding representatives of the previously unsampled genera (*Lepidozamia*, *Macrozamia* and *Microcycas*). We use this more extensive sampling to characterize the robustness of phylogenetic inference of higher-order cycad relationships. Our main goals are to investigate whether our large plastid data set permits well supported inference of an overall backbone of cycad phylogeny, whether this is consistent with the pruned backbone that represents the common elements of other recent studies (Fig. 1a), and the degree to which our data permit sturdy localization of the position of the three genera pruned from this agreement subtree.

2. Materials and methods

2.1. Recovery of plastid sequences, DNA sequence alignment

We used exemplar-based taxon sampling to represent the major branches of cycad phylogeny. Methods of DNA extraction, amplification and sequencing follow Graham and Olmstead (2000) and Rai et al. (2003). The additional taxa included here (Table 1) are *Lepidozamia hopei* (one of two species in the genus), *Macrozamia moorei* (one of 38 species in the genus) and the

monotypic *Microcycas* (*M. calocoma*). We surveyed a total of 17 genes per taxon for each exemplar. These genes code for proteins that function in photosynthesis (*atpB*, *rbcl* and ten photosystem II, *psb*, genes), the plastid translation apparatus (ribosomal protein genes *rpl2*, *rps7* and *3'-rps12*), and chlororespiration (*ndhB* and *ndhF*, coding for two of the subunits of plastid NADH dehydrogenase), plus their associated noncoding regions (three introns, one each in *rpl2*, *3'-rps12* and *ndhB*, and eight intergenic spacer regions that span genes in two photosystem II gene clusters [*psbE-psbJ*] and [*psbB-psbH*] and a cluster that includes *3'-rps12*, *rps7* and *ndhB*).

We added these new sequences to a previous alignment (Graham et al., 2006), using Se-AL version 1.0 (Rambaut, 1998) to manually adjust the alignment following criteria laid out in Graham et al. (2000). The alignment comprises 25,687 nucleotides, derived from an average of ~13.6 kb of unaligned DNA sequence data per cycad exemplar (range = 13.3–14.0 kb; *Ginkgo biloba* is 14.3 kb). For comparison, the entire plastid genome of *Cycas taitungensis* is ~163.4 kb (Wu et al., 2007). We offset several difficult-to-align regions in the intergenic spacer regions, following Graham et al. (2006). The resulting staggered regions are typically limited to single taxa, which are ignored in parsimony-based analysis (although subsets of the offset regions include aligned blocks involving multiple taxa). These offset regions should have only minimal effect on model-based methods (e.g., on estimation of base frequency parameter values). Of the total alignment, 471 sites are potentially parsimony informative, and 1395 are variable but parsimony uninformative.

Table 1

GenBank accession details for new cycad sequences examined here (see Rai et al., 2003 for other taxa)

| Taxon [collection no. (location)] | Gene or region | | | | | | | |
|---|----------------|-------------|-----------------------|-----------------|-----------------------|-----------------------|-------------|---|
| | <i>atpB</i> | <i>ndhF</i> | <i>psbB</i> , T, N, H | <i>psbD</i> & C | <i>psbE</i> , F, L, J | <i>rbcL</i> | <i>rpl2</i> | 3'- <i>rps12</i> , <i>rps7</i> <i>ndhB</i> , <i>trnL</i> |
| <i>Lepidozamia hopei</i> Regel [69428M (FTG)] | AY699124 | AY699157 | AY699145 | AY699172 | AY699148 | AY699175 | AY699151 | AY699154 |
| <i>Macrozamia moorei</i> F. Muell. [59302 (FTG)] | AY699125 | AY699158 | AY699146 | AY699173 | AY699149 | AY699176 | AY699152 | AY699155 |
| <i>Microcycas calocoma</i> (Miq.) A. DC. [77404T (FTG)] | AY699126 | AY699159 | AY699147 | AY699174 | AY699150 | AF531214 ^a | AY699153 | AY699156 |

^a Previously published (see GenBank accession for voucher information).

