Phylogeny of the cycads based on multiple single-copy nuclear genes: congruence of concatenated parsimony, likelihood and species tree inference methods

Dayana E. Salas-Leiva1, Alan W. Meerow2,*, Michael Calonje1,3, M. Patrick Griffith3, Javier Francisco-Ortega1,4, Kyoko Nakamura2, Dennis W. Stevenson5, Carl E. Lewis3 and Sandra Namoff6

1 Department of Biological Sciences, Florida International University, Miami, FL 33199, USA, 2 USDA-ARS-SHRS, National Germplasm Repository, Miami, FL 33158, USA, 3 Montgomery Botanical Center, Miami, FL 33156, USA, 4 Fairchild Tropical Botanic Garden, Coral Gables, FL 33156, USA, 5 New York Botanical Garden, Bronx, NY 10458, USA and 6 Rancho Santa Ana Botanic Garden and Claremont Graduate University, Claremont, CA 91711, USA

* For correspondence. E-mail alan.meerow@ars.usda.gov

Received: 27 March 2013 Returned for revision: 17 June 2013 Accepted: 15 July 2013

INTRODUCTION

With ten genera and 331 currently accepted species (Osborne et al., 2012), cycads (order Cycadales) can contribute to understanding the origin and evolution of seeds, cones and plant vegetative structures (Frohlich and Parker, 2000; Brenner et al., 2003a, b). Cycads also hold important clues that can be used to infer early molecular evolution trends of seed plants (Zhang et al., 2003; Wang et al., 2007), as well as provide evolutionary insights concerning ancestral plant–animal interactions such as ancient pollination and herbivory mechanisms (Schneider et al., 2002; Brenner et al., 2003a; Kono and Hiroshi, 2007; Terry et al., 2007; Hummel et al., 2008; Butler et al., 2009; Peñalver et al., 2012).

Cycads have a fossil record dating back to the Lower Permian of China, reaching their peak in abundance and diversity in the Mesozoic (Martínez et al., 2012). Due to their long evolutionary history and retention of ancestral characters such as flagellated sperm, cycads are considered the most primitive extant seed plant lineage (Brenner et al., 2003b) and are often characterized as ‘living fossils’. However, recent studies suggest that extant species are the result of Neogene speciation events (Crisp and Cook, 2011; Nagalingum et al., 2011).

Despite their extraordinary scientific importance and widespread horticultural popularity, phylogenetic relationships within Cycadales still are not fully resolved. At the infrageneric level, phylogenies incorporating molecular data have been published for Ceratozamia (González and Vovides, 2002; De Castro et al., 2006), Cycas (Chiang et al., 2009; Sangin et al., 2010; Xiao et al., 2010), Dioon (Moretti et al., 1993; González et al., 2008; Moynihan et al., 2012), Encephalartos (Treutlein et al., 2005; Rousseau, 2012) and Zamia (Caputo et al., 2004). There has also been some interest in using DNA sequences for bar-coding of cycad species (Sass et al., 2007; Nicolalde-Morejon et al., 2011). These phylogenetic reconstructions have been mostly based on restriction site data from the plastid genome, amplified fragment length polymorphisms (AFLPs) and sequences of several regions of plastid DNA and nuclear ribosomal DNA internal transcribed spacer (nrDNA ITS). Plastid sequences to date have

Key words: Biogeography, Cycadales, Bowenia, Stangeria, Dioon, gymnosperms, molecular systematics.
yield either phylogenies with few resolved nodes (Caputo et al., 2004; Treutlein et al., 2005; Sangin et al., 2010) or results highly incongruent with other molecular or morphological data (De Castro et al., 2006; González et al., 2008; Xiao et al., 2010; Moynihan et al., 2012).

At the supra-generic level, questions still remain unresolved, particularly concerning the phylogenetic placement of three genera: Bowenia, Dioon and Stangeria. Phylogenies based on morphology, ITS sequences or several plastid regions yielded incongruent results for these three genera (e.g. Stevenson, 1990, 1992; Hill et al., 2003; Rai et al., 2003; Bogler and Francisco-Ortega, 2004; Chow et al., 2005; Zgurski et al., 2008; Nagalingum et al., 2011; Griffith et al., 2012). However, all of these phylogenetic studies provided strong support for: (1) Cycas as sister to the rest of the Cycadales; (2) Encephalartos, Lepidozamia and Macrozamia forming a monophyletic group; and (3) Microcycas and Zamia (including Chigua) as a distinct clade (Figs 1 and 2).

Since Doyle’s (1992) admonishment of single gene phylogenies as ‘one character taxonomy’, systematists have striven to develop concatenated super-matrices of numerous gene sequences, thereby emulating the ‘total evidence’ approach of Kluge (1989, 2004). Over the last 15 years, the growing concern that gene histories are not necessarily congruent with species histories became more precisely understood (Avise and Wollenberg, 1997; Doyle, 1997; Maddison, 1997; Avise, 2000), and problems associated with concatenation, even with mixed models applied to the partitions (Nylander et al., 2004), have been a focus of discussion (Degnan and Rosenberg, 2006, 2009; Edwards et al., 2007; Kubatko and Degnan, 2007; Edwards, 2008). The heterogeneity of gene trees, and their conflict with species trees, can be caused by several phenomena, including: gene duplication/extinction, horizontal gene transfer and hybridization, and incomplete lineage sorting (Maddison, 1997; Degnan and Rosenberg, 2009). Further, one or several partitions of a concatenated super-matrix may bias the tree reconstruction results by the sheer number of phylogenetically informative characters (Edwards et al., 2007). Only recently has there been development of phylogenetic software tools optimized for estimating species trees from multiple gene trees without concatenation (e.g., Ané et al., 2007; Baum, 2007; Liu and Pearl, 2007; Liu, 2008; Liu et al., 2008, 2009; Wehe et al., 2008; Heled and Drummond, 2010; Maddison and Maddison, 2011). Such methods allow gene tree heterogeneity and estimate topologies which truly represent lineages of populations and species, rather than genes.

Single-copy nuclear genes (SCNGs) have been shown to provide a valid alternative to nrDNA and chlorolast DNA (cpDNA) regions for plant phylogenetic reconstructions (e.g. Popp and Oxelman, 2007; Roncal et al., 2008; Meerow et al., 2009; Rousseau-Gueutin et al., 2009; Albach and Meudt, 2010; Zhang et al., 2012; Zimmer and Wen, 2013). The utilization of multiple SCNGs can provide a whole selection of unlinked and independent characters which are extremely valuable in phylogenetics (Small et al., 2004). However, SCNGs have not been widely used mostly because of problems with isolation and amplification, and difficulty in distinguishing between paralogous and orthologous copies (Mort and Crawford, 2004).

