Systematics and biogeography of *Elaeocarpus* (Elaeocarpaceae)

Thesis submitted by

Sook-Ngoh PHOON

M.Sc. Universiti Putra Malaysia

in March 2015

for the degree of Doctor of Philosophy

Australian Tropical Herbarium

and the School of Marine and Tropical Biology

James Cook University

Cairns, Australia

STATEMENT ON THE CONTRIBUTION OF OTHERS

The chapters of this thesis are also manuscripts that are in preparation for submission. Several researchers have made contributions to these manuscripts as follows:

Chapter 2: Ferry Slik and Yumiko Baba shared their unpublished *psbA-trnH* sequences of Chinese *Elaeocarpus*, and the *trnL-trnF* region and *trnV-ndhC* sequences of the Australasian and the Pacific islands species, respectively; Mark Coode identified and confirmed the identification of the *Elaeocarpus* vouchers, provided knowledge on the systematics of *Elaeocarpus* and discussions; Darren Crayn provided theoretical background and discussions; John Sugau, Richard Chung, Kamariah Abu Salim, Rismita Sari, Shawn Lum, Nik Faizu Nik Hassan, Singapore Herbarium, Herbarium of the Forest Research Institute Malaysia, Sarawak herbarium, Sandakan Herbarium and Brunei Forestry Herbarium provided assistance during field trips; Missouri Botanic Gardens, Pang Chun Chiu, Rachun Pooma, Anthony van der Ent, Cameron Kilgour, researchers under the Flora of Peninsular Malaysia project donated plant tissue samples; Ashley Field, Andrew Thornhill, Claire Micheneau and Katharina Schulte, assisted in the data analyses and discussions; Philippa Griffin processed the Next-Generation raw data, formatted them for phylogenetic analysis and wrote section 2.2.6; and Gary Wilson made critical comments.

Chapter 3: Andrew Thornhill and Katharina Schulte provided advice on the molecular dating analysis and biogeographical reconstructions, respectively; Darren Crayn provided theoretical background and discussion; Mark Coode provided discussions and shared his unpublished data and knowledge of Elaeocarpaceae; Yumiko Baba provided discussions on Australian *Elaeocarpus*; Choon-Kit Tan and Kim-Wai Tee provided software and computing support; and Gary Wilson and Philippa Griffin provided critical comments.

Chapter 4: Mark Coode contributed part of the data of the Sumatran *Elaeocarpus cupreus* and provided discussions; Darren Crayn provided discussions; Kew Herbarium provided high resolution digital images of the type specimens; Singapore Herbarium, Herbarium of the Forest Research Institute Malaysia, Herbarium Bogoriense, Sarawak herbarium, Sandakan Herbarium and Brunei Forestry Herbarium kindly loaned the specimens of *polystachyus* group under their care; and Paul Gadek and Gary Wilson provided critical comments.

I was financially supported by the Malaysian Australian Alumni Council scholarship. Funding for this study came from the Australian Biological Resources Study Grant No. RFL211-42 to Darren Crayn, Yumiko Baba and myself. The Australian Tropical Herbarium provided significant in-kind and logistical support for this project.

ACKNOWLEDGMENTS

I would like to thank my principal supervisor, Professor Darren Crayn, for his guidance throughout my candidature, his advice and critiques greatly improved the quality of this research and provided a valuable learning experience with regard to addressing questions and answers from various perspectives; my co-supervisor, Professor Paul Gadek (James Cook University) and my collaborators, Dr Kamariah Abu Salim (University of Brunei), Dr Richard Chung (Herbarium of the Forest Research Institute Malaysia) and John Sugau (Sandakan Herbarium) for their support and contributions in the field works of this project; Melissa Harrison for her patience and assistance with molecular procedures, and together with Dr Claire Micheneau for their encouragement and discussions in exploring the Next-Generation Sequencing technologies; Dr Philippa Griffin for her contributions in the Next-Generation Sequencing data analysis and critiques on Chapter 2 and 3; Drs Andrew Thornhill, Ashley Field, Katharina Schulte, Caroline Puente Lelievre and Yumiko Baba for their valuable advice and help in data analyses; the Australian Tropical Herbarium staff (Andrea Lim, Gary Wilson, Frank Zich, Peter Bannink) and postgraduate students (Kaylene Bransgrove, Lalita Simpon, Agustina Arobaya) for their support throughout this study.

I owe my debt of gratitude to Mark Coode, the expert of Elaeocarpaceae, who shared his unpublished data and opinions generously and have been invaluable resources in this study. His enthusiasm for *Elaeocarpus* has brought much pleasure and joy to work on this genus throughout my study.

I would like to thank my family for their encouragement. Last but not least, my husband, Choon-Kit Tan, for his unconditional support, love and cheer, without which I could never have completed this work.

GENERAL ABSTRACT

The genus *Elaeocarpus* (family Elaeocarpaceae) comprises 350 – 400 species most of which are rainforest trees distributed in palaeo-tropical regions (except mainland Africa). The morphology of *Elaeocarpus* has been well documented and has been the basis for the intuitive infrageneric classification systems developed to date. The phylogenetics and evolutionary history of *Elaeocarpus* have received limited attention, however. Phylogenetic studies to date have been based on few markers and species sampling that was very restricted and biased towards Australian taxa. This thesis uses a much-expanded sample size including representatives from various biogeographical regions and a large DNA sequence dataset comprising over 3000 bp from four regions (plastid *psbA-trnH* intergenic spacer, *trnL-trnF* region and *trnV-ndhC* intergenic spacer, and nuclear *Xdh*) to address the following aims: (1) investigate phylogenetic relationships between and within *Elaeocarpus* using both non-parametric (maximum parsimony), and parametric model-based (maximum likelihood and Bayesian Inference) methods, (2) trace the transformation of seed morphological characters (embryo shape and endosperm ornamentation), which are considered as superior in the current infrageneric classifications, on the estimated phylogeny to identify morphological synapomorphies for molecular clades, and (3) estimate the divergence times of lineages, infer the origins and explain current distribution patterns of the genus using an uncorrelated lognormal relaxed molecular clock and five fossil calibration points (four in *Elaeocarpus* and one in *Sloanea*). Extending from the molecular evidence developed here, this study also aimed to delimit taxon boundaries within the *polystachyus* group (i.e. the *polystachyus* clade + *E. polystachyus* or the *Elaeocarpus polystachyus* complex) using morphometrics, a statistically testable and repeatable method and then to compare congruency between the results of the morphometrics and the alpha taxonomy. Finally, this study aimed to provide precursory evidence that morphological similarities within the *polystachyus* group are not correlated with ecological adaptations, and each

taxon does maintained its own morphological characteristics but the observation of these characters might have been limited by the alpha taxonomy method.

The results provide strong support for the monophyly of *Elaeocarpus* and for its sister relationship with *Aceratium*. Within the *Elaeocarpus* clade, *E. holopetalus* is resolved as a distinct lineage that is placed sister to the remainder of the taxa. Apart from *E. holopetalus*, a total of 13 main lineages or clades are resolved: *E. sedentarius*, the *obovatus*, section *Elaeocarpus*, *ganitrus*, group VI, *monocera*, group VII, group XI subgroup B, *acronodia*, *polystachyus*, *coilopetalum*, New Zealand and New Caledonian groups. All of the clades resolved in the present study are broadly congruent with the current infrageneric classifications, except the *obovatus* and the New Zealand clades, which are part of group V subgroup D.

The parsimony reconstructions of the ancestral states of two selected seed morphological characters, embryo shape and endosperm ornamentation indicate that both are homoplasious at higher taxonomic levels (i.e. genus level and above), although the curved embryo is homologous within *Elaeocarpus* (excluding *E. holopetalus*).

The large sample size with many representatives from various biogeographic regions and much-improved resolution of the phylogenetic relationships within *Elaeocarpus* provided a strong foundation to investigate the spatio-temporal evolution of this genus comprehensively for the first time. Elaeocarpaceae and its sister (Cunoniaceae + Cephalotaceae) diverged in the late Cretaceous, and diversification within the family (the crown age) is estimated to have begun at *c*. 83 Mya. Within the family, most of the infrafamilial lineages resolved are congruent with the current infrafamilial groupings (Coode 2004): the *Sloanea* alliance (*Vallea*, *Aristotelia* and *Sloanea*), the Tremandraceous genera (*Platytheca*, *Tetratheca* and *Tremandra*), and the *Elaeocarpus* alliance (*Sericolea*, *Aceratium* and *Elaeocarpus*). The exception is the *Crinodendron* alliance;

Crinodendron and *Peripentadenia* formed a clade but *Dubouzetia* was placed sister to a clade comprising the Tremandraceous genera and the *Elaeocarpus* alliance.

The results of historical biogeographic analysis using Fitch parsimony and Dispersal-Extinction-Cladogenesis methods, and molecular dating analysis using Bayesian relaxed-clock methods suggest that *Elaeocarpus* diverged from its sister – *Aceratium* – in the Eocene in Australia. Early diversification of *Elaeocarpus* in Australia occurred when the continent was still at high latitudes and largely covered with megathermal rainforests. Following this, migration events occurred to the surrounding regions, i.e. New Guinea, Central Malesia, West Malesia, New Zealand and the Pacific islands, and further northwards into continental Asia and Madagascar probably via West Malesia. Several reversal migrations are also postulated. Radiation of *Elaeocarpus* within New Guinea and Borneo, the two current centres of species diversity, may have coincided with mountain building in the Miocene. Geological and climatic changes and zoochorous dispersal mechanisms are hypothesised to have played major roles in shaping the present-day palaeo-tropical distribution patterns of *Elaeocarpus*.

The *polystachyus* group (i.e. the *Elaeocarpus polystachyus* complex) comprises six species (*E. cupreus*, *E. clementis*, *E. integripetalus*, *E. multinervosus*, *E. polyanthus* and *E. polystachyus*) and four varieties (*E. clementis* varieties *clementis*, *borneensis*, *clemensiae* and *kostermansii*). All are endemic to West Malesia and share a unique combination of morphological character states, including numerous, unawned stamens that are densely arranged in multiple tiers. Most members of this informal infrageneric group form a clade in the phylogenetic analysis, except *E. polystachyus*; while no DNA samples were available for *E. clementis* var. *kostermansii*, *E. integripetalus* and *E. polyanthus*. The group is morphologically well defined and phylogenetically broadly supported, but taxon boundaries within it are unclear. The results of the morphometric analysis and the alpha taxonomy are broadly congruent in

supporting six species, but the infraspecific taxa appear to be unsupported. Additionally, the hypothesis proposed in this study where morphological similarities within the *polystachyus* group are not correlated with ecological adaptations appeared to be supported by the morphometric evidence. This suggests that the morphological differences observed are predominantly genetically, rather than environmentally, controlled.