2.2. Phylogenetic analyses

We performed heuristic maximum parsimony (MP) and maximum likelihood (ML) searches using PAUP* (Swofford, 2002). We used *Ginkgo biloba* to root cycad phylogeny (see Rai et al., 2003; Chaw et al., 2005). For the MP analysis (using PAUP*), we treated all characters and character-state changes as equally weighted, and used TBR (tree-bisection-reconnection) branch swapping with 100 random addition replicates. PAUP* defaults were used for all other settings. We performed exploratory MP investigations on subpartitions of our plastid data for the following six cases: (1) All of the photosystem II genes combined (coding regions only); (2) All inverted repeat data combined (coding and noncoding regions); (3) *atpB*, *rbcL* and *ndhF* combined; (4) Codon positions 1 and 2 from all genes combined; (5) Codon position 3 combined; (6) All noncoding data combined. Cases 1–3 are independent from each other in terms of the data examined in each data partition, as are cases 4–6 (a 53 bp overlap in *psbC* and *psbD* was excluded from cases 4 and 5, as each nucleotide in this region belongs to two different codon position classes). A greatest agreement subtree of all of the MP trees inferred from the first five cases (data not shown) has an identical topology to the agreement subtree shown in Fig. 1a. The sixth partition was also congruent with this agreement subtree, except that it had a weak disagreement in a single element (a *Lepidozamia*–*Macrozamia* clade, with 66% bootstrap support). Setting aside this minor conflict, there was evidently a high level of congruence among the data partitions we assessed, and we therefore focus our analyses on the full plastid data set.

For ML searches, we chose the model of DNA sequence evolution with the hierarchical likelihood ratio test (hLRT) and the Akaike Information Criterion (AIC), using ModelTest (version 3.7; Posada and Crandall, 1998). When all 10 cycads and *Ginkgo* are considered, both assessment methods recover the same optimal DNA substitution model, GTR + Γ + I [i.e., the general-time-reversible (GTR) model, with among-site rate variation accounted for by considering the proportion of invariable sites (I), and the gamma (Γ) distribution, using four substitution-rate categories for the gamma shape parameter alpha (α)]. All ML analyses considered the full plastid data set. For the ML search, we fixed DNA substitution model parameters to values obtained for the optimal hLRT model in ModelTest, and used the heuristic search strategy used for the MP analyses, but with only a single random addition replicate. For MP and ML analyses we assessed branch support using the nonparametric bootstrap (Felsenstein, 1985), with 500 bootstrap replicates (but using one random addition replicate per bootstrap replicate). We use 'weak', 'moderate', and 'strong' in reference to clades that have bootstrap support values <70%, 70–89% and \geq 90%, respectively.

We used the greatest agreement subtree summarizing major analyses of other recent cycad studies (Fig. 1a) as a starting point to estimate the propensity of three of the genera (*Bowenia*, *Dioon* and *Stangeria*) to float along the remaining backbone of cycad phylogeny. Three sets of trees were constructed in MacClade ver. 4.03

(Maddison and Maddison, 2001), one for each genus. The added genus was attached to all 13 possible positions (branches) in the eight-taxon greatest agreement subtree (Fig. 1a), including the stem lineage connecting all other cycads to *Ginkgo*. The two other floating genera were ignored for each set of tests. We used the test of Shimodaira and Hasegawa (1999; implemented in PAUP*; Swofford, 2002) to simultaneously compare the resulting 13 possible nine-taxon trees for each floating genus (see Goldman et al., 2000), to assess which suboptimal positions for this taxon are significantly worse than the optimal one. We determined the optimal DNA substitution model for each taxon set using ModelTest, and estimated model parameters in PAUP* as part of the Shimodaira–Hasegawa test. The optimal DNA substitution models indicated for the hLRT and AIC comparisons disagree for two of these nine-taxon analyses (the ones involving *Stangeria* and *Dioon*). In these cases, the former method chose the GTR + Γ model and the latter the GTR + Γ + I model, and so we repeated the nine-taxon analyses under each DNA substitution model.