**Fig. 1.** Phylogeny of Cycadales based on cladistic analysis of 30 morphological characters (Stevenson, 1990) and formal classification (Stevenson, 1992).
Fig. 2. Molecular phylogenetic trees of Cycadales from various authors. Numbers below branches are bootstrap percentages ≥50 unless otherwise stated. (A) One of the two most-parsimonious trees from analysis of 17 plastid genes and non-coding regions (Rai et al., 2003). The second tree exchanged positions for Bowenia and Encephalartos (double-headed arrow). (B) One of the 81 most-parsimonious trees from analysis of concatenated plastid rbcL and trnL-F, and nrDNA ITS and 26S (Hill et al., 2003). (C) One of the 45 most-parsimonious trees from analysis of plastid matK (Chaw et al., 2005). Arrows indicate nodes that collapse in the strict consensus. With nrDNA ITS, the authors found the four most-parsimonious trees, the strict consensus of which was fully resolved at the generic level. Bootstrap values in parentheses are from the ITS analysis. (D) Single most-parsimonious tree found by Bogler and Francisco-Ortega (2004) using combined plastid trnL-F and nrDNA ITS2. (E) One of the two most-parsimonious trees found in the analysis of Zgurski et al. (2008) of the same 17 plastid genes and non-coding regions used by Rai et al. (2003). (F–H) Maximum likelihood trees from Nagalingum et al. (2011). (F) Plastid rbcL and matK. (G) Nuclear PHYP. (H) Combined plastid and nuclear genes. (I) Cladogram of Cycadales inferred from the maximum clade credibility chronogram of the gymnosperms using plastid matK and 25S nrDNA (Crisp and Cook, 2011). Numbers below branches indicate those branches with posterior probabilities <1.0. (J) Strict consensus of the 96 most-parsimonious trees based on analysis of combined plastid matK, rpoB and rps4, mitochondrial mtITS and nad1, nrITS, nuclear NEEDLY sequences and 130 morphological characters (Griffith et al., 2012).
These molecular markers have rarely been used to resolve phylogenetic relationships within the Cycadales (e.g. Nagalingum et al., 2011; Moynihan et al., 2012), although at least six of them have been isolated and identified for Cycas or Zamia [i.e. vicilin-like (Braun et al., 1996), the largest subunit of RNA polymerase II (Nickerson and Drouin, 2004), the second largest subunit of RNA polymerase II (Oxelman et al., 2004), APETALA/EREBP (Shigyo et al., 2006), Floricaula/LEAFY (Frohlich and Parker, 2000) and Cycas-AGAMOUS (hereafter CyAG; Zhang et al., 2004)].

In this study, we apply five SCNGs to the phylogeny of the order Cycadales. We specifically aim to evaluate several gene tree–species tree reconciliation approaches for developing an accurate phylogeny of the order, contrasting them with concatenated parsimony and maximum likelihood (ML) analyses, and hopefully resolve the erstwhile problematic phylogenetic position of Bowenia, Dioon and Stangeria. We discuss our results in the contexts of clade age estimation, the cycad fossil record and ancestral area reconstruction.

**MATERIALS AND METHODS**

**Taxonomic sampling**

Twenty representative taxa of Cycadales were used in this study (Table 1). Up to three species were sampled in each genus to allow branch and bound (B&B) parsimony searches and to keep the duration of other analyses within a reasonable time frame (genera with one or two species were completely sampled). Several studies have indicated that sequence length and the number of loci are much greater factors in phylogenetic accuracy than the number of taxa (Rosenberg and Kumar, 2001; Rokas and Carroll, 2005), and that a denser taxon sample does not always reduce a pre-existent topological conflict (Zhao et al., 2013). There is agreement among cycad and gymnosperm specialists that the cycad genera with multiple species (i.e. Bowenia, Ceratozamia, Cycas, Dioon, Encephalartos, Lepidozoama, Macrozamia and Zamia) are monophyletic (Stevenson, 1990, 1992; Hill et al., 2003; Rai et al., 2003; Bogler and Francisco-Ortega, 2004; Chaw et al., 2005; Zgurski et al., 2008; Crisp and Cook, 2011; Nagalingum et al., 2011). All broad geographic areas in the range of the genera are represented by the samples in our study, with the exception of the outgroup Cycas, which has a single species in Africa and Madagascar (C. thouarsii R. Br. ex Gaudi), but extends throughout Southeast Asia, India, China, Australia, and the western Pacific.

**DNA extraction, gene amplification and sequencing**

DNA extraction was performed using a FastDNA kit (MP Biomedicals, Santa Ana, CA, USA) or a DNeasy Plant Mini kit (Qiagen, Valencia, CA, USA) and 20–100 mg of dry leaf tissue. Five SCNGs (CyAG, COS26, GroES, GTP and HTS) were amplified using primers developed in our laboratories (see Supplementary Data Table S1). PCRs consisted of: 30 ng of total DNA, 1× PCR buffer, 0.2 mM of dNTPs, 0.2 μg mL⁻¹ of bovine serum albumin (BSA) or 1× TBT-PAR (Samarakoon et al., 2013), 0.2 μM of each primer and 0.05 U μL⁻¹ of Taq polymerase (New England Biolabs, Ipswich, MA, USA).
All the PCRs (except HTS) were carried out according to the following temperature profile: 95 °C for 2 min, 35 cycles of 95 °C for 30 s, annealing temperature (50–60 °C) for 1 min and 72 °C for 1 min, and final extension at 72 °C for 7 min. HTS was amplified using touchdown PCR (95 °C for 2 min; three cycles of 95 °C for 30 s, 58 °C for 1 min, 72 °C for 1 min; three cycles of 95 °C for 30 s, 57 °C for 1 min, 72 °C for 1 min; three cycles of 95 °C for 30 s, 56 °C for 1 min, 72 °C for 1 min; 30 cycles of 95 °C for 30 s, 54 °C for 1 min, 72 °C for 1 min; and final extension of 72 °C for 10 min). PCR products were checked using 1:2 % agarose gels with gel red (Biotium, Inc., Hayward, CA, USA) and a size standard ladder (100 bp New England Biolabs). Amplified products were cleaned up using exonuclease I (New England Biolabs) and shrimp alkaline phosphatase (USB Products- Affymetrix, Santa Clara, CA, USA), incubating at 37 °C for 1 h, followed by 80 °C for 20 min. The sequencing was done using ABI Big Dye Terminator v3-1 chemistry (Applied Biosystems, Carlsbad, CA, USA) followed by ethanol clean up. Labelled fragments were visualized on an ABI 3730 Automatic DNA Sequencer (Applied Biosystems). The nucleotide sequences were manually edited with Sequencher 4.9 (Gene Codes Corporation, Ann Arbor, MI, USA).

