The LINNEAN SOCIETY of London BOTANICAL Journal of the Linnean Society

Botanical Journal of the Linnean Society, 2014, **174**, 130–140. With 2 figures

The influence of tectonics, sea-level changes and dispersal on migration and diversification of Isonandreae (Sapotaceae)

JAMES E. RICHARDSON[1,2*], AZRUL M. BAKAR[1], JAMES TOSH[1], KATE ARMSTRONG[1], JENNY SMEDMARK[3], ARNE A. ANDERBERG[4], FERRY SLIK[5] and PETER WILKIE[1]

[1]*Royal Botanic Garden Edinburgh, 20a Inverleith Row, Edinburgh EH3 5LR, UK*
[2]*Universidad de Los Andes, Apartado Aéreo 4976, Bogotá, Colombia*
[3]*Bergius Foundation, Department of Botany, Stockholm University, SE-106 91 Stockholm, Sweden*
[4]*Department of Phanerogamic Botany, Swedish Museum of Natural History, PO Box 50007, SE-104 05 Stockholm, Sweden*
[5]*Plant Geography Lab, Xishuangbanna Tropical Botanic Garden, Chinese Academy of Sciences, Menglun, Mengla, 666303 Yunnan, China*

Received 8 June 2013; revised 30 July 2013; accepted for publication 14 August 2013

Internal transcribed spacer (ITS) ribosomal DNA sequence data were generated for 80 of the *c*. 200 species of Isonandreae and were added to data from African and Neotropical representatives in subfamily Sapotoideae and outgroups in Sapotaceae. Bayesian dating and ancestral area reconstruction indicated that Isonandreae are derived from within an African grade. Multiple Australasian species or lineages are derived from Sundanian lineages in South-East Asia with stem ages originating from the late Oligocene. Sri Lankan and Indian lineages are also derived from Sundanian lineages. Our results are consistent with migration from Africa into Sundania followed by numerous over-water dispersal events across Wallace's Line into Australasia and migration from Sundania to the Indian subcontinent. Pleistocene speciation indicates that sea-level changes during that epoch could have been responsible for some species diversification in Sundania. *Botanical Journal of the Linnean Society*, 2014, **174**, 130–140.

ADDITIONAL KEYWORDS: Australasia – dated phylogenetic tree – India – Pleistocene – Sapotoideae – South-East Asia – speciation – Wallace's Line.

INTRODUCTION

Tropical South-East Asia is one of the most species-rich regions on earth and is recognized as a biodiversity hotspot (Myers *et al*., 2000). Reconstructing the biogeographical history of taxa that occupy this region will enable a better understanding of how this biodiversity arose, and dated molecular trees allow us to determine whether evolutionary events coincide with historical tectonic or climatic events in a particular region (e.g. Donoghue & Moore, 2003; Pillon, 2012; Fay & Forest, 2013). A series of tectonic events have influenced diversification in the South-East Asian region. The Indian subcontinent separated from east Africa, Madagascar and Australia during the early Cretaceous at *c*. 120 Ma and collided with Eurasia in the Eocene from *c*. 50 Ma (Briggs, 2003). From the Oligocene (34–23 Ma) onward, the northward movement of the Australian Plate toward the Continental Asian Plate brought the emergent landmasses of these two continental shelves into close proximity (see Hall, 2009, for a review of South-East Asian tectonic history). These two plates converged around a line that was identified by Alfred Russel Wallace (1869) as a major biogeographical boundary between Sundania (largely Laurasian) and Australasia (largely Gondwanan) with a distinct fauna on either side. Plate convergence resulted in the meeting

*Corresponding author. E-mail: j.richardson@rbge.org.uk

of major continental areas, but also had an impact on the emergence/submergence of islands in the archipelago. New Guinea is thought to have emerged as recently as 10 Ma (Hall, 2009), which is remarkable given the high species diversity on this largest of all tropical islands (see Takeuchi, 2005, for comparison of species diversity on different islands in the Malay Archipelago). To add to the complexity, changes in sea level during the Pleistocene (Voris, 2000) have led to successive isolation and reconnection of various island groups in either Sundania or Australasia, although there has never been a direct land connection between these two areas that are separated by a deep oceanic trench through the Makassar Straits. Given each of the above geological and climatic events we can test a series of hypotheses on the origins and patterns of diversification of lineages that occupy the region using dated molecular trees of these groups.

Tribe Isonandreae Hartog are one of five recognized in Sapotaceae by Pennington (1991) and were placed in subfamily Sapotoideae by Swenson & Anderberg (2005). They are distributed from India and Sri Lanka, through Indochina and South-East Asia to Australia and the Pacific Islands. This distribution makes them an ideal model group to study diversification patterns in South-East Asia.

Isonandreae are composed of seven genera (*Palaquium* Blanco, *Aulandra* H.J.Lam, *Isonandra* Wight, *Madhuca* Buch.-Ham. ex J.F.Gmel, *Payena* A.DC., *Burckella* Pierre and *Diploknema* Pierre) and are defined on the basis of a complex combination of floral characters. Harley (1991) surveyed the pollen morphology of Sapotaceae extensively and defined a series of types. Although the fossil pollen record for Sapotaceae is copious, there are no reports of tribe Isonandreae in South-East Asia. The earliest reports of fossil pollen of Sapotaceae are from the Senonian (89–65 Ma) of China (Song, Wang & Fei, 2004), although the affiliation of these fossils with Sapotaceae is doubtful (Harley, 1991). Sapotaceae from the Palaeocene (65–57 Ma) to Early Miocene (*c*. 23 Ma) have been recorded in Sarawak (Muller, 1968), although Harley (1991) was again unconvinced that these are sapotaceous. Pollen from the Late Cretaceous (71 Ma) *Sapotaceoidaepollenites rotundus* that could be related to Mimusopeae (M. M. Harley, pers. comm. 2009) has been found at Otway Basin off the coast of South Australia (Stoian, 2002), and an Australasian origin for Sapotaceae or lineages within the family is thus possible. Pollen grains from the mid Eocene of England (Gruas-Cavagnetto, 1976) are similar to morphological types IIA and VIA described by Harley (1991) that are characteristic of tribes Mimusopeae, Isonandreae and Sideroxyleae. Despite the reasonably extensive fossil pollen record, it is not complete enough and insufficiently well characterized to provide a complete picture of the biogeographical history of the family.