We also used the Shimodaira–Hasegawa test to compare several additional hypotheses for the full 11-taxon data set (i.e., all 10 cycads plus *Ginkgo*), including the hypothesis that *Bowenia* and *Stangeria* comprise a clade that corresponds to the family Stangeriaceae (e.g., Hermsen et al., 2006). We performed topologically constrained ML searches in PAUP* corresponding to each hypothesis, using the model and model parameters determined for the unconstrained 11-taxon case, and then used the Shimodaira–Hasegawa test to simultaneously compare the scores of the shortest constrained trees to the optimal unconstrained 11-taxon ML tree.

3. Results

3.1. Phylogenetic results

The single ML tree and two MP trees inferred with the full plastid data set (Fig. 2) are completely congruent from the perspective of the branching order of the seven cycad genera retained in the greatest agreement subtree derived from other recent studies (cf. Figs. 1a and 2). Trees inferred from the MP and ML analyses here differ topologically from each other in the placement of *Bowenia* and *Stangeria*. Several of the internal (non-terminal) branches concerning these discordant relationships are extremely short (Fig. 2a) and are correspondingly poorly supported. However, *Dioon* is consistently placed here as the sister group of all cycads except *Cycas*, with moderate support in ML and MP analyses (Fig. 2). A deep basal split between *Cycas* and all other cycads consistently has 100% bootstrap support. A clade consisting of *Ceratozamia*, *Microcycas*, *Zamia* and *Stangeria* is recovered with moderate to strong support in MP and ML analyses (89–98% support; Fig. 2). Within this clade, *Microcycas* and *Zamia* are strongly supported as sister taxa, and two different arrangements of *Stangeria* and *Ceratozamia* are seen in the two MP trees, each with poor support; one of these arrangements, with *Stangeria* sister to the remaining three taxa (Fig. 2b), is also weakly supported in the ML analyses (53%; Fig. 2a). A clade

consisting of *Encephalartos*, *Lepidozamia* and *Macrozamia*, and a subclade consisting of *Encephalartos* and *Lepidozamia* are recovered with strong support (100% bootstrap support in MP and ML analyses). *Bowenia* is weakly supported either as the sister group of all cycads except *Cycas* and *Dioon* (MP trees; Fig. 2b and c), or as the sister group of the *Encephalartos*–*Lepidozamia*–*Macrozamia* clade (ML tree; Fig. 2a). The estimated branch support values from the MP and ML bootstrap analyses are broadly comparable for the well-supported branches (Fig. 2). One of our MP trees is consistent with the less inclusive taxon sampling for this gene sampling in Rai et al. (2003; cf. their Fig. 2 and Fig. 2c here).

3.2. Shimodaira–Hasegawa tests

When taxa pruned in the greatest agreement subtree topology (i.e., *Bowenia*, *Dioon* or *Stangeria*) are individually forced onto different possible branches of this tree, eight of their twelve possible suboptimal placements are rejected by the Shimodaira–Hasegawa test ($P < 0.01$ for each of the genera). However, four suboptimal branches cannot be rejected as possible placements in each case (Fig. 1b). The set of five plausible attachment points (both optimal and suboptimal) is the same for all three genera, although the optimal positions of the three floating genera in the nine-taxon trees (marked with arrowheads; Fig. 1b) differ from each other and mostly correspond to their positions in the 11-taxon ML tree (Fig. 2). *Bowenia* has two possible optimal placements in the nine-taxon case, that depend on the DNA substitution model used (labels 'i' and 'ii' in Fig. 1b). When *Bowenia* and *Stangeria* are constrained together in the full 11-taxon tree (i.e., with all taxa included), the shortest ML tree satisfying this constraint is not significantly worse than the optimal ML arrangement in Fig. 2a ($P = 0.094$). The constrained clade *Bowenia*–*Stangeria* is then the sister group of the *Ceratozamia*–*Microcycas*–*Zamia* clade (not shown). When *Lepidozamia* and *Macrozamia* are constrained together, an arrangement seen in the morphological analysis of Hermsen et al. (2006), the shortest ML tree satisfying this constraint is otherwise identical to the shortest tree (Fig. 2a), but is significantly worse ($P = 0.016$).