Phylogenetic methods

Cycas was used as the outgroup for all analyses, as we were only able to amplify DNA of Gingko with our CyAG primers. All previous phylogenetic analyses of Cycadales (e.g. Chaw et al., 2005; Zgurski et al., 2008; Nagalingum et al., 2011) have resolved Cycas as sister to all other genera in the order, thus its designation as the outgroup is appropriate. Sequences were aligned using MAFFT v.6 (Miyata et al., 2002) and/or manually in Sequencher 4.9. Nucleotide substitution models for each partition were evaluated with KAKUSAN 4.0 (Tanabe, 2007), and the corrected Akaike information criterion (AICc) (Sugiura, 1978) was used for model selection. Parsimony analysis of the concatenated matrix was conducted in PAUP v. 4.10b (Swofford, 2004) using a B&B search (Hendy and Penny, 1982) with simple addition, followed by generation of jackknife (JK) support values (1000 B&B iterations), and both total and partitioned Bremer indices, the latter with TreeRot v3.0 (Sorenson and Franzosa, 2007) and PAUP, using heuristic searches with 1000 rounds of random addition sequence and TBR branch swapping. B&B analyses were also conducted on each gene alone. The ML analyses were completed using Treefinder (Jobb, 2011). As the most likely model of nucleotide substitution was the same for all loci, a replicated (500 iterations) non-partitioned analysis was performed with bootstrap (1000 rounds). Species tree reconciliation analyses were performed with *BEAST (Heled and Drummond, 2010) as implemented in BEAST v.1.7.4 (Drummond et al., 2012); BEST (Liu, 2008), a modification of Mr. Bayes 3-1.2; and Bayesian Concordance Analysis (BCA; Ané et al., 2007) with BUCKy v.1.4.2 (Larget et al., 2010). A concordance factor can be defined as the proportion of the genome sampled that supports a given node in a species tree (Ané et al., 2007; Baum, 2007).

A partitioned analysis was run in *BEAST with 100 million Markov chain Monte Carlo (MCMC) iterations, under an uncorrelated relaxed clock (Drummond et al., 2006), with random starting trees for each partition generated under a constant population size coalescent model. A Yule model was applied as prior for the species tree likelihood. The MCMC statistics and trees were sampled every 1000 iterations. A maximum clade credibility (MCC) consensus species tree was created from 80 000 trees (100 000 trees saved minus 20 000 burn-in).

MrBayes v. 3.2.1 (Ronquist et al., 2012) was used to generate tree files for BCA in BUCKy. Two replicate analyses were run on each partition, with 10 million iterations and four chains, sampling every 500 iterations, and a burn-in of 1000 for summarizing posterior samples of both parameter values and trees. The resulting tree files from the two runs for each partition contained a total of 40 000 trees which were transformed to BUCKy infiles with a further burn-in of 10 000. BUCKy was run four times for 10 million MCMC updates each with four chains.

BEST was run twice concurrently for 100 million MCMC rounds with 32 chains on a 32-core parallel processing server. The log was written to and a tree sampled each 2500 iterations. The total number of trees sampled across both runs was 80 000, 50 % of which were discarded as further burn-in before generating an MCC consensus species tree.

Divergence age estimation

BEAST was used to perform an age estimation of divergence. Three fossils used for stratigraphic calibration points were from Hermen et al. (2006) as used similarly by Nagalingum et al. (2011) with the following priors: stem node of Bowenia [lower = 33-9, upper = 265-7 million years (My)], stem node of Lepidopteris [lower = 33-9, upper = 265-7 My] and stem node of Dioon (lower = 55-8, upper = 265-7 My). The only monophyly constraint was placed on the outgroup, Cycas. A non-partitioned analysis was run since KAKUSAN found the same highest likelihood model for all five loci (see the Results). A random starting tree was used, and a random local clock model was applied for determination of tree likelihood, as all other models resulted in zero tree likelihood for the starting trees. A total of 100 million iterations were run in BEAST, with log and tree samples every 2500th round. The total number of trees sampled was 40 000, 10 % of which were discarded as further burn-in before generating an MCC consensus species tree. Log output was evaluated in Tracer 1.5. TreeAnnotator in BEAST was used to generate MCC consensus trees from both BEAST and BEST output. All species trees were visualized with FigTree v.1.4 (Rambout, 2012).

Biogeographic analyses

Biogeographic analyses were conducted using RASP v.2.1a (Yu et al., 2010, 2011). Both the S-DIVA (Yu et al. 2010) and the dispersal–extinction–cladogenesis (DEC; Ree and Smith, 2008) methods were applied, using the MCC tree obtained from the BEAST age estimation analysis, and limiting ancestral area reconstructions to three to avoid all possible areas being assigned to deep nodes in the tree. Areas and their coding included in the analyses were A = Australia, B = China, C = Africa, D = Caribbean, E = Mexico, F = Northern Central America, G = South America and H = Southern Central America. Cycas angulata was used as the functional outgroup as only a single outgroup taxon can be designated in RASP.
RESULTS

Parsimony

CyAG was the only locus that alone produced a single fully resolved tree (Table 2; Supplementary Data Fig. S1). A B&B search of the five locus concatenated matrix found a single fully resolved tree (Fig. 3) identical to the CyAG topology. Each non-monotypic genus is monophyletic with 100% JK support at their crown nodes. Across 17 nodes of the ingroup, only two JK values were < 92% (nodes 2 and 16). Dioon is the first branch of the ingroup, with 100% JK value at the stem node. Bowenia is the next branch of the tree with 99% JK support at the stem. At the next node, two main clades are resolved with 63% JK support at the stem, the lowest support in the topology. One unites the Old World Zamiaceae (with the exception of Stangeria) with 100% JK support at the stem. Macrozamia is the first branch. Encephalartos and Lepidozamia are sister genera with 100% JK support. The second clade (92% JK support) unites the American Zamiaceae and Stangeria, with Ceratozamia as the first branch, and Stangeria as sister to Microcycas and Zamia with 100% JK support. Non-partitioned Bremer indices ranged from 2 to 364 across the 17 nodes of the ingroup, with all but three > 10 (Table 3). Both total and partitioned Bremer indices were weakest at nodes 2 and 16 (Table 3), which also received the lowest JK support.

Nucleotide substitution model

By the AICc, a general time-reversible (GTR) model with gamma correction was the best fit for all five loci. This model was applied to all MrBayes and BEAST analyses.