Given the geological history of South-East Asia we can postulate four hypotheses that would receive support through particular molecular tree topologies and node ages. (1) If Isonandreae had a Laurasian or Sundanian origin we would expect Gondwanan lineages (Indian or Australasian) to be derived from lineages currently occupying Laurasian land masses. Australasian lineages would have been increasingly likely to have migrated across Wallace's Line as the Asian and Australian Plates converged. Indian lineages would also be derived from Laurasian or Sundanian lineages if they had migrated into India as the subcontinent collided with Eurasia from the Eocene. (2) If Isonandreae had an Australasian origin, species from that region would form a basal grade from which Sundanian lineages would be derived, having diverged as a result of migration across Wallace's Line from the Miocene. (3) If Isonandreae had an Indian origin we would expect Indian/Sri Lankan species to form a basal grade from which Sundanian and Australasian species would be derived and that clades of Sundanian/Australasian species would diverge from their Indian ancestors when India collided with the Asian continent in the mid-Eocene. (4) If Isonandreae arrived as a result of a long-distance dispersal event from outside Sundania, India or Australasia it would be derived from a lineage that formed a grade of species that occupy the source area. Pleistocene sea-level changes (Voris, 2000) would have resulted in the successive isolation and reconnection of islands. This may have led to allopatric divergence and if so we would expect to see speciation events during the Pleistocene, although of course other events may have contributed to speciation during that time frame.

A Bayesian phylogenetic analysis of 46 nuclear (internal transcribed spacer, ITS) and plastid sequences of Sapotaceae [the *trnH-psbA* spacer, the *trnC-trnD* region (consisting of the *trnC-petN* spacer, the *petN* gene, the *petN-psbM* spacer, the *psbM* gene and the *psbM-trnD* spacer), the *trnC-psbM* region and the 3′ end of *ndhF*) including 20 representatives of Isonandreae (P. Wilkie, J. E. E. Smedmark, A. Anderberg, J. Tosh, K. Armstrong & J. E. Richardson, unpubl. data; Bayesian majority rule consensus tree available as online supplementary material as Supporting Information, Fig. S1) indicated strong support for the monophyly of the tribe. Separate analyses of the nuclear and plastid datasets revealed no examples of hard incongruence between trees (i.e. strongly supported topological differences with jackknife values > 95% or posterior probabilities > 0.99). The rDNA ITS was more variable than plastid genes, so this region was used to produce a phylogenetic tree

with an expanded species sample. This study aims to produce a dated molecular tree and reconstruct node distributions of Isonandreae using the ITS sequence data that will allow testing of all of the above hypotheses so that we may determine patterns in their diversification throughout the Malesian region.

MATERIALS AND METHODS

Data sets and taxon sampling

In total, 138 species of Sapotaceae were included in this study. At least 74 of the *c*. 200 species of Isonandreae were sequenced for ITS. Species of other tribes in subfamily Sapotoideae were included to reconstruct nodes that were used as calibration points to date the tree. As *Sarcosperma* Hook.f. has been shown to be the sister to all other Sapotaceae (Anderberg, Rydin & Källersjö, 2002; Anderberg & Swenson, 2003), *Sarcosperma laurinum* (Benth.) Hook.f. was used to root trees. To recover nodes that could be used to date the phylogenetic tree, sequences of a selection of taxa representing the major lineages in each of tribes Sideroxyleae and Mimusopeae were also utilized. Several species representing subfamily Chrysophylloideae (Anderberg & Swenson, 2003) were also included in the study. A list of accessions with voucher specimen details is provided online (Supporting Information, Table S1).

DNA extraction, amplification and purification

Total genomic DNA was extracted from leaves taken from herbarium specimens or dried in silica gel (Supporting Information, Table S1). About 20–25 mg of leaf tissue was ground using one or two 30-s cycles in a Mixer mill and subsequently treated with 1 mL 2% cetyltrimethylammonium bromide (CTAB) buffer (pre-heated at 65 °C) and 2 μL β-mercaptoethanol, following a modified version of the CTAB DNA protocol outlined by Doyle & Doyle (1990).

Amplifications of the ITS region were performed using the ITS5p/ITS8p primer pair (Möller & Cronk, 2001). PCR was carried out in 25-μL volume reactions containing 5.75 μL sterile distilled water, 0.5 μL genomic DNA, 2.5 μL 10× NH_4 reaction buffer, 0.75 μL each primer, 1.25 μL 50 mM $MgCl_2$, 2.5 μL 10 mM dNTPs, 10 μL 5 M betaine, 0.25 μL bovine serum albumin and 2.25 μL 5 u μL^{-1} Biotaq DNA polymerase buffer. The thermal cycling profile consisted of 5 min denaturation at 95 °C, followed by 35 cycles of 30 s at 95 °C for denaturation, 50 °C for 30 s for annealing and 72 °C for 30 s for extension, with a final extension period of 8 min at 72 °C. Extraction from herbarium specimens sometimes yields low quantities of degraded DNA. In these instances, a nested PCR was performed which involved using 1 μL of the first PCR product added to a second (nested) PCR reaction using the ITS1/ITS4 primer pair (White *et al*., 1990) with the same thermal cycling profile as above. Only four accessions, indicated in Supporting Information, Table S1, were sequenced using the nested PCR method. The PCR products were purified using the ExoSAP IT (GE Healthcare) enzymatic clean-up method following the manufacturer's instructions.

DNA sequencing, assembly and alignment

The purified DNA was used as a template for the sequencing reaction with 'Big Dye' terminators (v.3.1) and the Big Dye PCR profile. Sequencing was performed using 10 μM of the primers ITS5, ITS8, ITS1 and ITS4 in 10-μL reactions. Ten accessions did not produce clean sequences and were excluded from the analysis. Sequences of the complete ITS region (18S, ITS1, 5.8S, ITS2 and 26S) were assembled using Sequencher version 4.7. The sequences were then aligned using ClustalW in BioEdit (Hall, 1999) and exported as a Nexus file for phylogenetic analysis.

Phylogenetic analysis – maximum-parsimony (MP) and Bayesian inference

MP analysis was carried out using PAUP* version 4.0b10 (Swofford, 2002). All informative substitutions were used in the analyses, and indels were treated as missing data. The initial analysis consisted of an heuristic search of 10 000 replicates saving ten trees per replicate, with random stepwise addition, tree bisection-reconnection (TBR) branch swapping, MULTREES on, collapsing branches if maximum branch length is zero, and all characters unordered and equally weighted. All trees saved from this analysis were then used for a further heuristic search, activating steepest descent and MULTREES and saving a maximum of 10 000 trees. This strategy is considered to be sufficient to capture all topological variation (Sanderson & Doyle, 1992). No shorter trees were found in the additional analysis. A jackknife analysis as implemented in PAUP* ver. 4.0b10 (Swofford, 2002) was carried out that consisted of 10 000 replicates deleting 37% of characters, with TBR branch swapping and saving ten trees per replicate.