4. Discussion

The large gene sampling considered here has proven to be useful for characterizing a variety of deep phylogenetic splits in the seed plants, including the backbone of conifer phylogeny and multiple aspects of basal angiosperm phylogeny (e.g., Graham and Olmstead, 2000; Graham et al., 2006; Saarela et al., 2007; Rai et al., in press), permitting retrieval of well supported inferences of deep phylogenetic splits. The exemplar taxa that we surveyed represent all major cycad lineages, and broadly span the phylogenetic backbone of living and extinct cycads. Most of the fossil cycad genera belong within the crown clade, and they appear to be broadly dispersed among the extant lineages (Hermsen et al., 2006). If it could be obtained, a robust phylogeny of the major lineages of cycads would therefore have considerable value in aiding our understanding of the overall pattern of cycad evolution.

We characterized the common higher-order branching structure among seven of the ten or so currently accepted genera in other recent studies (Fig. 1a). The MP and ML inferences based on our plastid data set are completely congruent with this subtree topology (cf. Figs. 1a and 2), and also with each other in terms of their well supported structure (Fig. 2). The broad congruence among these multiple nearly independent data sets (our data and the previously published data sets based on morphology and plastid and nuclear data), supports the idea that the subtree in Fig. 1a accurately reflects the true branching relationships among these seven cycad genera. (It should be noted that our plastid data set

is largely non-overlapping with other recent molecular studies in terms of the regions examined, although *rbcl* is common to our study and Hill et al., 2003.) The three genera pruned to generate the agreement subtree (i.e., *Bowenia*, *Dioon* and *Stangeria*) generally have variable or weakly to moderately supported positions in recent studies, and here also (Fig. 2). However, we can rule out two-thirds of the possible suboptimal placements for each of these genera (the unlabeled branches in Fig. 1b), and so their positions should be considered to be partly localized. Despite the moderately strong bootstrap support for the placement of *Dioon* as sister to all cycads except *Cycas* (86–92% bootstrap support in MP vs. ML analyses; Fig. 2a and b), and of *Stangeria* as part of a four-taxon clade (89–98% bootstrap support; Fig. 2a and b) we cannot rule out four of 12 possible suboptimal placements of each genus on the cycad backbone, according to the Shimodaira–Hasegawa test. This may reflect the fact that this test is quite conservative (e.g., Susko, 2003).

However, it is clear that more data will be needed to firmly place *Bowenia*, *Dioon* and *Stangeria*. Until we can settle their placement with more confidence, it would be advisable to postpone further adjustments of higher-order cycad classification. It should be noted, however, that all of the feasible arrangements for the three floating genera found here (Fig. 1b) are consistent with a conservative taxonomic treatment of two families of cycads: Cycadaceae (for *Cycas* alone) and Zamiaceae (for all other extant genera), e.g., Hill et al. (2003), Chaw et al. (2005). The family Stangeriaceae (= *Bowenia* + *Stangeria*; Stevenson, 1992) was recovered as monophyletic in the morphological analysis of Hermsen et al. (2006) but has not been inferred in molecular studies. Trees inferred with this clade constrained are not significantly longer than the best ML tree here ($P = 0.094$). The sister group relationship between *Lepidozamia* and *Macrozamia* seen in Hermsen et al. (2006) was weakly supported here by the noncoding plastid data (66% bootstrap support; data not shown), but is firmly rejected when all plastid data are combined ($P = 0.016$). The *Lepidozamia*–*Macrozamia* clade was strongly supported in an analysis of the nuclear 26S rDNA gene (Hill et al., 2003). However, a combined analysis of the nuclear ITS and 26S rDNA locus in Hill et al. weakly supports the plastid arrangement. The apparent discordance between plastid data and the 26S locus (and perhaps between 26S and ITS regions) would be worth examining by considering additional nuclear loci, to determine whether either resolution is spurious. It would also be valuable to collect additional noncoding plastid data to assess whether weak support for the *Lepidozamia*–*Macrozamia* clade from this data partition is spurious.