TABLE OF CONTENTS

List of Figures

posterior probability. Genus abbreviations are *Aceratium* (Ac), *Crinodendron* (C), *Dubouzetia* (D), *Elaeocarpus* (E), *Peripentadenia* (Pe) and *Sericolea* (Se). * Branch collapses in the MP or ML 50 % majority rule consensus tree. The insert shows branch lengths proportional to the amount of nucleotide change and the section of the tree which has been enlarged.

Fig. 2.6 Maximum clade credibility tree from Bayesian analysis of the combined plastid regions *psbA-trnH*, *trnL-trnF* and *trnV-ndhC*, and the low-copy protein coding nuclear gene xanthine dehydrogenase (*Xdh*). Branch support values are to the left of the nodes in the following order: MP bootstrap/ ML bootstrap/ Bayesian posterior probability. Genus abbreviations are *Aceratium* (Ac), *Crinodendron* (C), *Dubouzetia* (D), *Elaeocarpus* (E), *Peripentadenia* (Pe) and *Sericolea* (Se). * Branch collapses in the MP or ML 50 % majority rule consensus tree. The insert shows branch lengths proportional to the amount of nucleotide change and the section of the tree which has been enlarged.

Fig. 2.7 Transformation of the shape of embryo reconstructed using Fitch parsimony and traced on the Bayesian phylogenetic tree of *Elaeocarpus* based on the combined plastid regions *psbA-trnH*, *trnL-trnF* and *trnV-ndhC*, and the low-copy protein coding nuclear gene – xanthine dehydrogenase (*Xdh*). White branch represents lineage(s) with straight embryo; black branch represents lineage(s) with curved embryo.

Fig. 2.8 Transformation of the ornamentation of endosperm reconstructed using Fitch parsimony and traced on the Bayesian phylogenetic tree of *Elaeocarpus* based on the combined plastid regions *psbA-trnH*, *trnL-trnF* and *trnV-ndhC*, and the low-copy protein coding nuclear gene – xanthine dehydrogenase (*Xdh*). White branch represents lineage(s) with entire endosperm; black branch represents lineage(s) with ruminate endosperm.

Fig. 3.1 Approximate distribution of Elaeocarpaceae. A, *Aristotelia* alliance. B, *Sloanea* alliance. C, *Crinodendron* alliance. D, *Elaeocarpus* alliance. E,

Tremandraceous genera (infrafamilial groupings and distribution follow Coode 2004).

Fig. 3.2 Eight main biogeographical areas representing the current distribution of species of *Elaeocarpus*, *Aceratium* and *Sericolea* defined for the ancestral area reconstruction. A, Madagascar; B, Continental Asia, including the northern part of the Sunda shelf from the Thailand-Peninsular Malaysia border northwards, and westwards to India and Sri Lanka; C, West Malesia including the southern part of the Sunda shelf from the Thailand-Peninsular Malaysia border southwards; D, Central Malesia (or Wallacea); E, New Guinea and the surrounding islands (e.g. New Britain); F, Australia; G, New Zealand; and H, the Pacific including New Caledonia.

Fig. 3.3 Chronogram of Elaeocarpaceae: maximum clade credibility tree from the BEAST analysis. Posterior estimates of divergence times were inferred using partitioned analyses based on four combined DNA regions, a UCLN model, and six calibration points, i.e. one secondary (labelled C1, below the node) and five macrofossils (four *Elaeocarpus* and one *Sloanea*) (labelled C2 – C6, below the nodes). Nodes are posterior mean ages (Mya), with blue node bars representing the 95 % HPD intervals (see Table 3.3 for details). Bayesian PP < 0.95 are indicated by asterisks above branches. All infrafamilial groups of Coode (2004) retrieved, i.e. *Aristotelia* alliance, *Sloanea* alliance, *Elaeocarpus* alliance and Tremandraceous genera, except the *Crinodendron* alliance, where *Dubouzetia*, *Peripentadenia* and *Crinodendron* did not formed a clade. Genus abbreviations are *Aristotelia* (Ar), *Crinodendron* (C), *Dubouzetia* (D), *Peripentadenia* (Pe), *Platytheca* (Pl), *Tetratheca* (Te), *Tremandra* (Tr), *Sloanea* (Sl) and *Vallea* (V). DNA extraction numbers were given to the right of taxon names. Geological time scale abbreviations: Cretaceous (CRE); Paleocene (PAL); Eocene (EOC); Oligocene (OLI); Miocene (MIO); Pliocene (PL); Pleistocene (P); Holocene (H). Insert showing the location of the enlarged section of the phylogenetic tree.

Fig. 3.4 Ancestral area reconstruction for *Elaeocarpus*. Letters on the left of the nodes represent Fitch parsimony results, and letters on the right are Dispersal-

Extinction-Cladogenesis results (see Materials and methods). For the latter, all possible ancestral range subdivision/inheritance scenarios are presented in Table 3.3. Letters to the left of the taxa names represent current distribution (as defined in Fig 3.2).

Fig. 4.1 Distribution of the *polystachyus* group: *E. clementis* (var. *borneensis* ▲, var. *clemensiae* ▼, var. *clementis* △ and var. *kostermansii* ▽), *E. cupreus* ❋, *E. integripetalus* ✚, *E. multinervosus* ○, *E. polyanthus* ◆ and *E. polystachyus* ◇. A, Distribution of the nine taxa in West Malesia; B, Distribution in Malay Peninsula; C, Distribution of in West Borneo; D, Distribution of in East Borneo; E, Distribution in Southeast Borneo; F, Distribution in North-Central Borneo.

Fig. 4.2 Flower sexuality in the *polystachyus* group. A, Typical bisexual flowers of *Elaeocarpus polystachyus*; B, Male flowers of *E. multinervosus* (photo: Coode); C, Female flowers of *E. multinervosus*.

Fig. 4.3 Phytogeographical areas at the floristic province and subprovince levels in the distribution range of the *polystachyus* group (modified from Wong 1998, with permission from Institute of Botany, Academia Sinica Monograph Series).

Fig. 4.4 Minimum spanning tree (MST) and Semi-strong Hybrid Multidimensional Scaling on the two-dimensional plane (axis 1 and 3) viewed through three-dimensional ordination space showing relationship of the nine taxa within the *polystachyus* group based on the combined vegetative and floral variables: *E. clementis* (var. *borneensis* ▲, var. *clemensiae* ▼, var. *clementis* △ and var. *kostermansii* ▽), *E. cupreus* ❋, *E. integripetalus* ＋, *E. multinervosus* ○, *E. polyanthus* ◆ and *E. polystachyus* ◇. A, Combined vegetative and floral dataset. Vectors of the most informative variables ($r^2 > 0.6$) as determined by PCC are shown in the direction of correlation. Variables are: TwigWidth (variable 1), PetioleLength (2), PetioleWidth (3), DistalPulvinusWidth (4), LeafLength (5), LeafWidth (6), LeafbaseWidestpoint (8), AbaxialMidribWidth (12), SecondaryVeins (13),

InflorescenceWidth (17), SepalWidth (21), PetalWidth (23), SepalNo (27), PetalNo (28), FlowerMery (29).

Fig. 4.5 Dendrograms of the nine taxa within the *polystachyus* group produced from the Agglomerative Hierarchical Clustering with Flexible UPGMA technique and distance estimated with Gower metric association measure based on the combined vegetative and floral dataset. OTUs represent the following taxa: *E. clementis* (var. *borneensis* ▲, var. *clemensiae* ▼, var. *clementis* △ and var. *kostermansii* ▽), *E. cupreus* ❋, *E. integripetalus* ┼, *E. multinervosus* ○, *E. polyanthus* ◆ and *E. polystachyus* ◇. Blue dotted line indicates grouping cut off point.

Fig. 4.6 Minimum spanning tree (MST) and Semi-strong Hybrid Multidimensional Scaling on the two-dimensional plane (axis 1 and 3) viewed through three-dimensional ordination space showing relationship of the nine taxa within the *polystachyus* group based on the combined vegetative and fruit variables: *E. clementis* (var. *borneensis* ▲, var. *clemensiae* ▼, var. *clementis* △ and var. *kostermansii* ▽), *E. cupreus* ❋, *E. integripetalus* ┼, *E. multinervosus* ○, *E. polyanthus* ◆ and *E. polystachyus* ◇. A, Vegetative + floral dataset. Vectors of the most informative variables (r^2 > 0.6) as determined by PCC are shown in the direction of correlation. Variables are: TwigWidth (variable 1), PetioleLength (2), PetioleWidth (3), DistalPulvinusWidth (4), LeafLength (5), LeafWidth (6), LeafbaseWidestpoint (8), AbaxialMidribWidth (12), SecondaryVeins (13) and InfructescencesWidth (36).

Fig. 4.7 Dendrograms of the nine taxa within the *polystachyus* group produced from the Agglomerative Hierarchical Clustering with Flexible UPGMA technique and distance estimated with Gower metric association measure based on the combined vegetative and fruit dataset. OTUs represent the following taxa: *E. clementis* (var. *borneensis* ▲, var. *clemensiae* ▼, var. *clementis* △ and var. *kostermansii* ▽), *E. cupreus* ❋, *E. integripetalus* ┼, *E. multinervosus* ○, *E. polyanthus* ◆ and *E. polystachyus* ◇. Blue dotted line indicates grouping cut off point.

Fig. 4.8. Minimum spanning tree (MST) and Semi-strong Hybrid Multidimensional Scaling on the two-dimensional plane viewed through three-dimensional ordination space showing correlation between morphological variation and floristic provinces: *E. clementis* in Borneo (var. *borneensis* △, var. *clemensiae* ▽, var. *clementis* ▲ and var. *kostermansii* ▼), *E. cupreus* in Sumatra ⨳ , *E. cupreus* in Malay Peninsula ◪ , *E. cupreus* in Borneo ✳, *E. integripetalus* in Sumatra ┼, *E. multinervosus* in Borneo ○, *E. polyanthus* in Borneo ◆ and *E. polystachyus* in Malay Peninsula ◇. A, Combined vegetative and floral dataset (axis 1 and 2). Vectors of the most informative variables (r^2 > 0.6) as determined by PCC are shown in the direction of correlation. Variables are: TwigWidth (variable 1), PetioleLength (2), PetioleWidth (3), DistalPulvinusWidth (4), LeafLength (5), LeafWidth (6), LeafbaseWidestpoint (8), AbaxialMidribWidth (12), SecondaryVeins (13), InflorescenceWidth (17), PetalWidth (23), SepalNo (27), PetalNo (28) and FlowerMery (29); B, Combined vegetative and fruit dataset (axis 2 and 3). Vectors of the most informative variables (r^2 > 0.6) as determined by PCC are shown in the direction of correlation. Variables are: TwigWidth (variable 1), PetioleLength (2), PetioleWidth (3), DistalPulvinusWidth (4), LeafLength (5), LeafWidth (6), LeafbaseWidestpoint (8), AbaxialMidribWidth (12), AbaxialSecondaryVeinsWidth (15) and InfructescenceWidth (36).