Bayesian analyses were conducted with MrBayes 3.1 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003), which uses the Markov chain Monte Carlo (MCMC) method (Larget & Simon, 1999) to search tree and parameter space. The Akaike information criterion (AIC) and hierarchical likelihood ratio test (hLRT) were used in MrModelTest (Nylander, 2004) to determine the most appropriate model for the data. Four MCMC chains with chain

temperature set at default were run simultaneously for ten million generations, sampling trees and parameters every 1000 generations. Tracer version 1.5 (Rambaut & Drummond, 2007) was used to determine that a burn-in period of 10% (1000 trees) sufficed and the remaining trees were saved. LogCombiner was used to combine trees in a single file. TreeAnnotator (Drummond & Rambaut, 2007) was used to select the maximum clade credibility (MCC) tree that has the maximum sum of posterior probabilities on its internal nodes and summarizes the node height statistics in the posterior sample. Clade support was represented by posterior probability (PP) values, with PP values of more than 0.95 indicating strong support. The Bayesian MCC tree was used to define calibration points and other key nodes in the dating analysis.

Molecular dating

Type VIA pollen grains (Harley, 1991) characteristic of Sideroxyleae from the Ypresian of England (Gruas-Cavagnetto, 1976) were used as a fossil constraint. We assigned a log normal prior to the stem node of the Sideroxyleae clade [offset: 52.2 Ma (mean: 0.001)]. A mean of 0.001 was chosen so that 95% of the probability is contained in an interval between the mid-point (52.2 Ma) and the upper boundary of the upper Eocene (48.6–55.8 Ma). A log normal prior was chosen to bias in favour of older age estimates as suggested by Ho (2007).

An XML input file was generated in BEAUti version v.1.4.8 (Drummond & Rambaut, 2007). The best performing evolutionary model was identified under two different model selection criteria, the hLRT and the AIC as implemented in MrModelTest (Nylander, 2004). Both selection criteria indicated that a general time reversible (GTR) model with site heterogeneity being gamma distributed and with invariant sites was optimal. A relaxed clock, uncorrelated lognormal model was chosen based on the assumption of the absence of a molecular clock. To specify informative priors for all the parameters in the model, the Yule tree prior was used (recommended as being appropriate for species-level phylogenies, Ho, 2007). The XML file was manually modified to include a starting tree that had been generated in PAUP. Two runs were performed for the dating analysis each with the MCMC chain length set to 10 000 000, to screen every 10 000 and sample every 1000 trees. The XML file was run in BEAST software version v.1.4.8 (Drummond & Rambaut, 2007). LogCombiner was used to combine tree files and to remove the burn-in, which was again determined using Tracer version 1.5 (Rambaut & Drummond, 2007). A burn-in of 10% (1000 trees) sufficed, the remaining trees were saved and the MCC tree was constructed using TreeAnnotator. MCC files were visualized using FigTree version 1.3.1 (Rambaut, 2009).

Reconstruction of ancestral distributions

Two approaches were used to reconstruct ancestral distributions at nodes. Terminals were assigned to the following areas that were chosen based largely on tectonic history (see map inset, Fig. 2): Sundania (including part of mainland Asia to the south of the Isthmus of Kra and islands to the west of Wallace's Line), Asia (north of the Isthmus of Kra), Australasia (including islands to the east of Wallace's Line), Africa (including Madagascar), South Asia (India and Sri Lanka), North America, Hawai'i and the Seychelles.

The first approach for reconstructing ancestral distributions was performed using RASP (Reconstruct Ancestral State in Phylogenies; Yan, Harris & He, 2011, http://mnh.scu.edu.cn/soft/blog/RASP) software, which implements Bayesian binary MCMC (BBM) time-events curve analysis (Yan *et al*., 2011) that allows multiple states to be assigned to terminals. Terminals were assigned states based on the overall distribution of the species, many of which are distributed in multiple areas. BBM suggests possible ancestral ranges at each node and also calculates probabilities of each ancestral range at nodes. One analysis was performed with 5000 000 cycles, ten chains, sampling every 100, with a temperature of 0.1 and with the maximum number of areas set to four. The root node was custom defined as Asian that we believe to be the most likely state for the crown node of the family given that the Asian taxa *Sarcosperma* Hook.f and *Eberhardtia* Lecomte form a grade in which the rest of the family is nested.

A discrete phylogeographical analysis (DPA) using a standard continuous-time Markov chain (CTMC) was also performed to determine node distribution probabilities (Lemey *et al*., 2009). This approach incorporates branch length information and uncertainty in the tree topology. It does not allow terminals to have multiple states, so each sample in this analysis was coded according to the locality of the individual. As recommended by Lemey *et al*. (2009), priors for migration rates used a Γ distribution for the relative rate parameter (shape parameter = 1.0) and an exponential distribution (mean = 1.0) for the geosite model parameter. The XML file for the dating analysis was edited to specify the locality of each accession for the discrete phylogeographical analysis. The analysis required four runs in BEAST of 10 000 000 generations each before effective sample size (ESS) values > 200 for each parameter were achieved. Log files for the separate analyses were

combined in LogCombiner and then imported into Tracer version 1.5 (Rambaut & Drummond, 2007) to check whether ESS values were adequate for each parameter (i.e. > 200). LogCombiner was then used to combine tree files and remove their burn-ins (first 1000 trees for each tree file) that were also determined using Tracer. TreeAnnotator was used to produce the MCC tree. MCC files were visualized using FigTree version 1.3.1 (Rambaut, 2009).

RESULTS

The ITS matrix contained 163 taxa and 1012 bp of sequence data of which 148 were variable but parsimony-uninformative and 407 were potentially parsimony-informative. Ten thousand trees were saved from the parsimony analysis with a length of 2279 steps. The trees from the MrBayes, parsimony and BEAST analyses were compared and there were no examples of hard incongruence. The dates obtained from BEAST analyses with or without ancestral state reconstructions were similar. Only the BEAST trees with ancestral state reconstructions are discussed but the Bayesian MCC tree and parsimony jackknife trees are provided as electronic supplementary information (Supporting Information, Figs S2, S3, respectively).

Figure 1 shows the MCC of the BEAST analysis with ancestral state reconstructions from the BBM analysis and indicates the relationships in Isonandreae and their relationships with other major clades. PP < 0.95 are indicated by dashed lines. Isonandreae are monophyletic with a PP of 1 and the tribe is nested in a grade of predominantly African taxa, most of which were placed in tribe Mimusopeae by Pennington (1991). Figure 2 shows the same MCC tree as in Figure 1, but with reconstructions based on the DPA analysis. PPs for geographical states of nodes for both analyses that are discussed in the text are indicated in Table 1.