The Shimodaira–Hasegawa tests allowed us to more fully characterize the uncertainty in the placement of the three taxa on the core backbone of cycad phylogeny from the perspective of our large plastid data set. We suspect that much of the uncertainty in the placement of these three taxa is a function of several very short internal (non-terminal) branches inferred when these taxa are included (Fig. 2a), in tandem with the generally very slow rate of plastid genome evolution in cycads. It should be noted that as crown-clade cycad genera are known from the Permian (e.g., Hermsen et al., 2006), even the shortest branches observed here (Fig. 2a) likely correspond to quite substantial time spans. Recent authors (e.g., Chaw et al., 2005) propose that the relatively long terminal branch subtending *Stangeria* in cycad phylogeny contributes to uncertainty over its placement. However, given the very low overall substitution rates observed in cycads (Rai et al., 2003; Wu et al., 2007) we feel that it is unlikely that long-branch attraction (or saturation in general) is much of a problem in inference of cycad phylogeny, at least from the perspective of plastid data. Very conservative data should minimize the effects of otherwise problematic long branches (e.g., Felsenstein, 1983).

Our plastid gene survey includes a substantial fraction of the plastid genome (approximately a ninth of the non-repetitive total),

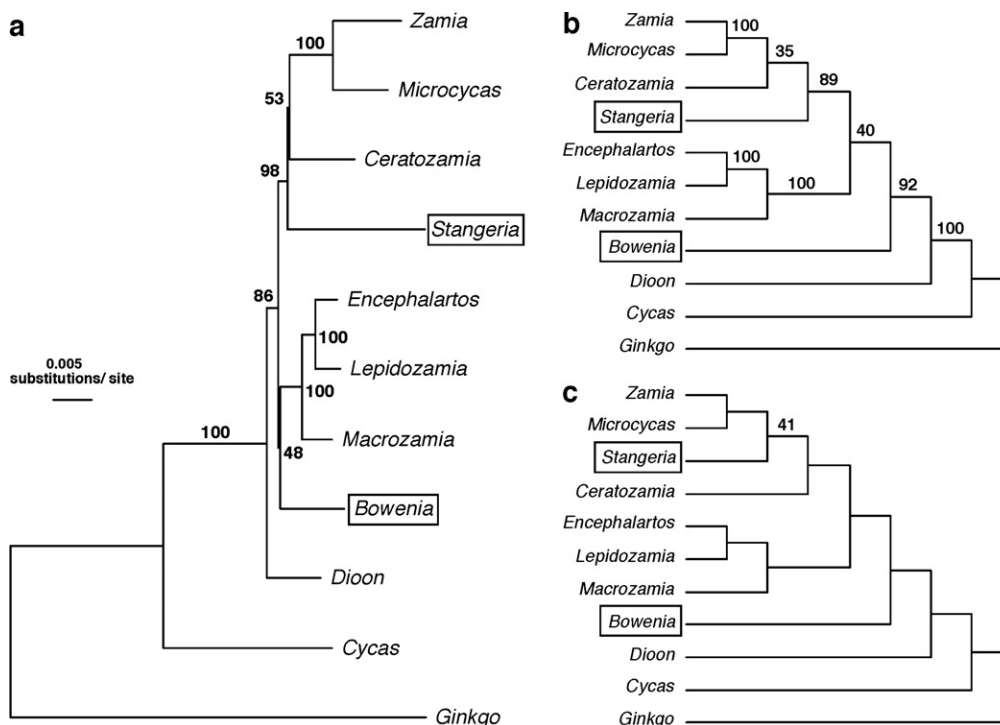


Fig. 2. Plastid-based phylogeny of the cycad genera inferred using 17 plastid genes and associated noncoding regions (three introns and eight intergenic spacer regions); *Ginkgo biloba* = outgroup. (a) Single tree found in ML analysis, depicted as a phylogram ($-\ln L = 32,574.23$). (b and c) Two shortest trees found using MP, depicted as cladograms (length = 2258 steps, CI = 0.864, RI = 0.567). The two highlighted genera (*Bowenia* and *Stangeria*) differ in position among the three trees. Bootstrap values are indicated near branches.

but still does not provide enough variation to rule out several alternative placements for *Bowenia*, *Dioon* and *Stangeria*. Because this genome is slowly evolving in cycads (and in *Ginkgo*; Rai et al., 2003; Wu et al., 2007) it would be valuable to collect more data from it to place these taxa with confidence. Our analyses clarify what work remains to be done. Clearly, additional data from multiple sources (e.g., nuclear data, additional morphological data) would be invaluable for inferring the remaining uncertainties concerning cycad higher-order phylogeny. However, it is becoming reasonably straightforward and inexpensive to obtain whole plastid genomes (e.g., Wu et al., 2007), and so an excellent case can be made for considerably expanding the plastid genomic sampling to encompass the entire plastid genome, for at least the current exemplar taxon sample.