Maximum likelihood analyses

Maximum likelihood resolved the same topology as parsimony, with similar support scores (Fig. 3). The mean tree score (log likelihood) for the analysis = −19 645.89. As with all other analyses, the most weakly supported node in the tree was number 16, which resolves a sister relationship between the encephalartoid and zamiod sub-clades.

Species tree estimation

All three gene tree–species tree reconciliation approaches converged on exactly the same tree topology as the concatenated parsimony analysis (Fig. 3). The effective sample size (ESS) scores from the *BEAST analysis were all > 100, with most > 1000. Posterior values in the MMC consensus species tree were < 0.99 in only two nodes: the crown node for Old World and New World Zamiaceae (plus Stangeria) after the branching of Bowenia (node 16 in Fig. 3), and the stem node of Ceratozamia (node 15).

The MrBayes runs to develop gene tree sets for BUCKy achieved stationarity based on the internal diagnostics of MrBayes and examination in Tracer. The nodes in the MCC tree from the BCA had concordance factor (CF) scores of 1 in all but two nodes (Fig. 3): node 16 (CF = 0.91) and node 15 (0.99).

<table>
<thead>
<tr>
<th>Locus ID Description</th>
<th>Alignment length</th>
<th>Tree length</th>
<th>Consistency index (CI)*</th>
<th>Retention index (RI)</th>
<th>Homoplasy index (HI)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>CyAG MADS-box transcrioption factor family AGAMOUS</td>
<td>1871</td>
<td>759</td>
<td>0.7955</td>
<td>0.8565</td>
<td></td>
</tr>
<tr>
<td>COS26 Uncharacterized protein</td>
<td>752</td>
<td>204</td>
<td>0.8367</td>
<td>0.8828</td>
<td></td>
</tr>
<tr>
<td>GroES GroES-like zinc-binding alcohol dehydrogenase family</td>
<td>955</td>
<td>389</td>
<td>0.8442</td>
<td>0.8870</td>
<td></td>
</tr>
<tr>
<td>GTP GTP-binding protein Era mRNA</td>
<td>644</td>
<td>222</td>
<td>0.8534</td>
<td>0.9045</td>
<td></td>
</tr>
<tr>
<td>HTS Histidyl-tRNA synthetase</td>
<td>1377</td>
<td>2369</td>
<td>0.8623</td>
<td>0.9025</td>
<td></td>
</tr>
</tbody>
</table>

* Excluding uninformative sites

Table 2. Locus ID, locus description, alignment length, polymorphic sites, and parsimony informative sites per locus, number of trees found, tree length, homoplasy index, consistency index, and retention index.
The ESS scores in Tracer from the BEST run were $>1000$ for all statistics except overall likelihood (LnL), all five gene tree likelihoods (TL) and gene mutation rate (GeneMu). These latter were $<100$, despite a 3 week run on 32 cores. The suboptimal scores decreased if the burn-in of the log files in Tracer was increased, which suggested that the run duration should have been longer. We then increased the number of MCMC iterations to 150 million for a second run with little improvement of these same ESS scores. Despite this fact, the topology of MCC consensus species tree from BEST was exactly the same as for all the other analyses, thus we ceased running BEST at this point.

The posterior probability (PP) was 1.0 at only three nodes of the BEST MCC consensus species tree: node 16 (PP = 0.56), and nodes 2 and 15 (0.9 and 0.97, respectively).

**Fig. 3.** Congruent tree topology found by concatenated branch and bound parsimony, likelihood, and three gene tree–species tree reconciliation methods for the five locus SCNG matrix across Cycadales with *Cycas* as outgroup. Only support values $>1$-0 or 100 % are shown. Numbers above branches are ML bootstrap (underlined) and parsimony jackknife support values for the single most-parsimonious tree; those at nodes are posterior probabilities (PP) from the *BEAST* (Heled and Drummond, 2010) maximum clade credibility (MCC) consensus tree; those below branches are concordance factor scores from BUCKy (Larget et al., 2010) and PP (italic) from the BEST (Liu, 2008) MCC consensus species tree. Bremer (decay) indices are listed in Table 3 and reference the node numbers in black circles.
To estimate node ages, we carried out BEAST 2 with 500,000,000 generations, sampling every 10,000 generations, and a burn-in of 10,000,000. We evaluated model performance with multiple independent runs using the Tracer tool (available online).

### Table 3. Total and partitioned Bremer indices for concatenated parsimony tree of Cycadales

<table>
<thead>
<tr>
<th>Node</th>
<th>Non-partitioned</th>
<th>CyAG</th>
<th>COS26</th>
<th>GroES</th>
<th>GTP</th>
<th>HTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>115</td>
<td>47.5</td>
<td>13.5</td>
<td>32</td>
<td>8.5</td>
<td>21.5</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>6</td>
<td>0.5</td>
<td>10</td>
<td>0</td>
<td>1.5</td>
</tr>
<tr>
<td>4</td>
<td>53</td>
<td>28</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>48</td>
<td>19</td>
<td>1</td>
<td>11</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>67</td>
<td>32</td>
<td>7</td>
<td>11</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>7</td>
<td>115</td>
<td>47</td>
<td>12</td>
<td>28</td>
<td>8</td>
<td>20</td>
</tr>
</tbody>
</table>

See Fig. 1 for the key to node numbers.

Negative numbers signify incongruous support at that node for that partition. A value of zero indicated neutral support.

### Age estimation

The ESS scores in Tracer for the age estimation analysis in BEAST were all >200, with many >1000. The mean estimated age of the crown node of Cycadales is 228.8 My (Fig. 4), with a 95 % HPD of 197.2–270.6. The crown node of Zamiaceae has a mean age of 91.1 My, with 95 % HPD of 70.1–110.8. The stem node of Bowenia is dated to a mean of 74.8 My (95 % HPD = 56.4–91.0), and the split between the Old World Zamiaceae (less Stangeria) and the New World (plus Stangeria) is estimated to have occurred approx. 3 My later (71.9 My; 95 % HPD = 55.7–87.2), followed 4–5 My later by the branching of Ceratozamia and Stangeria–Microcycas–Zamia (67.4 My; 95 % HPD = 52.0–82.7). The most recent common ancestor (MRCA) of Stangeria and Microcycas–Zamia is dated to 60.3 My (95 % HPD = 46.3–74.2) and that of Microcycas and Zamia at 36.5 My (95 % HPD = 27.6–45.6). Within the Encephalartaeae, the branching of Macrozamia is estimated at 39.9 My (95 % HPD = 33.9–49.2). The MRCA of Encephalartos and Lepidozamia has a mean age of 32.9 My (95 % HPD = 25.8–41.7). Crown nodes of all genera, with the exception of Cycas and Dioon (mid-Miocene), are dated to the late Miocene or the Pliocene (Fig. 4), except Macrozamia (Pleistocene), with fairly narrow HPD ranges. However, the sampling of the larger cycad genera is limited. We tested removing all but one of the calibration points, respectively, and found a <10 % change in node ages in each run, but a corresponding increase in the length of the 95 % HPD intervals (not shown).