Palaquium is not monophyletic with *Aulandra longifolia* H.J.Lam and *Diploknema butyracea* (Roxb.) H.J.Lam nested in it. *Madhuca*, *Isonandra* and *Diploknema* are polyphyletic. *Burckella* and *Payena* are strongly supported as monophyletic (PP 1). There is a strongly supported clade that contains most of the *Madhuca* spp. (PP 1). The 95% confidence intervals of the highest posterior densities (HPDs) at each node are indicated in Supporting Information, Figure S4.

The tree based on a subset of species but a larger number of genes (Supporting Information, Fig. S1) provided additional support for clades recovered in the ITS analysis. For example, the clade containing *Isonandra compta* Dubard, *I.* sp and *I. perakensis* King & Gamble had a PP of 1 in the combined analysis but had a PP < 0.5 in the BEAST analysis of ITS alone. There were no examples of hard incongruence between trees based on ITS compared with those from the combined analysis (Supporting Information, Fig. S1).

The stem age of Isonandreae is 40.5 Ma (HPD: 32.6–48.3 Ma; node D, Figs 1, 2) and the crown is estimated at 36.5 Ma (HPD: 29.0–43.6 Ma; node F, Figs 1, 2). The crown node of the clade that contains an African grade and Isonandreae has a probability of 1 (BBM) or 0.99 (DPA) of being from Africa. The stem node ages for lineages found to the east of Wallace's Line are shown in Table 1 at nodes G, K, N and O with mean ages ranging from 1.3 to 24.2 Ma. These lineages are all nested in a larger lineage composed of accessions that are found west of Wallace's Line. Sri Lankan lineages are also nested in otherwise Sundanian lineages at nodes H, I and L indicated in Table 1 with mean ages of 7.7, 19.8 and 22.4 Ma, respectively. The results indicate that the crown node of Sapotaceae is reconstructed as most probably Asian according to the BBM but most probably African according to the DPA, although the probability for the latter is low. The crown node of Isonandreae is reconstructed as being most probably Sundanian with a probability of being from that area of 0.69 (BBM) or 0.94 (DPA) and the most internal nodes in the tribe also have a higher probability of being from Sundania than from anywhere else: for example, crown of largest *Madhuca* clade (BBM 0.74 Sundania and 0.12 Sundania or Asia; DPA 0.96 Sundania), crown of *Palaquium* clade (BBM 0.56 Sundania and 0.41 Sundania or Asia; DPA 0.98 Sundania) and crown of *Payena* (BBM 0.84 Sundania or Asia; DPA 0.99 Sundania). Twenty-four species in Isonandreae evolved during the Pleistocene, i.e. from 2.58 Ma. That is 32% of the total of 74 sampled in this study.

DISCUSSION

Migration of Isonandreae

According to the BEAST analyses, Sapotaceae began to diversify from 84.5 Ma (HPD: 67.1–105.0 Ma; node A, Figs 1, 2) and the origin of the family is probably Asia as the rest of the family is derived in a grade made up of *Sarcosperma* and *Eberhardtia* that are Asian genera. Bell, Soltis & Soltis (2010) estimated the age of major clades of angiosperms using multiple fossils and found that the stem node of Sapotaceae was 74 (62–84) Ma using BEAST with fossils ages calibrated as exponential priors. In contrast Bremer, Friis & Bremer (2004) estimated a Sapotaceae stem node age of *c.* 102 Ma that is more consistent with the crown node age of the family estimated here. The age of the stem node of the Isonandreae/Mimusopeae clade is Mid Eocene, matching the age of type I and II