5. Conclusion

Our current plastid tree is consistent with the greatest agreement subtree inferred from other recent studies, and rejects two-thirds of the possible positions of three cycad genera (*Bowenia*, *Dioon* and *Stangeria*) that tend to float in these other studies. Despite uncertainty in the placements of *Bowenia*, *Dioon* and *Stangeria*, several cycad clades are strongly supported by the plastid data, including one that defines a sister group relationship between *Cycas* and all other cycads, a *Microcycas*–*Zamia* clade, and (*Macrozamia* (*Encephalartos*, *Lepidozamia*)). The monophyly of the family Stangeriaceae cannot be rejected with the large plastid data set, but a clade consisting of *Lepidozamia* and *Macrozamia*, observed in some other studies, is strongly rejected.

Acknowledgments

The authors thank Lisa Ostafichuck and Patricia Murray for technical assistance, and two anonymous reviewers for their com-

ments. This research was supported by an NSERC Discovery Grant to S.W.G. and by NSERC PGSM and Alberta Ingenuity Scholarships to J.M.Z.

References

- Bogler, D.J., Francisco-Ortega, J., 2004. Molecular systematic studies in cycads: evidence from *trnL* intron and ITS2 rDNA sequences. *Bot. Rev.* 70, 260–273.
- Brenner, E.D., Stevenson, D.W., Twigg, R.W., 2003. Cycads: evolutionary innovations and the role of plant-derived neurotoxins. *Trends Plant Sci.* 8, 446–452.
- Burleigh, J.G., Mathews, S., 2004. Phylogenetic signal in nucleotide data from seed plants: implications for resolving the seed plant tree of life. *Am. J. Bot.* 91, 1599–1613.
- Burleigh, J.G., Mathews, S., 2007a. Assessing among-locus variation in the inference of seed plant phylogeny. *Int. J. Plant Sci.* 168, 111–124.
- Burleigh, J.G., Mathews, S., 2007b. Assessing systematic error in the inference of seed plant phylogeny. *Int. J. Plant Sci.* 168, 125–135.
- Chaw, S.-M., Walters, T.W., Chang, C.-C., Hu, S.-H., Chen, S.-H., 2005. A phylogeny of cycads (Cycadales) inferred from chloroplast *matK* gene, *trnK* intron, and nuclear rDNA ITS region. *Mol. Phylogenet. Evol.* 37, 214–234.
- Chen, C.-J., Hill, K.D., Stevenson, D.W., 2004. Comments on *Cycas*, *Dyerocycas* and *Epicycas* (Cycadaceae). In: Walters, T., Osborne, R. (Eds.), *Cycad Classification: Concepts and Recommendations*. CABI Publishing, Cambridge, MA, pp. 57–68.
- Donaldson, J., Hill, K.D., Stevenson, D.W., 2003. Cycads of the world: an overview. In: Donaldson, J. (Ed.), *Cycads. Status Survey and Conservation Action Plan*. IUCN/SSC Cycad Specialist Group, IUCN, Gland, Switzerland and Cambridge, UK, pp. 3–19.
- Doyle, J.A., 2006. Seed ferns and the origin of angiosperms. *J. Torrey Bot. Soc.* 133, 169–209.
- Felsenstein, J., 1983. Parsimony in systematics: biological and statistical issues. *Annu. Rev. Ecol. Syst.* 14, 313–333.
- Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791.
- Gao, Z.-F., Thomas, B.A., 1989. A review of fossil cycad megasporophylls, with new evidence of *Crossozamia* Pomel and its associated leaves from the lower Permian of Taiyuan, China. *Rev. Palaeobot. Palynol.* 60, 205–223.
- Goldman, N., Anderson, J.P., Rodrigo, A.G., 2000. Likelihood-based tests of topologies in phylogenetics. *Syst. Biol.* 49, 652–670.
- Graham, S.W., Olmstead, R.G., 2000. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *Am. J. Bot.* 87, 1712–1730.
- Graham, S.W., Reeves, P.A., Burns, A.C.E., Olmstead, R.G., 2000. Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of