Fig. 4.9 Minimum spanning tree (MST) and Semi-strong Hybrid Multidimensional Scaling on the two-dimensional plane viewed through three-dimensional ordination space showing correlation between morphological variation and floristic subprovinces in Sumatra and Malay Peninsula: *E. cupreus* in Sumatra (S) ×, *E. cupreus* in Riau Pocket (RP) in vegetative + floral dataset ⨳or Lowland Flora in Malay Peninsula in vegetative + fruit dataset ✳, *E. integripetalus* in S ┼, *E. polystachyus* in Mountain Flora in Malay Peninsula ☒ , *E. polystachyus* in Lowland Flora in Malay Peninsula ◇, *E. polystachyus* in RP ◰ and *E. polystachyus* in Seasonal Asiatic Intrusion ⊟ . A, Combined vegetative and floral dataset (axis 1 and 2). Vectors of the most informative variables (r^2 > 0.6) as determined by PCC are shown in the direction of correlation. Variables are: PetioleWidth (variable 3), DistalPulvinusWidth (4), LeafLength (5), LeafWidth (6), LeafbaseWidestpoint (8), SecondaryVeins (13), SecondaryVeinAngle (14), PedicelLength (19), SepalWidth (21), PetalWidth (23),

SepalNo (27), PetalNo (28), FlowerMery (29), OvaryWidth (32) and FlowerSexuality (35); B, Combined vegetative and fruit dataset (axis 1 and 2). Vectors of the most informative variables (r^2 > 0.6) as determined by PCC are shown in the direction of correlation. Variables are: (variable 1), PetioleLength (2), PetioleWidth (3), DistalPulvinusWidth (4), LeafLength (variable 5), LeafWidth (6), LeafRatio (7), LeafbaseWidestpoint (8), AbaxialMidribWidth (12), SecondaryVeins (13), AbaxialSecondaryVeinsWidth (15), InfructescencesWidth (36) and FruitStalkWidth (38).

Fig. 4.10 Minimum spanning tree (MST) and Semi-strong Hybrid Multidimensional Scaling in the two-dimensional plane viewed through three-dimensional ordination space showing correlation between morphological variation and floristic subprovinces in Borneo: *E. clementis* in Lowland Flora in Borneo (B) ▲, *E. clementis* in East Coast Sabah Floristic Subprovince (ECSS) △, *E. clementis* in *Kerangas* Forest (K) ▼, *E. clementis* in Mountain Flora in Borneo (MFB) ▽, *E. clementis* in Seasonally Flooded Riverside (R) ▶, *E. clementis* in Riau Pocket (RP) ▷, *E. clementis* in Seasonal Asiatic-Australasian Intrusion in North Borneo (SAAI) ◥, *E. cupreus* in B ✳, *E. cupreus* in ECSS ×, *E. cupreus* in RP ⊠, *E. cupreus* typical form in B ⊞, *E. cupreus* typical form in RP ◪, *E. multinervosus* in RP ✴, *E. multinervosus* in B ○, *E. multinervosus* in K ●, *E. multinervosus* in SAAI ⊗ and *E. polyanthus* in RP ◆. A, Combined vegetative and floral dataset (axis 1 and 2). Vectors of the most informative variables (r^2 > 0.6) as determined by PCC are shown in the direction of correlation. Variables are: TwigWidth (variable 1), PetioleLength (2), PetioleWidth (3), DistalPulvinusWidth (4), LeafLength (5), LeafWidth (6), LeafbaseWidestpoint (8), AbaxialMidribWidth (12), SecondaryVeins (13) and InflorescenceWidth (17); B, Combined vegetative and fruit dataset (axis 1 and 3). Vectors of the most informative variables (r^2 > 0.6) as determined by PCC are shown in the direction of correlation. Variables are: TwigWidth (variable 1), PetioleLength (2), PetioleWidth (3), DistalPulvinusWidth (4), LeafLength (5), LeafWidth (6), LeafbaseWidestpoint (8), AbaxialMidribWidth (12), SecondaryVeins (13), AbaxialSecondaryVeinsWidth (15), InfructescenceWidth (36) and FruitStalkWidth (38).

List of Tables

Table 4.1 Comparison of diagnostic morphological characters within the *polystachyus* group complex showing continuous and overlapping character states and unclear taxon boundaries. *Elaeocarpus cupreus* (EC), *E. clementis* var. *borneensis* (ECb), *E. clementis* var. *clemensiae* (ECs), *E. clementis* var. *clementis* (ECc), *E. clementis* var. *kostermansii* (ECk), *E. integripetalus* (EI), *E. multinervosus* (EM), *E. polyanthus* (EP) and *E. polystachyus* (ES) (Regular text = information from Coode 1996c and unpublished data; italic text = observations made in this study).

Table 4.2 Definition of variables scored for morphometric analysis and values of the evaluation analyses (i.e. Kruskal-Wallis (p-value), Principal Component Correlation (r^2) and Monte-Carlo Attributes in Ordination (MCAO)) for taxon boundaries, floristic provinces and subprovinces are given. Values for variable 1 – 35 were extracted from the combined vegetative and floral dataset and variable 36 – 40 from the combined vegetative and fruit dataset. * Most informative variables.

List of Appendices

Appendix 1.1 Comparison of diagnostic morphological characters within the *polystachyus* group complex showing continuous and overlapped character states and unclear taxon boundaries.

Appendix 2.1 Details of samples used in the phylogenetic and molecular dating analyses. Provided for each sample is the taxon, authority, infrageneric assignment (according to Baba and Crayn 2012, Coode 1978, 1984, 1996a – d, 1998, 2001c, e, f, 2010, pers. comm. 2014, Coode and Weibel 1994 and Weibel 1968), voucher details and herbarium where deposited, DNA sample number and institution where extracted (G and CNS: extraction done at CNS; NSW: extraction done at NSW and donated for this study). + New sequence generated in this study; FS Unpublished sequences donated by Ferry Slik (pers. comm. 2011); NA Information not available or sequence not generated. GenBank accession numbers are provided for sequences obtained from that source. AK (Auckland War Memorial Museum), BRI (Queensland Herbarium), BRUN (Brunei Forestry Centre), CM (Carnegie Museum of Natural History Herbarium), CNS (Australian Tropical Herbarium), K (Royal Botanic Gardens, Kew), KEP (Forest Research Institute Malaysia, Kepong), KYO (Graduate School of Science, Kyoto University), MEL (Royal Botanic Gardens, Melbourne), MO (Missouri Botanical Garden), NOU (Institut de Recherche pour le Développement), NSW (National Herbarium of New South Wales), PERTH (Western Australian Herbarium), PTBG (National Tropical Botanical Garden, Kalaheo), SAN (Forest Research Centre, Sandakan), SAR (Department of Forestry, Kuching), UCBG (University of California Botanical Garden at Berkeley).

Appendix 2.2 Morphological character states scored for parsimony reconstruction. Coding of characters for mapping onto the molecular phylogenetic trees was as follows: straight = 0; curved = 1; entire = 0; ruminate = 1. CNS (Australian Tropical Herbarium), K (Royal Botanic Gardens, Kew), NA infrageneric grouping not applicable.

Appendix 2.3 Maximum clade credibility tree from Bayesian analysis of the *psbA-trnH* intergenic spacer. DNA extraction numbers or GenBank numbers are presented after taxon names. Branch support values are to the left of the nodes in the following order: MP bootstrap/ ML bootstrap/ Bayesian posterior probability. Genus abbreviations are *Aceratium* (Ac), *Crinodendron* (C), *Dubouzetia* (D), *Elaeocarpus* (E), *Peripentadenia* (Pe) and *Sericolea* (Se). * Branch collapses in the MP or ML 50 % majority rule consensus tree. The insert shows branch lengths proportional to the amount of nucleotide change and the section of the tree which has been enlarged.

Appendix 2.4 Maximum clade credibility tree from Bayesian analysis of the *trnL-trnF* region. DNA extraction numbers or GenBank numbers are presented after the taxon names. Branch support values are to the left of the nodes in the following order: MP bootstrap/ ML bootstrap/ Bayesian posterior probability. Genus abbreviations are *Aceratium* (Ac), *Crinodendron* (C), *Dubouzetia* (D), *Elaeocarpus* (E), *Peripentadenia* (Pe) and *Sericolea* (Se). * Branch collapses in the MP or ML 50 % majority rule consensus tree. The insert shows branch lengths proportional to the amount of nucleotide change and the section of the tree which has been enlarged.

Appendix 2.5 Maximum clade credibility tree from Bayesian analysis of the *trnV-ndhC* region. DNA extraction numbers or GenBank numbers are presented after the taxon names. Branch support values are to the left of the nodes in the following order: MP bootstrap/ ML bootstrap/ Bayesian posterior probability. Genus abbreviations are *Aceratium* (Ac), *Crinodendron* (C), *Dubouzetia* (D), *Elaeocarpus* (E), *Peripentadenia* (Pe) and *Sericolea* (Se). * Branch collapses in the MP or ML 50 % majority rule consensus tree. The insert shows branch lengths proportional to the amount of nucleotide change and the section of the tree which has been enlarged.

Appendix 2.6 Maximum clade credibility tree from Bayesian analysis of the *trnV-ndhC* region. DNA extraction numbers or GenBank numbers are presented after the taxon names. Branch support values are to the left of the nodes in the following order: MP bootstrap/ ML bootstrap/ Bayesian posterior probability. Genus abbreviations are *Aceratium* (Ac), *Crinodendron* (C), *Dubouzetia* (D), *Elaeocarpus* (E),

Peripentadenia (Pe) and *Sericolea* (Se). * Branch collapses in the MP or ML 50 % majority rule consensus tree. The insert shows branch lengths proportional to the amount of nucleotide change and the section of the tree which has been enlarged.

Appendix 3.1 Species distribution data.

Appendix 4.1 Herbarium specimens examined for morphometric analysis. Species: EC (*Elaeocarpus cupreus*), ECb (*Elaeocarpus clementis* var. *borneensis*), ECs (*Elaeocarpus clementis* var. *clemensiae*), ECc (*Elaeocarpus clementis* var. *clementis*), ECk (*Elaeocarpus clementis* var. *kostermansii*), EI (*E. integripetalus*), EM (*Elaeocarpus multinervosus*), EP (*Elaeocarpus polyanthus*) and ES (*Elaeocarpus polystachyus*); Herbaria: BO (Herbarium Bogoriense, Bogor), BRUN (Brunei Forestry Centre), K (Royal Botanic Gardens, Kew), KEP (Forest Research Institute Malaysia, Kepong), L (Nationaal Herbarium Nederland, Leiden University Branch, Leiden), SAN (Forest Research Centre, Sandakan), SAR (Department of Forestry, Kuching), SING (Singapore Herbarium); Characters scored: [a]specimen scored for vegetative and floral characters, [b]specimen scored for vegetative and fruit characters; [c]specimen scored for vegetative characters only; Floristic provinces: Borneo (B), Malay Peninsula (MP), Sumatra (S); Floristic subprovinces: Lowland Flora in Borneo (B), East Coast Sabah Floristic Subprovince (ECSS), *Kerangas* Forest (K), Mountain Flora in Borneo (MFB), Mountain Flora in Malay Peninsula (MFMP), Lowland Flora in Malay Peninsula (MP), Seasonally Flooded Riverside (R), Riau Pocket (RP), Sumatra (S), Seasonal Asiatic-Australasian Intrusion in North Borneo (SAAI), Seasonal Asiatic Intrusion (SAI); NA information not available.