### Biogeographic analysis

The most likely area for the root node of the phylogeny is Australia–Mexico–Northern Central America (AEF) by the S-DIVA method, and Australia–China–Northern Central America (AEF) by DEC (Fig. 5). While both methods hypothesize two dispersal events and one extinction event subsequently (Table 4), probabilities are low, especially with DEC (P = 0.0082), though an ancestral presence in China (or somewhere in greater Asia) is more likely than in Mexico. In either case, there is little confidence for any area optimization for the root node.

The crown node of Cycas (38) is situated in Australia and China with subsequent vicariance between the two areas by both methods with P = 1, but the minimal sampling of this genus weakens the optimization.

The crown node of Zamiaceae (37) is optimized quite differently between the two methods (Table 4), neither with high probability. S-DIVA calls Australia and Northern Central America, with three subsequent dispersal events [Africa, twice to Mexico, and one vicariance (Australia–Africa–Mexico|Mexico–Northern Central America)], while DEC considers Australia–Africa–Mexico as most likely, with subsequent single dispersal, vicariance and extinction events, the latter eliminating an early presence in Africa, dispersal to Northern Central America, and ultimately vicariance between Australia and Mexico–Northern Central America.

The stem node of Bowenia (36) is optimized differently by the two methods, but with similar probabilities (Table 4). S-DIVA hypothesizes an ancestral area comprising Australia, Africa and Mexico, with extinction eliminating Africa, a re-entry into Australia and vicariance between Australia and Australia–Mexico. DEC hypothesizes a simpler model: stasis in Australia. For the crown node of Bowenia (35), stasis in Australia is the optimal scenario with both methods (P = 1).

Node 34 (Fig. 5) is the ancestral node of the tribes Encephalartaeae and Zamiaeae–Ceratozamiaeae [Stevenson’s (1992) Zamioideae with the inclusion of Stangeria]. Both methods hypothesize two dispersal events and one vicariance event, but with different routes. From an ancestral Australia–Mexico, S-DIVA resolves two dispersals to Africa and the Caribbean, respectively, and vicariance between Australia and Africa–Mexico–Caribbean (P = 0.0556). DEC optimizes two dispersals into Africa from ancestral Australia and later vicariance between the two (P = 0.0362).

The stem node of Ceratozamia (33) is resolved as a vicariance between Mexico and Africa–Mexico (P = 0.033). DEC envisages three dispersals, from Africa to Mexico, the Caribbean and again to Africa, with vicariance between Mexico and Africa–Caribbean (P = 0.0362). The crown node for Ceratozamia (32) is optimized as stasis in Mexico by S-DIVA (P = 1) or dispersal from Mexico to Central America by DEC (P = 0.2876).

Both area optimization methodologies converged on the same scenario for the crown node of Stangeria–Zamia–Microcycas (30), which involves vicariance between Africa and the Caribbean. The two regions together comprise the ancestral area. P = 1 with S-DIVA, and 0.2214 with DEC. The crown nodes of Zamia and Microcycas (29) were similar with both methods but differed in a second dispersal event with DEC (P = 0.4425), from the Caribbean to South America, while that of S-DIVA only included Southern Central America (P = 1). The crown node scenarios for Zamia (28) reflected this, in that DEC retained South America in the ancestral area.
with DEC ($P = 0.4069$), with the single vicariance between the Caribbean and Southern Central America–South America, vs. only Southern Central America and the Caribbean with S-DIVA ($P = 1$).

Optimization of the crown node of Encephalarteeae (26) was exactly the same with both methods (S-DIVA $P = 1$, DEC $P = 0.3881$) with dispersal to Africa from ancestry in Australia. The split between Lepidozamia and Encephalartos (25) was also the same in both S-DIVA ($P = 1$) and DEC ($P = 0.7144$), with vicariance between Australia and Africa.

The remaining nodes (21–24) had exactly the same scenarios and probabilities ($P = 1$) with both methods (Fig. 5, Table 3), involving geographic stasis in the ancestral area except for a vicariance event between the two ancestral areas of Dioon (21).

**DISCUSSION**

Supra-generic arrangements in the Cycadales and the placement of Bowenia, Dioon and Stangeria

Complete congruence of species tree topology from three different gene tree–species tree reconciliation approaches with the single tree resolved by parsimony analysis of the concatenated matrix (Fig. 3) supports the following conclusions concerning phylogenetic placement for three formerly *sedis incertae* genera of the Cycadales: (1) Dioon is sister to all other members of sub-order Zamiineae *sensu* Stevenson (1992); (2) Bowenia is the next branch in Zamiineae; and (3) Stangeria is sister to *Microcycas/Zamia*. Our tree topology (Fig. 3) is identical to that of Bogler and Francisco-Ortega (2004) based on combined nrDNA ITS2 and plastid *trnL-F* (Fig. 2C), the ITS
Fig. 5. Chronogram from age estimation analysis in BEAST (Drummond et al., 2012) optimized with the most likely area on internal nodes as determined by S-DIVA (Yu et al., 2010) and DEC (Ree and Smith, 2008) biogeographic analyses. Green circles around nodes indicate vicariance, D = dispersal, E = extinction. See Table 4 for details.
Table 4. Comparison of S-DIVA (Yu et al., 2010) and DEC (Ree and Smith, 2008) biogeographic analyses on the five gene sequence phylogeny of Cycadales