- indels and inversions in basal angiosperm phylogenetic inference. *Int. J. Plant Sci.* 161, S83–S96.
- Graham, S.W., Zgurski, J.M., McPherson, M.A., Cherniawsky, D.M., Saarela, J.M., Horne, E.F.C., Smith, S.Y., Wong, W.A., O'Brien, H.E., Biron, V.L., Pires, J.C., Olmstead, R.G., Chase, M.W., Rai, H.S., 2006. Robust inference of monocot deep phylogeny using an expanded multigene plastid data set. In: Columbus, J.T., Friar, E.A., Porter, J.M., Prince, L.M., Simpson, M.G. (Eds.), *Monocots: Comparative Biology and Evolution (Excluding Poales)*. Rancho Santa Ana Botanic Garden, Claremont, CA, pp. 3–21.
- Hermesen, E.J., Taylor, T.N., Taylor, E.L., Stevenson, D.W., 2006. Cataphylls of the middle Triassic cycad *Antarcticycas schopfii* and new insights into cycad evolution. *Am. J. Bot.* 93, 724–738.
- Hill, K.D., Chase, M.W., Stevenson, D.W., Hills, H.G., Schutzman, B., 2003. The families and genera of cycads: a molecular phylogenetic analysis of Cycadophyta based on nuclear and plastid DNA sequences. *Int. J. Plant Sci.* 164, 933–948.
- Hilton, J., Bateman, R.M., 2006. Pteridosperms are the backbone of seed-plant phylogeny. *J. Torrey Bot. Soc.* 133, 119–168.
- Maddison, D.R., Maddison, W.P., 2001. *MacClade* version 4.03. Computer Program and Documentation. Sinauer Associates, Sunderland, MA.
- Posada, D., Crandall, K.A., 1998. MODELTEST: Testing the model of DNA substitution. *Bioinformatics* 14, 817–818.
- Rai, H.S., O'Brien, H.E., Reeves, P.A., Olmstead, R.G., Graham, S.W., 2003. Inference of higher-order relationships in the cycads from a large chloroplast data set. *Mol. Phylogenet. Evol.* 29, 350–359.
- Rai, H.S., Reeves, P.A., Peakall, R., Olmstead, R.G., Graham, S.W., in press. Inference of higher-order conifer relationships from a multi-locus plastid data set. *Botany*.
- Rambaut, A., 1998. *Se-Al* (Sequence Alignment Editor Version 1.0), Computer Program and Documentation. Department of Zoology, University of Oxford, UK.
- Saarela, J.M., Rai, H.S., Doyle, J.A., Endress, P.K., Mathews, S., Marchant, A.D., Briggs, B.G., Graham, S.W., 2007. Hydatellaceae identified as a new branch near the base of the angiosperm phylogenetic tree. *Nature* 446, 312–315.
- Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* 16, 1114–1116.
- Stevenson, D.W., 1992. A formal classification of the extant cycads. *Brittonia* 44, 220–223.
- Susko, E., 2003. Confidence regions and hypothesis tests for topologies using generalized least squares. *Mol. Biol. Evol.* 20, 862–868.
- Swofford, D.L., 1991. When are phylogenetic estimates from molecular and morphological data incongruent? In: Miyamoto, M.M., Cracraft, J. (Eds.), *Phylogenetic Analysis of DNA Sequences*. Oxford University Press, New York, pp. 295–333.
- Swofford, D.L., 2002. *PAUP**. *Phylogenetic Analysis Using Parsimony (and Other Methods)*. Version 4.0b10. Computer Program and Documentation. Sinauer Associates, Sunderland, MA.
- Wu, C.-S., Wang, Y.-N., Liu, S.-M., Chaw, S.-M., 2007. Chloroplast genome (cpDNA) of *Cycas taitungensis* and 56 cp protein-coding genes of *Gnetum parvifolium*: insights into cpDNA evolution and phylogeny of extant seed plants. *Mol. Biol. Evol.* 24, 1366–1379.