List of Supplementary Figures

Supplementary Figure 4.1 Minimum spanning tree (MST) and Semi-strong Hybrid Multidimensional Scaling based on the combined vegetative and floral dataset on the three-dimensional ordination space showing relationship of the nine taxa within the *Polystachyus* group: *E. clementis* (var. *borneensis* ▲, var. *clemensiae* ▼, var. *clementis* △ and var. *kostermansii* ▽), *E. cupreus* ❋, *E. integripetalus* ┼, *E. multinervosus* ○, *E. polyanthus* ◆ and *E. polystachyus* ◇.

Supplementary Figure 4.2 Minimum spanning tree (MST) and Semi-strong Hybrid Multidimensional Scaling based on the combined vegetative and fruit dataset on the three-dimensional ordination space showing relationship of the nine taxa within the *Polystachyus* group: *E. clementis* (var. *borneensis* ▲, var. *clemensiae* ▼, var. *clementis* △ and var. *kostermansii* ▽), *E. cupreus* ❋, *E. integripetalus* ┼, *E. multinervosus* ○, *E. polyanthus* ◆ and *E. polystachyus* ◇.

Supplementary Figure 4.3 Minimum spanning tree (MST) and Semi-strong Hybrid Multidimensional Scaling based on the combined vegetative and floral dataset on the three-dimensional ordination space showing correlation between morphological variation and floristic provinces: *E. clementis* in Borneo (var. *borneensis* △, var. *clemensiae* ▽, var. *clementis* ▲ and var. *kostermansii* ▼), *E. cupreus* in Sumatra ⨳, *E. cupreus* in Malay Peninsula ◪ , *E. cupreus* in Borneo ❋, *E. integripetalus* in Sumatra ┼, *E. multinervosus* in Borneo ○, *E. polyanthus* in Borneo ◆ and *E. polystachyus* in Malay Peninsula ◇.

Supplementary Figure 4.4 Minimum spanning tree (MST) and Semi-strong Hybrid Multidimensional Scaling based on the combined vegetative and fruit dataset on the three-dimensional ordination space showing correlation between morphological variation and floristic provinces: *E. clementis* in Borneo (var. *borneensis* △, var. *clemensiae* ▽, var. *clementis* ▲ and var. *kostermansii* ▼), *E. cupreus* in Sumatra ⨳, *E. cupreus* in Malay Peninsula ◪, *E. cupreus* in Borneo ❋, *E. multinervosus* in Borneo ○, *E. polyanthus* in Borneo ◆ and *E. polystachyus* in Malay Peninsula ◇.

Supplementary Figure 4.5 Minimum spanning tree (MST) and Semi-strong Hybrid Multidimensional Scaling based on the combined vegetative and floral dataset on the three-dimensional ordination space showing correlation between morphological variation and floristic subprovinces in Sumatra and Malay Peninsula: *E. cupreus* in Sumatra (S) ×, *E. cupreus* in Riau Pocket (RP) ✱, *E. integripetalus* in S ┼, *E. polystachyus* in Mountain Flora in Malay Peninsula ☒, *E. polystachyus* in Lowland Flora in Malay Peninsula ◇, *E. polystachyus* in RP ◪ and *E. polystachyus* in Seasonal Asiatic Intrusion ⊟.

Supplementary Figure 4.6 Minimum spanning tree (MST) and Semi-strong Hybrid Multidimensional Scaling based on the combined vegetative and fruit dataset on the three-dimensional ordination space showing correlation between morphological variation and floristic subprovinces in Sumatra and Malay Peninsula: *E. cupreus* Lowland Flora in Malay Peninsula ✱, *E. polystachyus* in Mountain Flora in Malay Peninsula ☒, *E. polystachyus* in Lowland Flora in Malay Peninsula ◇, *E. polystachyus* in Riau Pocket ◪ and *E. polystachyus* in Seasonal Asiatic Intrusion ⊟.

Supplementary Figure 4.7 Minimum spanning tree (MST) and Semi-strong Hybrid Multidimensional Scaling based on the combined vegetative and floral dataset on the three-dimensional ordination space showing correlation between morphological variation and floristic subprovinces in Borneo: *E. clementis* in Lowland Flora in Borneo (B) ▲, *E. clementis* in East Coast Sabah Floristic Subprovince (ECSS) △, *E. clementis* in *Kerangas* Forest (K) ▼, *E. clementis* in Mountain Flora in Borneo (MFB) ▽, *E. clementis* in Seasonally Flooded Riverside (R) ▶, *E. clementis* in Riau Pocket (RP) ▷, *E. clementis* in Seasonal Asiatic-Australasian Intrusion in North Borneo (SAAI) ◀, *E. cupreus* in B ✱, *E. cupreus* in ECSS ×, *E. cupreus* in RP ☒, *E. cupreus* typical form in B ⊞, *E. cupreus* typical form in RP ◪, *E. multinervosus* in RP ✶, *E. multinervosus* in B ○, *E. multinervosus* in K ●, *E. multinervosus* in SAAI ⍟ and *E. polyanthus* in RP ◆.

Supplementary Figure 4.8 Minimum spanning tree (MST) and Semi-strong Hybrid Multidimensional Scaling based on the combined vegetative and fruit dataset on the

three-dimensional ordination space showing correlation between morphological variation and floristic subprovinces in Borneo: *E. clementis* in Lowland Flora in Borneo (B) ▲, *E. clementis* in East Coast Sabah Floristic Subprovince (ECSS) △, *E. clementis* in *Kerangas* Forest (K) ▼, *E. clementis* in Mountain Flora in Borneo (MFB) ▽, *E. clementis* in Riau Pocket (RP) ▷, *E. clementis* in Seasonal Asiatic-Australasian Intrusion in North Borneo (SAAI) ◀ , *E. cupreus* in B ❋, *E. cupreus* in ECSS ×, *E. multinervosus* in RP ⨳ , *E. multinervosus* in B ○, *E. multinervosus* in K ●, *E. multinervosus* in SAAI ⨳ and *E. polyanthus* in RP ◆.

Chapter 1: General Introduction

The main focus of this thesis is to improve the understanding of the systematics and evolutionary history of the genus *Elaeocarpus* L. Despite being the largest genus in the family Elaeocarpaceae, studies to date have been limited in scope or in the range of methods and data utilised. Therefore, the objectives of this thesis are hierarchically three-fold: firstly, to develop a phylogenetic framework using molecular data to investigate the monophyly of *Elaeocarpus* and its relationships with *Aceratium* DC. and *Sericolea* Schltr.; secondly, to infer the evolutionary history of *Elaeocarpus* in time and space to develop and test hypotheses regarding its diversity and current distribution patterns; and lastly, to resolve groups and their relationships within the West Malesian *polystachyus* group (i.e. the *Elaeocarpus polystachyus* Wall. ex Müll.Berol. complex) to test the current taxonomy. Three different methods are explored to achieve these objectives: molecular phylogenetics, molecular dating and historical biogeographical analysis, and morphometric analysis. Molecular evidence is used to estimate the phylogeny, evolutionary divergence times and historical biogeography of the genus, while morphological evidence is used to resolve morphological groups within the *polystachyus* group.

1.1 Approaches in systematics

Similarity among organisms is of two kinds: homology (similarity due to inheritance from a common ancestor) and homoplasy (convergence). Homology is useful for reconstructing the pattern of ancestor-descendent, or phylogenetic, relationships among organisms (a key objective of systematics) whereas convergence may be misleading.

Phenetics and phylogenetics are the two main philosophical approaches used in systematics (Simpson 2006). Both seek to discover 'natural groups'. The phenetic approach uses overall similarity to infer the membership of groups but does not distinguish between homology and homoplasy – all variation is considered potentially useful for discovering natural groups. The phylogenetic approach differs from

phenetics in that it ignores homoplasy, and further distinguishes two types of homology. Homologous character states that are retained unchanged from a common ancestor (pleisomorphies) are not considered informative about evolutionary relationships. Derived character states or evolutionary novelties (apomorphies) are considered informative about evolutionary relationships and therefore are used to group taxa in the phylogenetic approach.

Evolutionary relationships between groups are illustrated as phylogenetic trees (Simpson 2006), which can then serve as a basis for formulating and testing evolutionary hypotheses in various fields of comparative biology. However, any phylogenetic tree is simply an estimate of the one true phylogeny, and it can be challenging to determine its accuracy.

Two major analytical approaches are commonly used to infer phylogenetic trees: distance-, and character-based methods. Distance-based methods, such as Neighbour Joining and the Unweighted Pair Group Method with Arithmetic Mean (UPGMA), calculate a phylogenetic tree based on a matrix of pairwise evolutionary distances among the entities (operational taxonomic units or OTUs). Because the characters are converted to distances and not interpreted directly information on the distribution of character states and the relationship between individual characters and the tree is lost. Distance methods are computationally fast.

Character-based methods utilise the character data directly. They may be of two types, non-parametric and parametric. The most commonly used non-parametric method is Maximum Parsimony, which seeks the most parsimonious tree topology (i.e. the one reflecting the least evolutionary change) among the universe of possible trees. The most parsimonious tree is preferred as it involves the fewest *ad hoc* assumptions about the evolutionary process (Fitch 1971; Felsenstein 2004; Simpson 2006). However, parsimony algorithms may underestimate the amount of evolutionary change on the tree leading to an inaccurate estimate of the true phylogeny.

Parametric methods use a specific model of nucleotide substitution in the phylogeny estimation and generate a phylogram, i.e. a tree with branch lengths scaled to the amount of inferred genetic change (Bromham 2008). The models used to represent the nucleotide evolution process can range from a simple model one-parameter model (e.g. Jukes-Cantor model) to a complex model where all possible nucleotide substitutions are separately parameterised (e.g. General time reversible model) (Strimmer 1997). It is important to choose the model that is the best fit to the particular dataset being analysed, and this choice is often made using the likelihood ratio test. Maximum Likelihood and Bayesian Inference are among the most popular parametric methods for phylogeny estimation, although they differ markedly many ways.

Maximum Likelihood calculates the likelihood that a given tree produced the observed data matrix given the model of evolution used (Felsenstein 2004). The best-fit model should account for differences in the rate of evolution among characters and failure to determine this model may produce an inaccurate estimate of the phylogeny (Bromham 2008). The confidence of the results is usually assessed using bootstrapping (Felsenstein 1985).