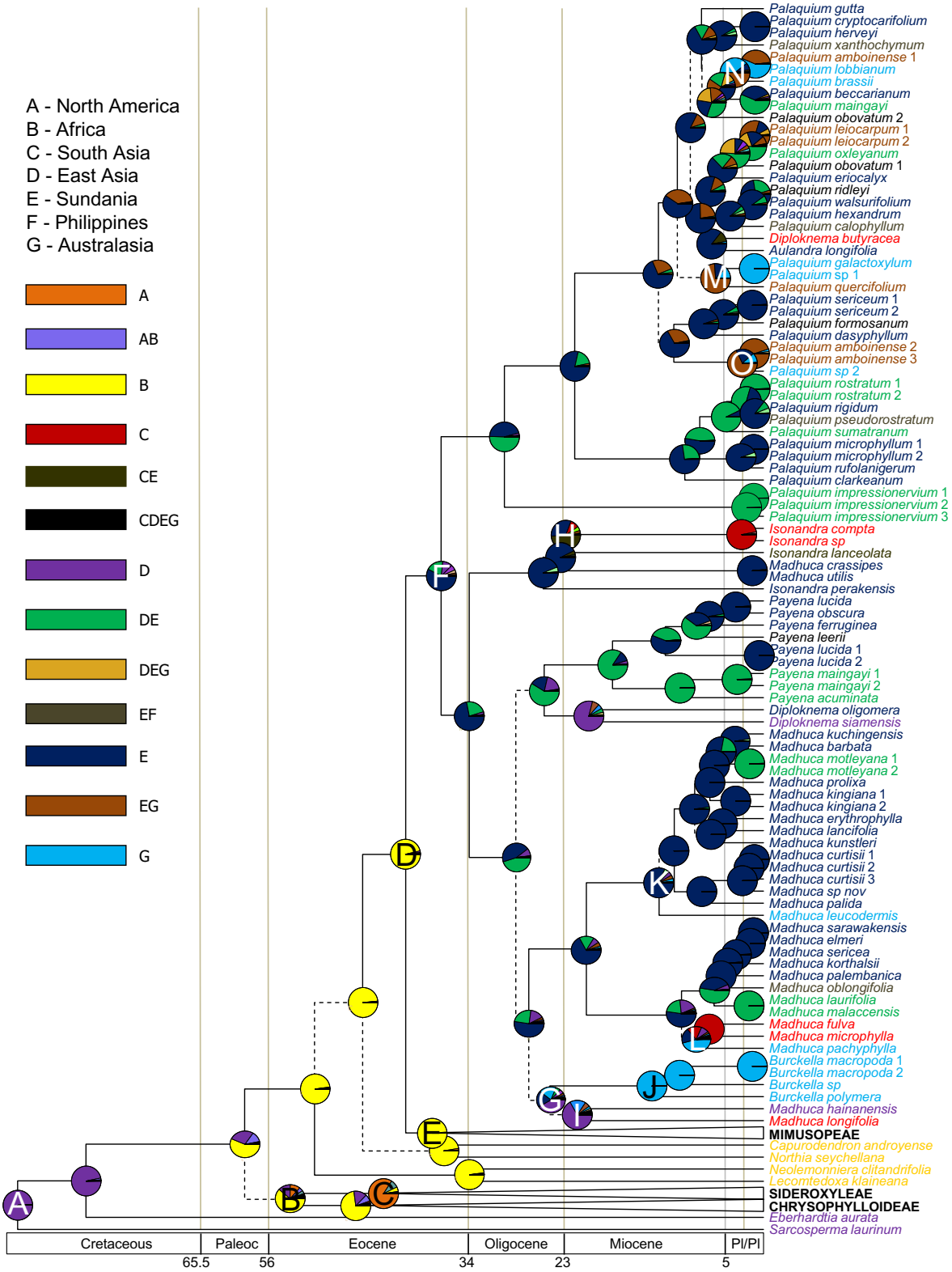

Figure 1. The maximum clade credibility (MCC) tree from the posterior sample generated from the BEAST analysis showing the relationships in Isonandreae and their relationships to other major clades. Calibration nodes and nodes discussed in the text are indicated. Broken lines indicate nodes with PP values < 0.95. Pie charts on nodes indicate ancestral state reconstructions based on the BBM.

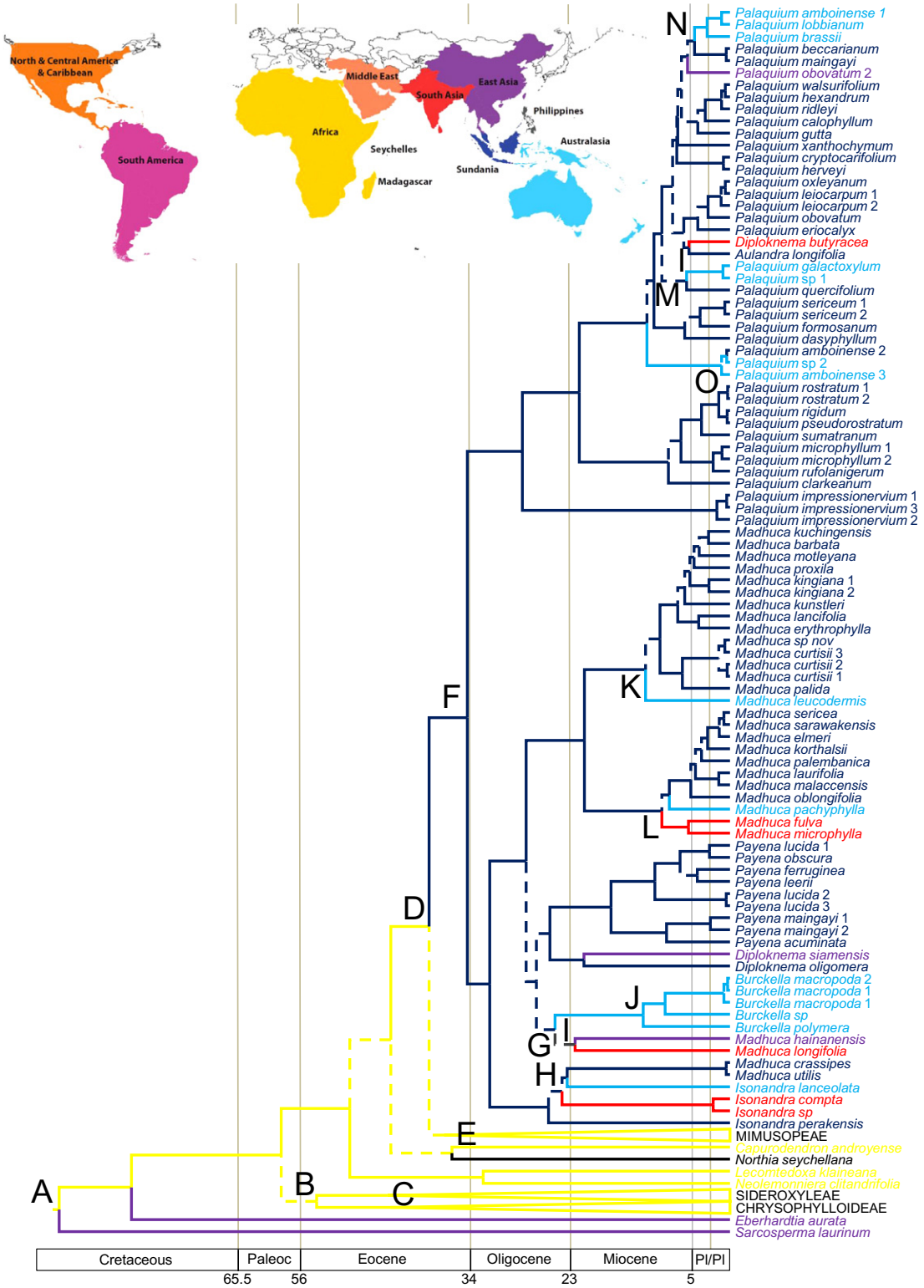

Figure 2. The maximum clade credibility (MCC) tree from the posterior sample generated from the BEAST analysis showing the relationships in Isonandreae and their relationships to other major clades. Calibration nodes and nodes discussed in the text are indicated. Broken lines indicate nodes with PP values < 0.95. Lines are coloured based on the the results of the discrete phylogeographical analysis: purple, Asia; yellow, Africa; light blue, Australasia; red, India and Sri Lanka; dark blue, Sundania.