<table>
<thead>
<tr>
<th>Node</th>
<th>S-DIVA</th>
<th>DEC</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>Events: D: 0, V: 1, E: 0</td>
<td>Events: D: 0, V: 1, E: 0</td>
</tr>
<tr>
<td></td>
<td>Event route: EF → E</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>$P = 1$</td>
<td>$P = 1$</td>
</tr>
<tr>
<td>22</td>
<td>Events: D: 0, V: 0, E: 0</td>
<td>Events: D: 0, V: 0, E: 0</td>
</tr>
<tr>
<td></td>
<td>Event route: A → A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>$P = 1$</td>
<td>$P = 1$</td>
</tr>
<tr>
<td>23</td>
<td>Events: D: 0, V: 0, E: 0</td>
<td>Events: D: 0, V: 0, E: 0</td>
</tr>
<tr>
<td></td>
<td>Event route: A → A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>$P = 1$</td>
<td>$P = 1$</td>
</tr>
<tr>
<td>24</td>
<td>Events: D: 0, V: 0, E: 0</td>
<td>Events: D: 0, V: 0, E: 0</td>
</tr>
<tr>
<td></td>
<td>Event route: C → C</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>$P = 1$</td>
<td>$P = 1$</td>
</tr>
<tr>
<td>25</td>
<td>Events: D: 0, V: 1, E: 0</td>
<td>Events: D: 0, V: 1, E: 0</td>
</tr>
<tr>
<td></td>
<td>Event route: AC → C</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>$P = 1$</td>
<td>$P = 1$</td>
</tr>
<tr>
<td>26</td>
<td>Events: D: 1, V: 0, E: 0</td>
<td>Events: D: 1, V: 0, E: 0</td>
</tr>
<tr>
<td></td>
<td>Event route: A → A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>$P = 1$</td>
<td>$P = 1$</td>
</tr>
<tr>
<td>27</td>
<td>Events: D: 1, V: 0, E: 0</td>
<td>Events: D: 1, V: 0, E: 0</td>
</tr>
<tr>
<td></td>
<td>Event route: H → H</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td>$P = 1$</td>
<td>$P = 1$</td>
</tr>
<tr>
<td>28</td>
<td>Events: D: 0, V: 1, E: 0</td>
<td>Events: D: 0, V: 1, E: 0</td>
</tr>
<tr>
<td></td>
<td>Event route: DH → D</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td>$P = 1$</td>
<td>$P = 1$</td>
</tr>
<tr>
<td>29</td>
<td>Events: D: 1, V: 0, E: 0</td>
<td>Events: D: 1, V: 0, E: 0</td>
</tr>
<tr>
<td></td>
<td>Event route: D → D</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>$P = 1$</td>
<td>$P = 1$</td>
</tr>
<tr>
<td>30</td>
<td>Events: D: 0, V: 1, E: 0</td>
<td>Events: D: 0, V: 1, E: 0</td>
</tr>
<tr>
<td></td>
<td>Event route: CD → C</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>$P = 1$</td>
<td>$P = 1$</td>
</tr>
<tr>
<td>31</td>
<td>Events: D: 1, V: 0, E: 0</td>
<td>Events: D: 1, V: 0, E: 0</td>
</tr>
<tr>
<td></td>
<td>Event route: E → E</td>
<td>E</td>
</tr>
<tr>
<td></td>
<td>$P = 1$</td>
<td>$P = 1$</td>
</tr>
<tr>
<td>32</td>
<td>Events: D: 0, V: 0, E: 0</td>
<td>Events: D: 0, V: 0, E: 0</td>
</tr>
<tr>
<td></td>
<td>Event route: E → E</td>
<td>E</td>
</tr>
<tr>
<td></td>
<td>$P = 1$</td>
<td>$P = 1$</td>
</tr>
<tr>
<td>33</td>
<td>Events: D: 0, V: 1, E: 0</td>
<td>Events: D: 0, V: 1, E: 0</td>
</tr>
<tr>
<td></td>
<td>Event route: CDE → E</td>
<td>CD</td>
</tr>
<tr>
<td></td>
<td>$P = 0.333$</td>
<td>$P = 0.333$</td>
</tr>
<tr>
<td>34</td>
<td>Events: D: 2, V: 1, E: 0</td>
<td>Events: D: 2, V: 1, E: 0</td>
</tr>
<tr>
<td></td>
<td>Event route: AE → CDEA → CDE</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>$P = 0.0556$</td>
<td>$P = 0.0556$</td>
</tr>
<tr>
<td>35</td>
<td>Events: D: 0, V: 0, E: 0</td>
<td>Events: D: 0, V: 0, E: 0</td>
</tr>
<tr>
<td></td>
<td>Event route: A → A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>$P = 1$</td>
<td>$P = 1$</td>
</tr>
<tr>
<td>36</td>
<td>Events: D: 1, V: 0, E: 1</td>
<td>Events: D: 1, V: 0, E: 1</td>
</tr>
<tr>
<td></td>
<td>Event route: ACE → AE → AE</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>$P = 0.0417$</td>
<td>$P = 0.0417$</td>
</tr>
<tr>
<td>37</td>
<td>Events: D: 3, V: 1, E: 0</td>
<td>Events: D: 3, V: 1, E: 0</td>
</tr>
<tr>
<td></td>
<td>Event route: AF → AE</td>
<td>E</td>
</tr>
<tr>
<td></td>
<td>$P = 0.0625$</td>
<td>$P = 0.0625$</td>
</tr>
<tr>
<td>38</td>
<td>Events: D: 0, V: 1, E: 0</td>
<td>Events: D: 0, V: 1, E: 0</td>
</tr>
<tr>
<td></td>
<td>Event route: AB → A</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>$P = 1$</td>
<td>$P = 1$</td>
</tr>
<tr>
<td>39</td>
<td>Events: D: 2, V: 0, E: 1</td>
<td>Events: D: 2, V: 0, E: 1</td>
</tr>
<tr>
<td></td>
<td>Event route: AEF → AF → AFA → ABF</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>$P = 0.0625$</td>
<td>$P = 0.0082$</td>
</tr>
</tbody>
</table>

Global events

- Global dispersal: 12
- Global vicariance: 8
- Global extinction: 2

See Fig. 3 for node identity.

D = dispersal, V = vicariance, E = extinction, P = probability of events.
tree of Chaw et al. (2005) and the combined analysis of Griffith et al. (2012), but with improved support values.

Several morphological synapomorphies support the clades recovered in our study. For instance, the large sister clade to Dioon includes the vast majority of cycad genera, all of which have stomata on their sporangia (Dehgan et al., 1993). The presence of lateral lobes in the megasporophylls and vascular bundles in the pith defines the clade comprised of Encephalartos, Lepidozamia and Macrozamia (Stevenson, 1990). Unequivocally branched trichomes are shared only by Microcycas and Zamia (Stevenson, 1990). The sister relationship of African Encephalartos and Australian Lepidozamia is supported by common mucilage chemistry that is unique to these two genera (De Luca et al., 1982).

We are not aware of morphological synapomorphies supporting two of the main clades resolved in the species tree (Fig. 3): (1) Ceratozamia–Stangeria–Microcycas–Zamia; and (2) Stangeria–Microcycas–Zamia. The molecular topologies of Bogler and Francisco-Ortega (2004), Chaw et al. (2005), Zgurski et al. (2008), Nagalingum et al. (2011) and Crisp and Cook (2011) supported the Ceratozamia–Stangeria–Microcycas–Zamia clade. Not all of the gene topologies recovered in our study support this clade; indeed, based on the partitioned Bremer index analysis (Table 3, node 15), support for this group came mostly from three of the loci (i.e. CyAG, COS26 and GroES). Only the molecular studies by Bogler and Francisco-Ortega (2004), Chaw et al. (2005), Zgurski et al. (2008) in part, and the combined molecular and morphology analysis of Griffith et al. (2012) recovered the clade composed of Stangeria, Microcycas and Zamia. The partitioned Bremer index analysis (Table 3, node 14) showed that this group is mostly supported by the CyAG data set.