Bayesian analysis uses the likelihood function and a model of nucleotide substitution but differs from ML in that the Markov chain Monte Carlo (MCMC) simulation is used to sample trees from the set of credible trees in proportion to their true posterior probability distribution. Thus Bayesian Inference estimates the probability that the hypothesis is true rather than just the likelihood of the data given the tree and model (Larget and Simon 1999; Gamerman and Lopes 2006). Several Markov chains are often run and tree sampling is initiated when the runs have reached convergence (Ronquist *et al*. 2011). Unfortunately, sometimes one or more chains may be trapped in local optima and the sampled trees will not correctly approximate the posterior density (Altekar *et al*. 2004). The metropolis-coupled MCMC is used as a strategy to improve mixing of Markov chains when multiple local peaks in the posterior density are present (Altekar *et al*. 2004). Assessing convergence of the multiple runs is a common drawback in the Bayesian analysis, as the currently

available tools often provide only a rough estimation (Wilgenbusch *et al*. 2004). Since all methods have their own advantages and disadvantages, it is important to analyse the data with different methods and compare congruencies of the results of each method according to the research objectives and the nature of the data.

Assessing confidence in the inferred phylogenetic tree is an essential part of most phylogenetic studies. Bayesian methods assess confidence inherently and report it as posterior probabilities of nodes on the tree. For other methods however, random resampling procedures such as the bootstrap and the jackknife are commonly used. The bootstrap samples with replacement, building a fictional dataset of the same size as the original (Felsenstein 1985, 2004), whereas the jackknife deletes a proportion of the original dataset (Farris *et al*. 1996; Felsenstein 2004). These datasets are then subject to phylogenetic analysis then the process is repeated many times. Confidence in particular nodes is reported as the bootstrap or jackknife proportion, which is the frequency of that node in the collection of strict consensus trees produced by analysis of the bootstrapped or jackknifed datasets.

1.1.1 Applications of molecular data in phylogenetic systematics

Since the late 1950s, rapid development of technologies in the molecular sciences has driven the application of molecular data to systematics (Graur and Li 2000). Molecular data has several advantages in the study of phylogenetics compared to morphological data. Firstly, DNA is strictly heritable, generally has limited environmental influences (Hillis 1987), and characters take discrete states (i.e. A, C, G, T, gap) allowing simpler homology assessment of molecular traits compared with morphological traits. Therefore the ancestor-descendant relationships among organisms can be reconstructed more accurately, even when they are distantly related (Graur and Li 2000). Finally, molecular characters are more numerous within organisms than morphological ones, providing a vast resource base for phylogenetic estimation (Hillis 1987).

Plant DNA is found in three different genomes, mitochondrial, plastid and nuclear; the latter two are most commonly utilised for molecular phylogenetic studies in angiosperms. Plastid DNA is often favoured over nuclear because its simpler genetic structure (haploid and non-recombinant) facilitates data interpretation (Small *et al*. 2004). However, plastid DNA is uniparentally inherited, usually maternally, and thus reflects the evolutionary history of one parental lineage only which is a disadvantage for estimating the true phylogenetic tree, particularly for complicated lineages resulting from incomplete lineage sorting or hybridisation (Small *et al*. 2004; Lihová *et al*. 2006). Moreover, a phylogenetic tree based on any single molecular marker is simply a gene tree and it may or may not accurately estimate the true species tree.

Nuclear DNA (nDNA) is potentially a more informative data source for phylogenetic inference because it is biparentally inherited, giving both of the paternal and maternal evolutionary histories. Being independent of plastid DNA it also provides an independent estimate of the phylogeny (Small *et al*. 2004). Additionally, plastid DNA is relatively slowly evolving and may fail to resolve relationships in some groups - studies of Hodges and Arnold (1994), Gielly *et al*. (1996), Sang *et al*. (1997) and Whitten *et al*. (2000) on various angiosperms found that analyses of sequence data from the nuclear ribosomal DNA (nrDNA) provided improved resolution of clades that showed low plastid DNA gene sequence divergence.

Despite the advantages of nDNA or nrDNA, available markers from this genome that can be used on a broad taxonomic scale without problems are limited. nDNA (including nrDNA) is often subject to the presence of paralogous gene copies, secondary structure (including compensatory mutation), pseudogenes and homoplasy from incomplete concerted evolution. These phenomena potentially violate the assumptions of neutrality and independence of characters (Feliner and Rossello 2007) and may mislead the estimation of phylogeny (Sanderson and Doyle 1992).

Paralogous genes owe their similarity to gene duplication events rather than speciation. Amplifying a mixture of templates using Sanger sequencing can be challenging because when more than one copy is amplified, this sequencing technique

will result in sequence reading difficulties (Small *et al*. 2004; Griffin *et al*. 2011). The most common approach to overcome the problem of paralogy is by bacterial cloning to segregate the different gene copies and amplify only a single amplicon from each colony, which can then be sequenced (Archibald *et al*. 1999). However, this technique is time and cost inefficient often leading to fewer individuals and genes being investigated than the recommended 20 – 40 clones per individual (Small *et al*. 2004; Griffin *et al*. 2011).

Next-generation sequencing (NGS) approaches, i.e. high-throughput sequencing technologies, have the capability to interpret a mixture of templates (Church 2006; Hall 2007). These different copies can then be segregated bioinformatically. There are various NGS methods and platforms, e.g. single-molecule real-time sequencing (PacBio RS II by Pacific Biosciences of California Inc. 2013), ion semiconductor (Ion Torrent™ sequencing by Life Technologies Corporation 2014), pyrosequencing (454 Sequencing by Roche Diagnostics Corporation 2014), sequencing by synthesis (illumina by Illumina Inc. 2013) and sequencing by ligation (SOLiD® Sequencing Chemistry by Life Technologies Corporation 2014). The suitability of the method or platform depends upon the aims, time and budget of the research project. Griffin *et al*. (2011) demonstrated that the pyrosequencing platform is more time and cost efficient for retrieving orthologous gene sequences compared to cloning.

1.1.2 Applications of molecular phylogenies in evolutionary biology

Molecular dating incorporates DNA sequences and palaeo-evidence (such as fossil, palaeogeography and palaeoclimate) to estimate the time of divergence of two lineages from their most recent common ancestor (MRCA) (i.e. speciation) (Weir and Schluter 2008). Development of this field was initiated by the molecular clock hypothesis proposed by Zuckerkandl and Pauling (1962), which suggests that if sequence evolution is constant over time, then the amount of genetic difference between any two taxa is proportional to the time since they last shared the MRCA. However, the strict molecular clock hypothesis of Zuckerkandl and Pauling (1962) does not accommodate variation of the molecular evolutionary rate across organisms as

shown by other researchers (Wu *et al*. 1985; Herbert 1996; Kumar and Subramanian 2002; Waterston *et al*. 2002; Yi *et al*. 2002). Hence, the local or relaxed molecular clock was introduced by Drummond *et al.* (2006) to 'relax' the assumption of a strictly constant rate by allowing the molecular evolutionary rate to vary among lineages.

Calibration of the molecular clock is the most crucial step in estimating evolutionary timescales, regardless of whether a strict or relaxed clock is used. Accurate calibration can be done by determining the absolute timescale of relevant evolutionary events, such as the estimated age of fossils that can be used to determine the divergence time between two lineages or a known historical geological event (the formation of mountain range or breaking up of continents that drove the speciation process) (Ho 2008). Once the molecular clock is calibrated, the evolutionary rate can then be used by extrapolation to estimate the timing of evolutionary events within other lineages for which no direct fossil evidence is available. Weir and Schluter (2008) showed that the evolutionary rate is highly variable among different bird species - most lineages evolve at an average rate of 1 % per million years, but rates among lineages varied four-fold. Therefore, the evolutionary rate determined for one lineage should be very cautiously applied to another even when the two are relatively closely related. It is advisable to use a relaxed clock, recalibrated, to accommodate any variation in the rate.

Divergence times in the Elaeocarpaceae were estimated in previous studies (Wikström *et al*. 2001; Crayn *et al*. 2006) using the program r8s (Sanderson 2003). This program incorporates both Non-Parametric Rate Smoothing (NPRS) and Penalised Likelihood (PL) methods to estimate molecular evolutionary rates and divergence times on a given phylogenetic tree. NPRS estimates ages through a smoothing criterion (Sanderson 1997), whereas PL is a semiparametric approach (Sanderson 2002) that uses both parametric and non-parametric methods in the divergence time estimation. On the other hand, a recently developed Bayesian evolutionary analytical program, Bayesian Evolutionary Analysis Sampling Trees (BEAST) (Drummond and Rambaut 2008), uses strict or relaxed clock models. The strict clock model uses a lognormal distribution with a mean and variance selected from other distributions

(e.g. gamma), but often only one evolutionary rate is applied to the entire phylogenetic tree (Brown and Yang 2011). The relaxed clock molecular model (also known as local clock approach) estimates each of the branch rates either independently or together with its ancestral branch rate (Sanderson 1997; Kishino *et al*. 2001). BEAST also simultaneously estimates phylogeny, node ages and substitution rates, and these features allow uncorrelated substitution rates among branches on the phylogenetic tree as well as flexibility on the rate at each branch (Drummond and Rambaut 2008). Therefore, BEAST has significant advantages over r8s and is likely to more accurately estimate divergence times.

Phylogenetic reconstructions provide opportunities for the study of evolutionary events including historical biogeographical events. Biogeography is the study of the distributions of organisms and is usually divided into two traditions: ecological biogeography and historical biogeography (Crisci *et al*. 2003). Ecological biogeography investigates ecological factors or processes that are acting in the present time (Cox and Moore 1993), whereas historical biogeography aims to deduce evolutionary processes, such as speciation and extinction, or changes in distribution patterns of species across geographical areas, over an evolutionary time scale (Baldwin and Sanderson 1998; Crisci *et al*. 2003; Albert and Reis 2011).

Historical biogeographic patterns can be explained by three major processes: vicariance, dispersal and extinction. Vicariance is a process by which a population of organisms is split into disjunct populations due to the formation of physical barrier, such as results from continental drift, glaciation or orogeny and leads to various forms of speciation (Nelson and Ladiges 1996; Crisci *et al*. 2003; Leiberman 2005). Dispersal is the movement of organisms to a new area, either through the disappearance of geographic barriers allowing genetic interchange of those previously isolated organisms (Leiberman 2005) or through stochastic transport of seeds or propagules from their point of origin to different habitats via dispersal agent (Albert and Reis 2011). On the other hand, extinction is the disappearance of a lineage, which is irreversible (Albert and Reis 2011). Each of these historical biogeographical hypotheses can be tested using different approaches, i.e. dispersalism, phylogenetic

biogeography, panbiogeography, cladistic biogeography and parsimony analysis of endemicity, either independently or integrated as a single approach, depending on the aims and the nature of the data (Morrone and Crisci 1995).