Table 1. Calibration points and information on other key nodes

Node	Mean node age (Ma)	HPD interval (Ma)	Most probable state BBM/DPA
A. Sapotaceae crown	84.5	67.1–105.0	Asia 0.99/Africa 0.52
B. Sideroxyleae stem	53.6	52.2–56.6	Africa 0.66/0.93
C. Sideroxyleae crown	44.5	38.3–50.3	North America 0.89/0.91
D. Isonandreae stem	40.5	32.6–48.3	Africa 0.96/0.99
E. Mimusopeae crown	38.8	NA	Africa 1.0/0.99
F. Isonandreae crown	36.5	29.0–43.6	Sundania 0.69/0.94
G. *Burckella* stem node, west to east dispersal	24.2	NA	Sundania 0.29/NA
H. *Isonandra compta*, *Isonandra* sp., stem, dispersal/migration to Indian subcontinent	22.4	NA	Sundania or India 0.49/ Sundania 0.95
I. *Diploknema butyracea* MRCA, dispersal/migration to Indian subcontinent	19.8	12.8–26.7	Asia 0.71/Sundania 0.99
J. *Burckella* crown node	12.7	7.7–18.2	Australasia 1.0/ Australasia 0.97
K. *Madhuca leucodermis* MRCA, west to east dispersal	10.1	6.6–13.8	Sundania 0.84/Sundania 0.97
L. *Madhuca fulva*, *Madhuca microphylla*, stem, dispersal/migration to Indian subcontinent	7.7	NA	Australasia 0.53/Sundania 0.92
M. *Palaquium galactoxylum*, *Palaquium* sp. 1 stem, west to east dispersal	5.5	1.5–10.0	Sundania or Australasia 0.70/ Sundania 0.96
N. *Palaquium lobbianum*, *Palaquium brassii* stem, west to east dispersal	4.9	2.9–7.2	Sundania 0.34/Sundania 0.94
O. *Palaquium* sp. 2, west to east to dispersal	1.3	0.1–2.9	Sundania or Australasia 0.63/NA

The mean minimum age estimates and their 95% highest posterior density intervals are indicated. NA, not available; MRCA, most recent common ancestor.

pollen grains (Harley, 1991) characteristic of Mimusopeae and Isonandreae from the Mid Eocene of England (Gruas-Cavagnetto, 1976). The Isonandreae clade is nested in a grade, the early diverging lineages of which are African, consistent with a migration of this lineage from Africa into Asia from the time of the age of the stem node of the tribe at 40.5 Ma (HPD: 32.6–48.3 Ma; node D, Figs 1, 2). This migration could have occurred via Laurasia during this period as temperatures were high enough to support megathermal taxa that are now predominantly restricted to lower latitudes. The historical presence of Sapotaceae in Laurasia is demonstrated by pollen characteristic of Mimusopeae, Isonandreae and Sideroxyleae in the Eocene of England (Gruas-Cavagnetto, 1976; Harley, 1991).

The results are consistent with the hypothesis that Isonandreae began to diversify first in Laurasia or Sundania. Gondwanan (i.e. Indian or Australasian) Isonandreae are nested in a Sundanian lineage supporting the fact that early diversification in the tribe was in Sundania. Migration from Sundania into New Guinea/Australia, across Wallace's Line, occurred on five occasions between *c*. 24.2 and 1.3 Ma. The number of migrations may of course be increased by adding to our limited sampling to the east of Wallace's Line. This west to east migration is also evident in studies of Aglaieae (Meliaceae) by Muellner *et al*. (2008), *Pseuduvaria* Miquel (Annonaceae) by Su & Saunders (2009) and *Begonia* L. (Thomas *et al*., 2012). These results are consistent with previous hypotheses on the origin and patterns of migration of South-East Asian lineages (e.g. Good, 1960; Van Balgooy, 1976; Van Welzen & Slik, 2009), i.e. that migration of angiosperms across Wallace's Line was predominantly from west to east. This has recently been reinforced in a re-assessment of patterns of migration based on a review of phylogenetic studies for the region (Richardson, Costion & Muellner, 2012).

Migration could have occurred at any point between stem and crown nodes with the arrival time only being confirmed at the crown node by diversification. The age of the crown node of *Burckella*, 12.7 Ma (HPD: 7.7–18.2 Ma; node J, Figs 1, 2), indicated that it had definitely reached New Guinea by this time. New Guinea was believed to be submerged until the Mid Miocene (Hall, 2009), so the age of the onset of diversification of *Burckella* is consistent with dispersal onto the island following its emergence. The stem node of *Burckella* is *c*. 24.2 Ma (node G, Figs 1, 2) and migration from Sundania to New Guinea could have occurred at any time from that point onwards. It is possible that this lineage could have migrated to New Guinea via a now submerged island and/or that the

sister lineage of extant *Burckella* may now be extinct. An extinct sister lineage would have decreased the age of the stem node of *Burckella*.

Many west-to-east migrations documented in the literature were accompanied by substantial radiation in the east (e.g. Muellner *et al*., 2008; Thomas *et al*., 2012). Extensive radiation has also occurred in *Palaquium* with approximately half of its 110 species occurring to the east of Wallace's Line. Smaller radiations have occurred in *Madhuca*, a genus of 100 species, which has only 14 species to the east of Wallace's Line and *Payena* with only one of 20 species to the east of the line. *Burckella* has about 14 species, all to the east of Wallace's Line in the Western Pacific from the Moluccas and New Guinea to Fiji, Samoa and Tonga. Our results indicate only one poorly supported example of a back migration from east of Wallace's Line to the west, although additional sampling may of course reveal more examples of this. The absence or low frequency of back-dispersal has been noted in Malesian fanged frogs (*Limnonectes*, Evans *et al*., 2003) and in *Pseuduvaria* (Annonaceae; Su & Saunders, 2009) and was also highlighted by Thomas *et al*. (2012) in his study of *Begonia*, who suggested that it may be due to niche pre-emption (Silvertown, 2004; Silvertown, Francisco-Ortega & Carine, 2005). Prior filling of niche space may inhibit later arrivals by close relatives. Niche space would have been filled in the more ancient terranes of Sundania in comparison with much of the area to the east of Wallace's Line that only emerged recently.

Despite this relative lack of back migration it is clear that Isonandreae have migrated over water across Wallace's Line on numerous occasions. One criticism of using fossil-calibrated dated trees has been that they only give minimum age estimates that may result in under-estimation of the true ages (Heads, 2005). If this were true, then many phylogenetic splits that have been attributed to trans-oceanic long-distance dispersal on the basis of fossil calibrated dated trees (e.g. Bartish *et al*., 2011) could in fact have been older and therefore may have been caused by tectonic events such as the break-up of Gondwanaland. If we are underestimating the ages of splits between Sundania and Australasia, then we could also be underestimating the ability of Sapotaceae to disperse across oceans because these events would have occurred earlier over an even wider expanse of ocean when these two regions were more geographically distant from one another. This is also true of any other tectonic convergence event such as that between the North and South American plates (see Cody *et al*., 2010; J. E. Richardson, G. A. Mondragon, J. Serrano, J. A. Hawkins, I. V. Bartish, U. Swenson, M. Gonzalez, J. Chave, J. Vieu & S. Madriñán, unpubl. data).