The last comprehensive systematic arrangement of the Cycadales is that of Stevenson (1992), based on cladistic analyses of morphological traits (Stevenson, 1990), with clades formally labelled from the rank of sub-order to sub-tribe (Fig. 1). More recently, Christenhusz et al. (2011), following the plastid-based sequence phylogeny of Zgurski et al. (2008), treated the Cycadales as consisting of two families: Cycadaceae including only the genus Cycles, and Zamiaceae including the rest of the genera with no recognized subfamilial ranks. No available molecular-based phylogenies (Fig. 2) support the placement of Stangeria and Bowenia within the same clade [i.e. Stangeriaceae in Stevenson’s (1992) classification]. In contrast, all available molecular phylogenies support the Cycadinae, Encephalartae and Zamiaceae [excluding Chigua, which has been sunk back into Zamia (Chaw et al., 2005; Lindstrom, 2009)] as three monophyletic groups, but are highly discordant concerning the monophyly of the other suprageneric taxa proposed by Stevenson (1992). Interestingly, all molecular studies placed the tribes Ceratozamiae and Zamiaceae in the same clade (Fig. 2); however, several of the recovered topologies showed Stangeria as an anomalous member of this clade (Hill et al., 2003; Rai et al., 2003; Bogler and Francisco-Ortega, 2004; Chaw et al., 2005; Zgurski et al., 2008; Nagalingum et al., 2011, in part).

The chronogram of Nagalingum et al. (2011) was based on a single nuclear gene (PHYP), which resolved Macrozamia as sister to a clade of Stangeria and the American Zamioidae sensu Stevenson (1992), but including Dioon. Only when two chloroplast regions were included in the matrix, despite their much lower taxonomic coverage, did Macrozamia resolve as it does in all of our gene trees and species trees (Fig. 3, Supplementary Data Fig. S1), as sister to the remainder of Encephalartae. However, none of the trees shown in Nagalingum et al. (2011) is identical to ours (Fig. 3).

Age estimations and biogeographic associations

The earliest known cycad fossils date to the Early Permian of north China, Crossozamia (Norstog and Nichols, 1997) and Pseudotenis (Pott et al., 2010), approx. 300–280 million years ago (Mya). Our mean crown age for modern Cycadales of approx. 230 My (95 % HPD = 271–179; Fig. 4), may thus be a reasonable estimate. This would also support a putative ancestral distribution in China for the order (Tang, 2004), which appears at the crown node of our optimization with DEC, at least in part (Fig. 5, Table 4). North American fossils from this era, previously assigned to Cycadales, are now considered to represent different but related orders (Anderson et al., 2007).

The Mesozoic was characterized by an increased occurrence and diversity of fossil cycads (Hermsen et al., 2009), broadly distributed throughout the relatively uniform climate of the supercontinent Pangaea. No relationships to living cycad genera have been proposed for these fossil genera (Anderson and Anderson, 1989). Fossils identified as Cycas, sister to the rest of the extant Cycadales, have been described from the late Cretaceous of Greenland (Osborne, 2002), and from the early Cenozoic of Japan and China (Liu et al., 1991). The lack of Cycas fossils from the Southern hemisphere suggests that the genus was absent from southern Pangaea. A crown age of 12 My for Cycas (Fig. 4) supports the hypothesis of Hill (1999) that the modern presence of the genus in Africa, Australia and the Pacific islands probably represents a relatively recent dispersal from the ancestral area.

The dating of the stem node of Dioon in the Cretaceous (Fig. 2), with area optimizations (Fig. 5) of Australia–Northern Central America (S-DIVA) or Australia–Africa–Mexico (DEC), probably reflects a broad distribution of an ancestral cycad flora across Laurasia, the northern half of Pangaea, which began to split apart in the Jurassic, of which Dioon remains a surviving relict. The lineage (tribe Diooeae) has fossils from the Triassic of western Laurasia (Lyssoxyylon) and southwestern Gondwana (Michelliolou) (Artabe et al., 2005). South American (Bororoa) elements of this tribe disappeared after the Cenozoic (Artabe et al., 2005).

Gondwana, the southern portion of Pangaea, also hosted a great diversity of cycads and cycad-like plants in the area that would become Australia (Delevoryas, 1975). Bowenia, the stem node of which is estimated at 75 My, may represent a surviving remnant of this diversity.

During the late Cretaceous period through the early Tertiary, the remaining major clades of extant cycads [Encephalartaeae and Zamioidae sensu Stevenson (1992) + Stangeria] were established. The dispersal/vicariance is dated at approx. 72 My (Fig. 4), with an ancestral distribution in Australia with DEC or Australia–Mexico with S-DIVA (Fig. 5, Table 4). This split may represent the termination of direct exchange between Laurasia and Gondwana.
Evaluated on the basis of extant distribution of genera, the Encephalartea would appear to represent an eastern Gondwanan lineage the stem age of which (approx. 72 Mya) suggests was initially isolated by the break up of that continent. While the sister relationships of Australian endemic Lepidozamia and the African endemic Encephalartos is optimized by both S-DIVA and DEC as vicariance between Africa and Australia (Fig. 5, Table 4), by the late Paleogene, the stem age of this clade, the two continents were already situated well apart from each other (Raven and Axelrod, 1974). This relationship does not fit any of the Southern hemisphere biogeographic scenarios recognized by Sammartin and Ronquist (2004), and for now remains an enigma. However, fossil remains for the Encephalartea (Cantrill, 2000; Artabe et al., 2004, 2005; Martínez et al., 2012) have been found from the Jurassic of India (Fascevarioxylon), the Cretaceous of Antarctica (Centriclexs) and Argentina (Neochamberlainia, Worsdellia and Wintuccycas), and the Tertiary of Argentina (Menuccia). It has been suggested that the tribe originated in the Triassic of western Laurasia (Charmorgia; Artabe et al., 2005), but this latter interpretation implies Dioon as a member of the group (Zamiaceae subfamily Encephalartoidae), which it most certainly is not (Fig. 3).