1.1.3 Applications of morphometrics in resolving species complexes

Morphology can provide valuable information for systematic studies provided that an appropriate approach is selected to avoid ambiguous interpretation of the character states. Morphometrics is a method that can provide a repeatable and statistically reliable framework for resolving species complexes (for examples see Blackith and Reyment 1971; Marcus 1990; Reyment 1991). This method assumes that variables are independent, but non-independent variables can used if they are transformed into ratios. The data are analysed using various kinds of cluster analysis (Sneath and Sokal 1973) or multivariate approaches such as principal component analysis (Glassburn 1995). Additionally, geographical or environmental factors may be employed as proxies to understand correlations between morphological variation and ecological differences observed (Möller *et al*. 2007). There are two common approaches in morphometrics: traditional morphometrics and geometric morphometrics. The former captures landmarks (measurement points) as distances (length or width) or angles. The latter captures geometrical relationships between landmarks, i.e. pairs of measurements made from a common landmark or a set of three measurements corresponding to a triangle of landmarks (Rohlf and Marcus 1993).

1.2 Diversity, biology and systematics of *Elaeocarpus*

The Elaeocarpaceae is a medium sized angiosperm family comprising 12 genera (*Aceratium*, *Aristotelia* L'Hér., *Crinodendron* Molina, *Dubouzetia* Brongn. & Gris., *Elaeocarpus*, *Peripentadenia* L.S.Sm., *Platytheca* Steetz, *Sericolea*, *Sloanea* L., *Tetratheca* Sm., *Tremandra* R.Br. and *Vallea* L.) and *c*. 550 species (Coode 2004). The family is widely distributed in the neo- and palaeo-tropics, and in the subtropics and temperate regions of the southern hemisphere, but is absent from mainland Africa

(Coode 2004). Molecular phylogenetic studies have consistently resolved the family as monophyletic if including the formerly recognised family Tremandraceae (e.g. Savolainen *et al*. 2000; Crayn *et al*. 2006; Magallón *et al.* 2015), and the most recent and comprehensive studies suggest that its sister group comprises Cephalotaceae and Cunoniaceae. The former is monotypic, herbaceous, carnivorous (pitcher-bearing) and endemic to swamps in southwestern Australia, and the latter is a family of mostly southern hemisphere trees and shrubs (Heibl and Renner 2012; Magallón *et al.* 2015).

Elaeocarpus is the most species-rich genus in the family, comprising 350 – 400 species (inclusive of new, currently undescribed species) (Coode pers. comm. 2014) that inhabit tropical and subtropical forests, ranging from lowland to montane areas of Madagascar, Asia, Australia and the Pacific islands (Coode 2004). New Guinea, with *c*. 85 species, and Borneo, with *c*. 70 species, are the major centres of species diversity, followed by the Philippines (*c*. 50), Sulawesi (*c*. 40), Peninsular Malaysia (*c*. 30), Sri Lanka and India (*c*. 30), New Caledonia (29), Australia (*c*. 27), Fiji (22), China (*c*. 20), Thailand (*c*. 18), Java (*c*. 15), Madagascar (8), Japan (2), New Zealand (2), Lord Howe Island (1) and Hawai'i (1) (Coode 2004, pers. comm. 2014; Maynard *et al.* 2008; Baba and Crayn 2012).

Most *Elaeocarpus* species are evergreen trees or shrubs, although a few species can occur as epiphytes or lianes, and some are briefly deciduous. Architecturally, many species exhibit *Terminalia*-like branching (i.e. sympodial), with the branches arising from the main axis at irregular intervals and bearing simple leaves in an alternate, spiral or distichous phyllotaxy (Fig. 1.1A). The inflorescences are usually racemose (Fig. 1.1B) and the flowers symmetrical, bisexual (Fig. 1.1C) or unisexual (Fig. 1.1D). The petals are sometimes sepaloid with entire (Fig. 1.1C & D), lobed or fimbriate margins (Fig. 1.1E). The disk (Fig. 1.1E) is variously developed taking a range of shapes and sizes. The number of stamens ranges between 5 – 200; the filaments are straight to sigmoid; the anthers are sometimes with setae or when a connective is present, it is often extended as an awn or rarely a hook. The ovary is 2 – 9-locular, containing 2 – 16 ovules per loculus. The drupes are usually iridescent blue

Fig. 1.1 Morphology of *Elaeocarpus*. A, *Elaeocarpus mastersii* King showing *Terminalia*-like branching of *Elaeocarpus*; B, *Elaeocarpus stipularis* Blume showing the racemose inflorescence; C, Bisexual flowers of *E. polystachyus* Wall. ex Müll. Berol., with entire petals; D, Female flowers of *E. multinervosus* R.Knuth, where the sterile stamens are much reduced in size; E, Senescing flowers of *E. robustus* Roxb. showing the fimbriate petals and the yellow disk surrounding the base of ovary; F, Iridescent blue fruits of *E. angustifolius* Blume; G, Straight embryo and entire endosperm of *E. stipularis*; H, Ruminate endosperm of *E. mastersii*.

(Fig. 1.1F), due at least in some species, to surface diffraction instead of pigmentation (Lee 1991). The seeds are enclosed in a hard, often sculptured, stone that is formed from the inner mesocarp (Dettman and Clifford 2000). Embryos are of two main types: straight (Fig. 1.1G) with broad cotyledons or curved with broad or narrow cotyledons. Some curved embryos have ruminate endosperm (Fig. 1.1H).

Seed dispersal of some *Elaeocarpus* is known to be facilitated by animals (zoochory). Frugivorous birds, such as pigeons, cassowaries, parrots, Kiwi and the possibly extinct wattled crow Huia (*Heteralocha acutirostris* Cabanis), are known to disperse various species of this genus in India (Aggarwal 2002), Australia, New Guinea (Crome 1975, 1976), Admiralty Islands and New Zealand (Ridley 1990). The fruits of some *Elaeocarpus* are eaten by various mammals, such as rodents, bats, pigs, civet cats, macaques and Sloth bear (*Melursus ursinus* Shaw, a native in South India and Sri Lanka) in Asia (Ridley 1990; Aggarwal 2002), Malesia (Ridley 1990; personal observations 2011, 2012) and Australasia (Rossetto *et al*. 2008; R. Kooyman pers. comm. 2014). *Elaeocarpus* are commonly found near to rivers or streams, but dispersal via water (hydrochory) has not been recorded.

Elaeocarpus is particularly diverse with high endemism in West Malesia, which is a phytogeographic area located on the southern tip of the Sunda Shelf. West Malesia comprises Sumatra, part of Southern Thailand, Peninsular Malaysia, Singapore, Java, Borneo and Palawan Island in the Philippines (Van Steenis 1949; Van Welzen and Slik 2009) (Fig. 1.2). About one third of all *Elaeocarpus* species occur there, at least 80 % of them endemic, while species complexes are also common (Coode and Weibel 1994; Coode 1996a – c, 1998, 2001d, e, 2010). An example is the *polystachyus* group or the *E. polystachyus* complex comprising six morphologically similar species: *E. polystachyus* in the Malay Peninsula, *E. integripetalus* Miq. in Sumatra, *E. cupreus* Merr. in the Malay Peninsula and Borneo, and *E. clementis* Merr., *E. multinervosus* R.Knuth and *E. polyanthus* Ridl. in Borneo (Coode 1996c). The taxon boundaries between these six species are not well defined; their diagnostic characters (listed in Appendix 1.1) show apparently continuous variation or overlap within the group. Repeatable and statistically reliable approaches, such as morphometrics and

molecular phylogenetics, may be more useful to resolve the taxon boundaries within this complex.

Infrageneric classifications of *Elaeocarpus* have been established for the species in China, Malesia, Papuasia, Australasia and the Pacific islands (Table 1.1), mainly following the concepts of Brongniart and Gris (1861), Schlechter (1916), Weibel (1968), Coode (1978, 1996a – c, 2001a, b, d – f, 2003, 2010) and Coode and Weibel (1994). However, the names of the infrageneric groups are not standardised. Hence, some groups are named (Brongniart and Gris 1861; Smith 1944, 1945; Weibel 1968; Coode and Weibel 1994; Coode 1995, 1996a – c, 1998, 2001a – f, 2002, 2003, 2005, 2007a – b, 2010), some numbered (Coode 1984) and some are both named and numbered (Schlechter 1916; Coode 1978, 1980) (Table 1.1), albeit the names and numbers are synonymous.

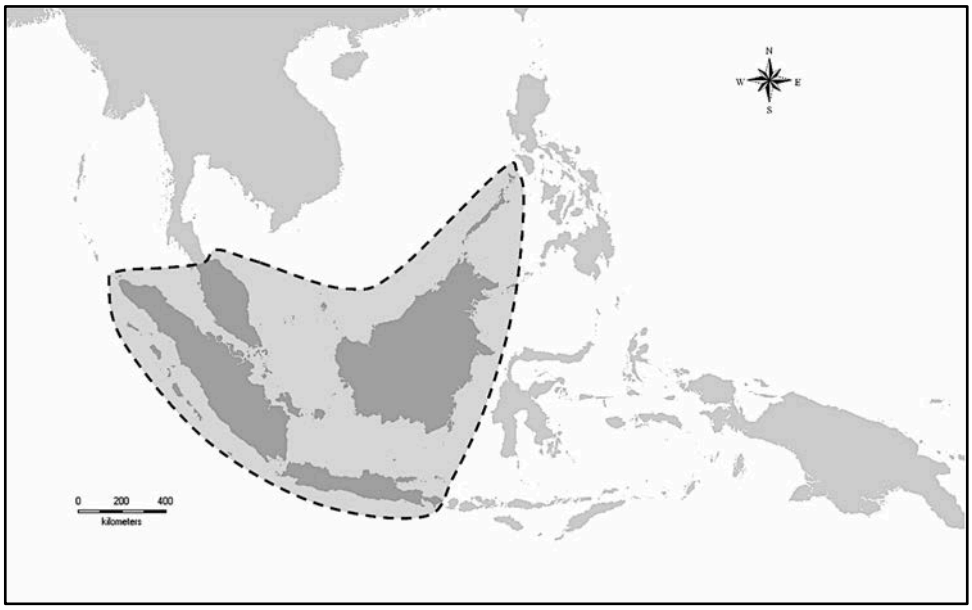


Fig. 1.2 West Malesia, a phytogeographic area comprised of Sumatra, part of Southern Thailand, Peninsular Malaysia, Singapore, Java, Borneo and Palawan Island in the Philippines.