Our results show Sri Lankan lineages nested in Sundanian ones, indicating an origin in Sundania and either dispersal or overland migration to Sri Lanka. Migration occurred from Malesia into Sri Lanka (India) three times at *c*. 22.4 Ma (node H, Figs 1, 2), 19.8 Ma (node I, Figs 1, 2) and 7.7 Ma (node L, Figs 1, 2). More sampling, particularly of Indo-Chinese taxa, will be needed to determine whether Sri Lankan lineages arrived overland or by long-distance dispersal. An 'Out of India' scenario was suggested for Crypteroniaceae by Rutschmann *et al*. (2004) with overland migration from India occurring as the subcontinent collided with Laurasia. Based on the fossil record, Morley (1998) suggested an over water invasion of Sundania from India during the Mid Eocene when India and South-East Asia were positioned at similar latitudes and in the same climatic zone. In the Mid and Late Eocene evidence appeared for the presence in South-East Asia of pollen of taxa characteristic of the Palaeocene and Early Eocene of India, such as numerous palms, bombacoid Malvaceae, Sapindaceae, Thymeleaceae, Proteaceae, Linaceae, Olacaceae, Polygalaceae and Ctenolophonaceae (Morley, 1998, 2000; Lelono, 2000; Collinson & Hooker, 2003), suggesting dispersal of Indian elements to South-East Asia in these periods. Our results provide strong evidence to indicate that migration *into* sub-continental India and Sri Lanka from South-East Asia has also occurred during a different period of time than those migrations into Sundania indicated by the fossil record.

Pleistocene speciation

Pleistocene climatic events resulted in changes in land area (and therefore the potential area of occupation of terrestrial organisms) as a consequence of changing sea levels (Voris, 2000). Sea levels were much lower during glacial periods and consequently many present day islands, particularly in Sundania, would have been connected. Inter-glacial periods would have seen contraction of land areas into island 'refugia', and this contraction may have resulted in allopatric speciation. According to the mean ages of splits from a most recent common ancestor, 24 species evolved in Isonandreae during the Pleistocene, indicating that sea-level changes could have had resulted in substantial speciation, although it is, of course, possible that other events might have caused speciation during that epoch. In contrast, only three speciation events occurred in Neotropical Chrysophylloideae (J. E. Richardson, G. A. Mondragon, J. Serrano, J. A. Hawkins, I. V. Bartish, U. Swenson, M. Gonzalez, J. Chave, J. Vieu & S. Madriñán, unpubl. data) during the Pleistocene, although this was based on a sample of only 25% of species compared with the

40% of Isonandreae sampled here. If, on complete sampling, this pattern proves to be the case then it would indicate that for Sapotaceae, South-East Asia has been a more active laboratory of speciation than the Neotropics during the Pleistocene and that, despite Sapotaceae occupying similar biomes in both regions, speciation patterns and processes may differ between them.

ACKNOWLEDGEMENTS

Felix Forest and the DNA bank managers at the Royal Botanic Gardens, Kew, are thanked for supplying DNA aliquots and D. Ndiade Bourobou (CENAREST, Rougier and IRAF) is thanked for the DNA aliquot of *Baillonella toxisperma*. Charlie Danny Heatabun and the staff of the Indonesian Institute for Sciences (LIPI) are thanked for facilitating fieldwork in Indonesia.

REFERENCES

Anderberg AA, Rydin C, Källersjö M. 2002. Phylogenetic relationships in the order Ericales *s.l.*: analyses of molecular data from five genes from the plastid and mitochondrial genomes. *American Journal of Botany* **89:** 677–687.

Anderberg AA, Swenson U. 2003. Evolutionary lineages in Sapotaceae (Ericales): a cladistic analysis based on *ndhF* sequence data. *International Journal of Plant Science* **164:** 763–773.

van Balgooy MMJ. 1976. Phytogeography. In: Paijmans K, ed. *New Guinea vegetation*. Canberra: CSIRO, 1–22.

Bartish IV, Richardson JE, Antonelli A, Swenson U. 2011. Vicariance or long-distance dispersal: historical biogeography of the pantropical subfamily Chrysophylloideae (Sapotaceae). *Journal of Biogeography* **38:** 177–190.

Bell CD, Soltis DE, Soltis PS. 2010. The age and diversification of the angiosperms re-revisited. *American Journal of Botany* **97:** 1296–1303.

Bremer K, Friis EM, Bremer B. 2004. Molecular phylogenetic dating of asterid flowering plants shows early Cretaceous diversification. *Systematic Biology* **53:** 496–505.

Briggs JC. 2003. The biogeography and tectonic history of India. *Journal of Biogeography* **30:** 381–388.

Cody S, Richardson JE, Rull V, Ellis C, Pennington RT. 2010. The great American biotic interchange revisited. *Ecography* **33:** 1–7.

Collinson ME, Hooker JJ. 2003. Paleogene vegetation of Eurasia: framework for mammalian faunas. In Reumer JFW & Wessels W, eds. *Distribution and migration of Tertiary mammals in Eurasia, a volume in honour of Hans De Bruijn*. *Deinsea* **10:** 41–83.

Donoghue MJ, Moore BR. 2003. Toward an integrative historical biogeography. *Integrative and Comparative Biology* **43:** 261–270.

Doyle JJ, Doyle JL. 1990. Isolation of plant DNA from fresh tissue. *Focus* **12:** 13–15.

Drummond AJ, Rambaut A. 2007. BEAST: bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7:** 214.

Evans BJ, Brown RM, McGuire JA, Supriatna J, Andayani N, Diesmos A, Melnick DJ, Cannatella DC. 2003. Phylogenetics of fanged frogs: testing biogeographical hypotheses at the interface of the Asian and Australian faunal zones. *Systematic Biology* **52:** 1–29.

Fay MF, Forest F. 2013. Biogeography – different geographical and taxonomic scales. *Botanical Journal of the Linnean Society* **171:** 301–303.

Good R. 1960. The geographical relationships of the angiosperm flora of New Guinea. *Bulletin of the British Museum (Natural History) Botany* **2:** 203–226.

Govaerts R, Frodin DG, Pennington TD. 2001. *World checklist and bibliography of Sapotaceae*. Kew: Royal Botanic Gardens.

Gruas-Cavagnetto C. 1976. Étude palynologique du Paléogène du sud de l'Angleterre. *Cachiers de Micropaléontologie* **1:** 1–49.

Hall R. 2009. SE Asia's changing palaeogeography. *Blumea* **54:** 148–161.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41:** 95–98.

Harley MM. 1991. The pollen morphology of the Sapotaceae. *Kew Bulletin* **46:** 379–491.

Heads M. 2005. Dating nodes on molecular phylogenies: a critique of molecular biogeography. *Cladistics* **21:** 62–78.