The asteroid impact at the Cretaceous/Tertiary (K–T) boundary, approx. 65.5 Mya (Schulte et al., 2010), resulted in extinction of nearly a third of terrestrial vegetation and great declines in species abundance (Nichols and Johnson, 2008). The extinction rate in tropical North America may have been as high as 60–70 % (Nichols and Johnson, 2008). The effects of the K–T boundary impact on cycads are not as clear, but the time period represented a major diversification time for the Cycadales. The branch lengths of nodes dated from approx. 75–60 Mya are very short (Fig. 4), suggesting rapid diversification (Kubatko and Degnan, 2007), perhaps in this case associated with drastic environmental changes. Tang (2012) hypothesizes that modern cycad genera in the Americas represent lineages that evolved from the survivors of the K–T extinction. The stem nodes of Ceratozamia (67 Mya) and the Zamia–Microcycas–Stangeria clade (60 Mya) may reflect the influence of this catastrophic event. North American leaf remains identify cycads have been dated to the first half of the Paleogene, apparently representing taxa that were not known from before the K–T boundary (Mustoe, 2008).

Fossils from the early Cenozoic have been variously related to modern taxa, e.g. Dioonopsis, which has been classified as Dioon, Ceratozamia or Zamia, but has unique cuticular characteristics (Erdei et al., 2012). An undescribed fossil from an Eocene deposit in Oregon resembles Dioon (Manchester, 1981; Tang, 2012), at least superficially. Younger fossils from Oligocene and Miocene strata in Europe bear the diagnostic cuticle morphology of Ceratozamia (Kvaček, 2002, 2004). Pseudodioon, a Miocene fossil from Turkey, bears macro-morphological characteristics of Dioon and anatomical features of Cycas (Erdei et al., 2010). Eostangeria, described from Paleocene deposits in Wyoming, and Eocene fossils from Europe and Oregon, greatly resembles Stangeria but has divergent cuticle morphology (Kvaček and Manchester, 1999; Uzunova et al., 2001). A recent fossil from Patagonia, Argentina pushes a leaflet mid-ribbed zamioid lineage to the early Cretaceous (Passalia et al., 2010).

In the late Eocene, there were major extinction events linked to global cooling and a decrease in atmospheric carbon dioxide (Jaramillo et al., 2006; Zachos et al., 2008; Kunzig, 2011). These climatic changes led to latitudinal shifts of the main vegetation belts of the planet and probably the elimination of cycads from higher latitudes of North America and Eurasia (Tang, 2012).

Our results indicate that extant diversification of cycad species occurred in the relatively recent past (Fig. 4), in agreement with Nagalingum et al. (2011). No crown node of any modern genus of cycad in our chronogram has a mean date estimate >12 Mya. Crisp and Cook (2011) also showed that most of the current species diversification within the Cycadales occurred relatively recently, although many of their age estimates placed these speciation events in the Eocene and Oligocene rather than in the Miocene.

The sister relationship between the South African endemic monotypic Stangeria eriopus and Zamia–Microcycas, dated to the early Paleocene but with a 95 % HPD that extends back into the late Cretaceous (Fig. 4), is also enigmatic, implying an African–Caribbean vicariance at the ancestral node (Fig. 5, Table 4). Mesodescolea, a Cretaceous fossil genus described from Argentina, has some features in common with Stangeria (Artabe et al., 2004), as does the Laurasian Eostangeria (Kvaček and Manchester, 1999; Uzunova et al., 2001). Eostangeria also has elements in common with Zamia (Uzunova et al., 2001). We conclude that S. eriopus represents the only surviving branch of a lineage that was once more widely distributed. The available fossil evidence suggests that while Gondwana once enjoyed a cosmopolitan tropical cycad flora from east to west (Sabato, 1990; Artabe and Stevenson, 1999; Tang, 2006), the South American elements of this flora were extirpated as Africa and South America separated, while in Africa and Australia, related taxa evolved and persisted.

The split between the Cuban endemic Microcycas and the more broadly distributed Zamia is estimated in our chronogram at approx. 36 Mya when Cuba was already a well-established and isolated island (Graham, 2003a, b). This is more or less the same time in the late Eocene/early Oligocene to which the stem nodes of the modern Encephalarteeae are dated (Fig. 4), a period of global climate change (Jaramillo et al., 2006) that seems to have resulted in cycad cladogenesis in both hemispheres. Tang (2002) hypothesized that modern cycads colonized the Greater Antilles from Mesoamerica during an interval during the late Cretaceous/early Cenozoic when they may have formed a land bridge with the south of Mesoamerica (Pinchell and Kennan, 2009), ground zero for the K–T asteroid impact at about the same time.

Massive extinctions of the Cycadales meant that many areas where several of the genera existed in the past do not currently have any of their living representatives. This might explain why the inferences in the historical biogeographic reconstructions for many deeper nodes of our five gene phylogeny were poorly supported (Fig. 5, Table 4).

Concluding remarks

Our five SCNG phylogeny of Cycadales provides the most congruent estimate of the phylogeny of the order yet presented.
An emerging picture for the biogeography and evolutionary history of Cycadales is supported by both the recent molecular phylogenies and the fossil record. Our results suggest that the current supra-generic classification of the order needs to be revisited. Further research is underway to determine if there are morphological and anatomical features to support some of the clades detected in our study and others with congruent results (Bogler and Francisco-Ortega, 2004; Chaw et al., 2005; Griffith et al., 2012). The extant diversity of cycad species is recent and post-dates the Cretaceous–Paleogene boundary. Our topology, resolved by both concatenation of five loci and three methods of gene tree–species tree reconciliation, suggests massive extinctions prior to the most recent diversification events, and the elimination of certain lineages from entire geographic areas, including close relatives of modern cycads. This hypothesis is supported by the fossil record. Biogeographic reconstructions of cycad ancestral areas are compromised by these extinction episodes and the fact that current distributions do not mirror those from the past.

SUPPLEMENTARY DATA

Supplementary data are available online at www.aob.oxfordjournals.org and consist of the following. Table S1: primer sequences for five single copy genes used across the order Cycadales. Figure S1: trees from individual SCNG locus B&B parsimony analyses of Cycadales.

FUNDING

This work was supported by National Science Foundation Grant DEB 1050340 and by a Christiane Tyson Research Fellowship to D.S.-L. through Montgomery Botanical Center.

ACKNOWLEDGEMENTS

We thank S. Wang, M. Alford and T. Samarakoon for providing a sample and recipe for the PCR adjuvant TBT-PAR 5X. This is a contribution number 261 from the Tropical Biology Program of Florida International University.

LITERATURE CITED


Bremer ED, Stevenson DW, McCombie RW, et al. 2003b. Expressed sequence tag analysis in Cycas, the most primitive living seed plant. Genome Biology 4: R78.