Weibel (1968) was the first to use seed morphological features, i.e. embryo shape (straight or curved) and endosperm condition (broad or narrow, entire or ruminate), as taxonomic markers in *Elaeocarpus*. Subsequently Coode (1978, 1996a –

c, 2001a, b, d – f, 2003, 2010) and Coode and Weibel (1994) adopted these characters in various Flora treatments. They proposed significant changes in the infrageneric classifications of *Elaeocarpus*, where all of the previously established groups except Group I (= *Lobopetalum* Schltr.) and Group II (= *Dactylosphaera* Schltr.) of Brongniart and Gris (1861) were re-defined, giving a total of 15 infrageneric groups and 27 subgroups (Table 1.1). These infrageneric classifications are based on an intuitive approach and have not been examined using a phylogenetic approach to test the monophyly of the groups and to determine the morphological synapomorphies that support them. Coode (1985, 1987) considered the homology of morphological traits and polarity of character states and used his results to reconstruct a preliminary phylogenetic tree of Elaeocarpaceae genera, with a focus on *Aristotelia*, *Crinodendron*, *Dubouzetia*, *Peripentadenia* and *Vallea*. The other genera (including *Elaeocarpus*) were briefly assessed and discussed (Coode 1985, 1987). A robust phylogenetic analysis of the morphological traits has not yet been undertaken.

Several molecular phylogenetic studies on the family Elaeocarpaceae as well as the genus *Elaeocarpus* have been undertaken. Molecular studies by Maynard (2004, using sequence data from the internal transcribed spacer of nuclear ribosomal DNA, or ITS), Crayn *et al*. (2006, using plastid *trnL-trnF* region and nuclear ITS sequence data) and Baba (2013, using plastid *trnL-trnF* region, *trnV-ndhC* spacer and nuclear ITS sequence data) found that *Elaeocarpus*, *Aceratium* and *Sericolea* form a robust clade placed in a recently diverging position in the family. *Aceratium* and *Sericolea* are each monophyletic, but the monophyly of *Elaeocarpus* and the phylogenetic relationships within it are poorly resolved. Moreover, the studies of Maynard (2004) and Baba (2013) focused on the Australian *Elaeocarpus* with limited representatives from the surrounding regions, therefore, further assessments of the monophyly and relationships within *Elaeocarpus* are needed by expanding the range of DNA markers used and sampling size.

Crayn *et al*. (2006) estimated the age of the most recent common ancestor (MRCA) of the *Elaeocarpus* alliance clade at 46 ± 3 Ma, *Aceratium-Sericolea* 33 ± 2 Ma, *Aceratium* 18 ± 1 Ma and *Sericolea* 12 ± 1 Ma using Non-Parametric Rate Smoothing

(NPRS) and Penalised Likelihood (PL) analysis of a combined *trnL-trnF* region and ITS sequence dataset as implemented in the program r8s (Sanderson 2003). Niissalo (2011), on the other hand, estimated the age of the MRCA of the *Elaeocarpus* alliance clade at 41 Ma (estimated from the node of topology) using a relaxed clock model based on an ITS dataset as implemented in the program Bayesian Evolutionary Analysis Sampling Trees (BEAST) (Drummond and Rambaut 2008). In both studies, the age of *Elaeocarpus* was fixed at 30 Ma based on a conservative estimate of the minimum age of the *Elaeocarpus* macrofossils that were assigned to the crown group node. Furthermore, the sample size of *Elaeocarpus* in both studies was relatively small (13 species in Crayn *et al*. (2006); 7 species in Niissalo (2011)), hence the age of *Elaeocarpus* has not been adequately examined.

Table 1.1 Infrageneric classifications of *Elaeocarpus*.

Authors	**Regions**	**Sections/Groups**
Brongniart and Gris (1861)	New Caledonia	sect. *monocera* Brongn. & Gris
		sect. *dicera* Brongn. & Gris
		sect. *ganitrus* Brongn. & Gris
Bentham and Hooker (1862)	General description of	sect. *monocera* Brongn. & Gris
	Elaeocarpus	sect. *ganitrus* Brongn. & Gris (including sect. *dicera* Brongn. & Gris)
		sect. *craspedum* Bentham & Hooker *f*.
		sect. *acronodia* Bentham & Hooker *f*.
Masters (1874)	Asia (India, Sri Lanka,	sect. I = *ganitrus* Brongn. & Gris
	China & Southeast Asia)	sect. II = *dicera* Brongn. & Gris
		sect. III = *monocera* Brongn. & Gris
		sect. IV = *acronodia* Bentham & Hooker *f*.
Schlechter (1916)	Papuasia	sect. I = *lobopetalum* Schltr.
		sect. II = *dactylosphaera* Schltr.
		sect. III = *chascanthus* Schltr.
		sect. IV = *fissipetalum* Schltr.

		sect. V = *ptilanthus* Schltr.
		sect. VI = *oreocarpus* Schltr.
		sect. VII = *blepharoceras* Schltr.
		sect. VIII = *papuanthus* Schltr.
		sect. IX = *coilopetalum* Schltr.
Smith (1944, 1953)	Papuasia & Pacific islands	sect. *lobopetalum* Schltr.
		sect. *dactylosphaera* Schltr.
		sect. *chascanthus* Schltr.
		sect. *fissipetalum* Schltr.
		sect. *ptilanthus* Schltr.
		sect. *oreocarpus* Schltr.
		sect. *blepharoceras* Schltr.
		sect. *papuanthus* Schltr.
		sect. *coilopetalum* Schltr.
		sect. *monocera* Brongn. & Gris
		sect. *dicera* Brongn. & Gris
		sect. *ganitrus* Brongn. & Gris
Weibel (1968)	Malesia	sect. *Elaeocarpus* (= sect. *chascanthus* Schltr.)

		sect. *fissipetalum* Schltr. sect. *oreocarpus* Schltr. sect. *blepharoceras* Schltr. sect. *coilopetalum* Schltr. sect. *monocera* Brongn. & Gris sect. *dicera* Brongn. & Gris sect. *ganitrus* Brongn. & Gris sect. *acronodia* Weibel
Coode (1978, 1980)	Papuasia	group I = *lobopetalum* Schltr. group II = *dactylosphaera* Schltr. group III = sect. *Elaeocarpus*; sect. *chascanthus* Schltr. *sensu lato* subgroup A = sect. *chascanthus* Schltr. *sensu stricto* subgroup B = sect. *chascanthus* auct. non Schltr.: A.C. Sm. (1944) *pro parte* group IV = *blepharoceras* Schltr. *sensu stricto* group V = sect. *fissipetalum* Schltr.; sect. *ganitrus* Brongn. & Gris; sect. *ptilanthus* Schltr. subgroup A = sect. *ganitrus* Brongn. & Gris *sensu stricto* subgroup B = sect. *ganitrus* auct. non Brongn. & Gris: Schltr. (1916), A.C. Sm.

(1944) *pro parte*

subgroup C

subgroup D = sect. *fissipetalum* Schltr.

subgroup E = sect. *fissipetalum* auct. non Schltr.: A.C. Sm. (1944) *pro parte*

group VI = sect. *monocera* Brongn. & Gris; sect. *papuanthus* Schltr.; sect. *blepharoceras* Schltr. *pro parte*; sect. *oreocarpus* auct. non Schltr.: A.C. Sm. (1944) *pro parte*

subgroup A = sect. *monocera* sensu A.C. Sm. (1944) *pro parte*

subgroup B

subgroup C = sect. *papuanthus* Schltr.

subgroup D = sect. *blepharoceras* Schltr. *pro parte*

subgroup E = sect. *oreocarpus* auct. non Schltr.: A.C. Sm. (1944) *pro parte*

group VII = sect. *oreocarpus* Schltr. *sensu stricto*

group VIII = sect. *coilopetalum* Schltr.; sect. *blepharoceras* auct. non Schltr.: A.C. Sm. (1944) *pro parte*

subgroup A = sect. *blepharoceras* auct. non Schltr.: A.C. Sm. (1944) *pro parte*; sect. *coilopetalum* auct. non Schltr.: A.C. Sm. (1944) *pro parte*

subgroup B = sect. *coilopetalum* auct. non Schltr.: A.C. Sm. (1944) *pro parte*

		subgroup C = sect. *coilopetalum* Schltr. pro parte; sect. *blepharoceras* auct. non Schltr.: A.C. Sm. (1944) *pro parte*
		subgroup D = sect. *coilopetalum* Schltr. *pro majore parte*
		group IX
Coode (1984)	Australia & New Zealand	group IV
		group V
		subgroup A
		subgroup D
		group VI
		subgroup B
		subgroup E
		subgroup F
		group VII
		group X
		group XI
		subgroup A
		subgroup B
		group XII

Coode and Weibel (1994) and Coode (1995, 1996a – c, 1998, 2001a – f, 2002, 2003, 2005, 2007a – b, 2010)	Malesia	*acronodia* Weibel group *coilopetalum* Schltr. group sect. *Elaeocarpus* (group) *fissipetalum* Schltr. group *ganitrus* Brongn. & Gris group *monocera* Brongn. & Gris group *coloides* Coode subgroup *debruynii* Coode subgroup *monocera* subgroup *obtusus* Coode subgroup *verticellatae* Coode subgroup *myrtoides* Coode group *polystachyus* Coode group

1.3 Thesis outline

The findings of this thesis are presented in three interdependent data chapters outlined below, each written in a manuscript style.

In **Chapter 2**, the phylogeny of *Elaeocarpus* is inferred using both plastid and nuclear DNA markers to test hypotheses regarding monophyly of the genus and the relationships within it. The relationships between *Elaeocarpus*, *Aceratium* and *Sericolea* and the main lineages within *Elaeocarpus* are elucidated using MP, ML and Bayesian Inference analysis of independent and combined sequence data matrices. Furthermore, tree topologies generated from the plastid and nuclear sequence data are used to: compare the degree of support inferred from MP, ML and Bayesian analyses; and determine the congruence of phylogenies estimated from uniparentally (*psbA-trnH* spacer, *trnL-trnF* region and *trnV-ndhC* spacer) and biparentally (*Xdh*) inherited genes. Additionally, possible evolutionary transformations of seed morphological characters in the lineages within *Elaeocarpus* are discussed.

In **Chapter 3**, the historical biogeography and evolutionary divergence times of *Elaeocarpus* are estimated based on the inferred molecular phylogeny presented in Chapter 2. An uncorrelated lognormal Bayesian relaxed molecular clock method with five fossil calibration points is used to estimate diversification times of the main clades within the family with a focus on the genus *Elaeocarpus*. The origin and subsequent diversification of *Elaeocarpus* are discussed.

In **Chapter 4**, a morphometrics approach is used to investigate taxon boundaries within the *polystachyus* group (i.e. the *E. polystachyus* complex), in the light of the phylogenetic relationships inferred in Chapter 2 (i.e. the *polystachyus* clade + *E. polystachyus*). Since the *polystachyus* group studied here is morphologically relatively simple, the traditional morphometrics method is, therefore, considered as appropriate approach. Congruency between the results of the morphometrics and the alpha taxonomy (Coode 1996c) are assessed. Precursory evidence on the

morphological variation versus ecological adaptations is provided, and the relationships within the group with the taxonomic implications are discussed.