Ho SYW. 2007. Calibrating molecular estimates of substitution rates and divergence times in birds. *Journal of Avian Biology* **38:** 409–414.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: bayesian inference of phylogeny. *Bioinformatics* **17:** 754–755.

Larget B, Simon D. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Molecular Biology and Evolution* **16:** 750–759.

Lelono E. 2000. *Palynological study of the Eocene Nanggulan Formation of Central Java, Indonesia*. London: University of London.

Lemey P, Rambaut A, Drummond AJ, Suchard MA. 2009. Bayesian phylogeography finds its roots. *PLoS Computational Biology* **5:** e1000520.

Möller M, Cronk QCB. 2001. Evolution of morphological novelty: a phylogenetic analysis of growth patterns in *Streptocarpus* (Gesneriaceae). *Evolution* **55:** 918–929.

Morley RJ. 1998. Palynological evidence for Tertiary plant dispersals in the SE Asian region in relation to plate tectonics and climate. In: Hall R, Holloway JD, eds. *Biogeography and geological evolution of SE Asia*. Leiden: Backhuys Publishers, 211–234.

Morley RJ. 2000. *Origin and evolution of tropical rainforests*. Chichester: Wiley.

Muellner AN, Pannell CM, Coleman A, Chase MW. 2008. The origin and evolution of Indomalesian, Australasian and Pacific island biotas: insights from Aglaieae (Meliaceae, Sapindales). *Journal of Biogeography* **35:** 1769–1789.

Muller J. 1968. Palynology of the Pedawan and Plateau Sandstone Formations (Cretaceous–Eocene) in Sarawak. *Micropalaeontology* **14:** 1–37.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J. 2000. Biodiversity hotspots for conservation priorities. *Nature* **403:** 853–858.

Nylander JAA. 2004. *MrModeltest v2. Program distributed by the author*. Uppsala: Evolutionary Biology Centre, Uppsala University.

Pennington TD. 1991. *The genera of Sapotaceae*. Kew: Royal Botanic Gardens.

Pillon Y. 2012. Time and tempo of diversification in the flora of New Caledonia. *Botanical Journal of the Linnean Society* **170:** 288–298.

Rambaut A. 2009. FigTree v1.3.1. Computer program Available at: http://tree.bio.ed.ac.uk/software/figtree/ (last accessed March 2012).

Rambaut A, Drummond AJ. 2007. Tracer v1.4, Available at: http://beast.bio.ed.ac.uk/Tracer

Richardson JE, Costion CM, Muellner AM. 2012. The malesian floristic interchange. In: Gower DJ, Richardson JE, Rosen BR, Rüber L, Williams ST, eds. *Biotic evolution and environmental change in Southeast Asia*. Cambridge: Cambridge University Press, 138–163.

Ronquist F, Huelsenbeck JP. 2003. MRBAYES 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* **19:** 1572–1574.

Rutschmann F, Eriksson T, Schönenberger J, Conti E. 2004. Did Crypteroniaceae really disperse out-of-India? Molecular dating evidence from *rbc*L, *ndh*F and *rpl*16 intron sequences. *International Journal of Plant Science* **165 (**4 Suppl.): 69–83.

Sanderson MJ, Doyle JJ. 1992. Reconstruction of organismal and gene phylogenies from data on multigene families: concerted evolution, homoplasy and confidence. *Systematic Biology* **41:** 4–17.

Silvertown J. 2004. The ghost of competition past in the phylogeny of island endemic plants. *Journal of Ecology* **92:** 168–173.

Silvertown J, Francisco-Ortega J, Carine M. 2005. The monophyly of island radiations: an evaluation of niche pre-emption and some alternative explanations. *Journal of Ecology* **93:** 653–657.

Song ZC, Wang WM, Fei H. 2004. Fossil pollen records of extant angiosperms in China. *The Botanical Review* **70:** 425–458.

Stoian LM. 2002. *Late Cretaceous–Late Eocene palynofloras from drillhole Troas 1, offshore Otway Basin, South Australia*. Report Book 2002/10. Adelaide: Government of South Australia.

Su YCF, Saunders RMK. 2009. Evolutionary divergence times in the Annonaceae: evidence of a late Miocene origin of *Pseuduvaria* in Sundania with subsequent diversification in New Guinea. *BMC Evolutionary Biology* **9:** 153.

Swenson U, Anderberg AA. 2005. Phylogeny, character evolution and classification of Sapotaceae (Ericales). *Cladistics* **21:** 101–130.

Swofford DL. 2002. *PAUP*: phylogenetic analysis using parsimony (* and other methods),* version 4.0b10. Sunderland, MA: Sinauer Associates.

Takeuchi WN. 2005. Introduction to the flora of Papua. In: Marshall AJ, Beehler BM, eds. *The ecology of Papua*. Singapore: Periplus Editions, 269–302.

Thomas DC, Hughes M, Phutthai T, Ardi WH, Rajbhandary S, Rubite R, Twyford AD, Richardson JE. 2012. West to east dispersal and subsequent rapid diversification of the mega-diverse genus *Begonia* (Begoniaceae) in the Malesian archipelago. *Journal of Biogeography* **39:** 98–113.

Voris HK. 2000. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography* **27:** 1153–1167.

Wallace AR. 1869. *The Malay Archipelago*. New York: Harper & Brothers.

van Welzen PC, Slik JWF. 2009. Patterns in species richness and composition of plant families in the Malay Archipelago. *Blumea* **54:** 166–171.

White TJ, Bruns T, Lee S, Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, eds. *PCR protocols: a guide to methods and applications*. London: Academic Press, 315–322.

Yan Y, Harris AJ, He X-J. 2011. RASP (Reconstruct Ancestral State in Phylogenies) 2.1 beta. Available at: http://mnh.scu.edu.cn/soft/blog/RASP

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1. Bayesian majority rule consensus tree of an analysis of a subset of Isonandreae species based on ITS, the *trnH-psbA* spacer, the *trnC-trnD* region (consisting of the *trnCpetN* spacer, the *petN* gene, the *petN-psbM* spacer, the *psbM* gene and the *psbM-trnD* spacer), the *trnC-psbM* region and the 3′ end of *ndhF*). Bayesian support values are given above branches and bootstrap ones below.
Figure S2. Parsimony jackknife consensus tree of ITS data with support values below branches.
Figure S3. MCC tree of the Bayesian analysis of ITS data with support values indicated to the right of nodes.
Figure S4. MCC tree from the BEAST analysis of ITS data with mean and HPDs of node ages.
Table S1. Voucher information and ITS GenBank numbers for accessions sampled. Nomenclature follows Pennington (1991); see Govaerts *et al*. (2001) for synonyms.