Chapter 5 highlights the key findings of this thesis, then concludes with discussions and future research directions.

The following people made valuable contributions to this chapter: Mark Coode identified and confirmed the identification of the *Elaeocarpus* vouchers; Philippa Griffin processed the Next-Generation raw data, formatted them for phylogenetic analysis and wrote section 2.2.6; and Ferry Slik and Yumiko Baba shared their unpublished *psbA-trnH* sequences of Chinese *Elaeocarpus*, and the *trnL-trnF* region and *trnV-ndhC* sequences of the Australasian and the Pacific islands species, respectively.

Chapter 2: Phylogeny of *Elaeocarpus* (Elaeocarpaceae) based on DNA sequence data from four loci and the implications for taxonomy and the evolution of seed morphology

ABSTRACT

Phylogenetic relationships within the genus *Elaeocarpus* (Elaeocarpaceae) were investigated using over 3000 bp of DNA sequence from four regions: the plastid *psbA-trnH* spacer, *trnL-trnF*, and *trnV-ndhC* spacer regions, and the nuclear xanthine dehydrogenase gene (*Xdh*). The four DNA datasets were analysed separately and in combination using maximum parsimony (MP) and Bayesian methods. The phylogenetic trees constructed using both phylogenetic methods, based on the combined data, strongly support the monophyly of *Elaeocarpus*. The trees generated using MP were less well resolved, but relationships were similar to those obtained using the other method. Bayesian analyses recovered trees with short branch lengths within *Elaeocarpus*, showing 13 labelled lineages or clades broadly corresponding to the current infrageneric classification. This study also traced the evolutionary changes in two seed morphological characters, embryo shape (straight or curved) and endosperm ornamentation (entire or ruminate), which have been used as the diagnostic characters in infrageneric classifications. Results of the parsimony reconstruction of these character states indicated that both states are homoplasious at higher taxonomic level (i.e. genus level and above), although the curved embryo within *Elaeocarpus* is homologous by excluding *E. holopetalus*.

2.1 Introduction

The Elaeocarpaceae is a medium sized angiosperm family comprising 12 genera (*Aceratium* DC., *Aristotelia* L'Hér., *Crinodendron* Molina, *Dubouzetia* Brongn. & Gris., *Elaeocarpus* L., *Peripentadenia* L.S.Sm., *Platytheca* Steetz, *Sericolea* Schltr., *Sloanea* L., *Tetratheca* Sm., *Tremandra* R.Br. and *Vallea* L.) and *c*. 550 species (Coode 2004). Most members of the family are trees or shrubs, although a few species can occur as epiphytes or lianes. The family is widely distributed in the neo- and palaeo-tropics, and in the subtropics and temperate regions of the southern hemisphere, but is absent from mainland Africa (Coode 2004).

Elaeocarpus L. is the most species-rich genus in the family Elaeocarpaceae, comprising 350 – 400 species (inclusive of new, currently undescribed species) (Coode pers. comm. 2014). The genus includes shrubs and large trees inhabiting mostly the palaeo-tropical and palaeo-subtropical forests ranging from lowland to montane areas of Madagascar, Asia, Australia and the Pacific islands (Coode 2004). New Guinea with *c*. 85 species and Borneo with *c*. 70 species are the major centres of species diversity (Coode 2004).

2.1.1 Infrageneric classifications of Elaeocarpus *using a classical taxonomic approach*

While a number of regional treatments and checklists of *Elaeocarpus* exist, such as those for Papuasia (Schlechter 1916; Smith 1944; Weibel 1968, 1971; Coode 1978), Australia and New Zealand (Coode 1984), Pacific Islands (Smith 1953) and Malesia (Coode 1978, 1996a – c, 2001b, 2001e – f, 2010; Coode and Weibel 1994), the genus has never been fully monographed. Therefore the overall species-level taxonomy of *Elaeocarpus* and the patterns of morphological variation within it are inadequately known.

Several infrageneric classifications of *Elaeocarpus* have been established for the species occurring in China, Malesia, Papuasia, Australasia and the Pacific islands (Table 1.1). No similar classifications exist for species occurring outside of these

regions. The first of these infrageneric classifications of *Elaeocarpus* was that of Brongniart and Gris (1861) for the New Caledonian species. This classification, based on reproductive morphological characters (i.e. stamens, anthers, ovary, number of loculi per ovary, number of ovules per loculus and number of seeds per fruit), was broadly followed by Bentham and Hooker (1862), Masters (1874) and Tang (1992) with minor modifications to accommodate morphological variation in the local species in India (including Nepal, Bhutan, Bangladesh and Sri Lanka) and Southeast Asia.

Schlechter (1916), in his revision for the Papuasian species, proposed a new infrageneric classification which emphasised the number of ovules per loculus as the superior diagnostic character of those used in Brongniart and Gris' (1861) classification. His classification was broadly followed by Smith (1944, 1953), Weibel (1968) and Tang (1992) for species occurring in Papuasia, Pacific islands and Malesia.

Coode (1978, 1996a – c, 2001a – b, 2001d – f, 2003, 2010) and Coode and Weibel (1994) proposed significant changes to the classification of Brongniart and Gris (1861) in a series of revisions of the Malesian species. Combinations of vegetative and reproductive characters were used to define groups within *Elaeocarpus*. These characters included branching pattern, the length and/or indumentum on the shoot, petiole, lamina and inflorescence, the number of flowers per inflorescence, flower-mery and sexuality, the presence of a keel or pocket on the petals, the degree of development of the disk, stamen morphology, the number of loculi per ovary and the number of ovules per loculus, the shape of transverse section of the fertile loculus and the number of seeds per fruit, the shape of the fruit stones, the shape of the embryo (e.g. straight (Fig. 1.1G) or curved (Fig. 1.1H)) and the condition of the endosperm (e.g. entire (Fig. 1.1G) or ruminate (Fig. 1.1H)). Of all the morphological characters used, they considered the seed characters, i.e. the shape of the embryo, the cotyledon (broad or narrow) and the condition of the endosperm, to be superior and hence all of the groups previously established by Brongniart and Gris (1861) were re-defined in terms of these characters, except group I (= *lobopetalum* Schltr.) and group II (= *dactylosphaera* Schltr.) (Table 1.1). While these characters are often useful in defining group boundaries, exceptions do occur in some groups, for example members of the

widespread *coilopetalum* group have curved embryos and the endosperm can be either entire or ruminate (Coode 1998). Hence, the transformation of these characters regarded as superior by Weibel and Coode needs to be examined to test the homology of the states and hence their usefulness in diagnosing monophyletic taxa.

A total of 15 infrageneric groups and 27 subgroups of *Elaeocarpus* are recognised in Coode's informal system (Table 1.1). Some of these groups were recognised in previous classifications (Brongniart and Gris 1861; Smith 1944, 1945; Weibel 1968; Coode and Weibel 1994; Coode 1995, 1996a – c, 1998, 2001a – f, 2002, 2003, 2005, 2007a – b, 2010) and the names adopted, some are numbered (Coode 1984) and some are both named and numbered (Schlechter 1916; Coode 1978, 1980). This classification is based on an alpha taxonomic approach, and the monophyly of the infrageneric groups and the phylogenetic relationships between them have not been adequately investigated.

2.1.2 Molecular phylogenetic relationships of Elaeocarpaceae and Elaeocarpus

Phylogenetic relationships within Elaeocarpaceae and the genus *Elaeocarpus* have been investigated in several studies using phylogenetic analysis of nuclear and plastid DNA sequences (Maynard 2004; Crayn *et al*. 2006; Baba 2013). Of these studies, Crayn *et al*. (2006) focused on relationships among the genera using nuclear ITS and plastid *trnL-trnF* sequences from representatives of all accepted genera, including the three formerly placed in Tremandraceae (*Platytheca*, *Tetratheca*, *Tremandra*). The results indicate that *Elaeocarpus*, *Aceratium* and *Sericolea* form a recently diverged lineage in the family consistent with the grouping retrieved in Coode's (1987) cladistic analysis based on morphology (and later named the *Elaeocarpus* alliance, Coode 2004), *Aceratium* and *Sericolea* were each resolved as monophyletic and strongly supported as sisters. The great majority of *Elaeocarpus* sampled formed a robust clade but excluding *Elaeocarpus sedentarius* D.J.Maynard & Crayn (treated as *Elaeocarpus* sp. "Rocky Creek"), therefore the monophyly of *Elaeocarpus* was not supported. *Elaeocarpus* is morphologically distinguishable from *Aceratium* and *Sericolea* as its leaves are usually alternate or spirally-arranged, and the

fruit is a drupe with woody stone (inner mesocarp), while the radially fibrous outer mesocarp is usually caducous or absent when dried (except in some Australian (*E. johnsonii* F.Muell. ex C.T.White and *E. sedentarius*) and New Guinean (*E. blepharoceras* Schltr. and *E. womersleyi* Weibel) species, where the fibres are persistent when dried), whereas *Aceratium* and *Sericolea* have leaves that are usually opposite when mature and the fruits are either drupes with a persistent, radially fibrous outer mesocarp overlying a weakly woody inner mesocarp (*Aceratium*), or berries (*Sericolea*) (Coode 2004).

The phylogenetic relationships among *Elaeocarpus* species are not well understood. A few molecular phylogenetic analyses have been undertaken to date but these were limited in terms of both number of loci and taxon sampling. Crayn *et al*. (2006) included *trnL-trnF* and ITS sequence data of 13 *Elaeocarpus* species (mostly from Australia) in an analysis of intergeneric relationships within Elaeocarpaceae *sensu lato*. Two analyses have focused on relationships within the genus. Maynard (2004) analysed ITS sequence data for 29 *Elaeocarpus* taxa mostly from Australia, and Baba (2013) analysed *trnL-trnF*, *trnV-ndhC* intergenic spacer and ITS sequence data for 59 *Elaeocarpus* taxa. Together these molecular studies resolved with some confidence the phylogenetic relationships between most Elaeocarpaceae genera. However, the low resolution of the DNA markers and the limited representation of *Elaeocarpus* species from biogeographical regions other than Australia left the monophyly and interspecific relationships of *Elaeocarpus* unresolved and the current infrageneric system untested.

2.1.3 Aims

The aims of this study were: (1) to evaluate the monophyly and relationships of within the *Elaeocarpus* alliance, (2) to determine the main lineages within *Elaeocarpus* and evaluate the monophyly of groups in Coode's system, and (3) to determine the evolutionary transformations in embryo and endosperm characters and their usefulness in diagnosing monophyletic groups of *Elaeocarpus*.

2.2 Materials and methods

2.2.1 Taxon sampling and DNA extraction

The sampling strategy aimed to maximise representation of different morphological groups and biogeographic regions. A total of 176 samples were studied including the outgroups. The ingroups samples represent 114 species of *Elaeocarpus* (including three putative new species from Australia): 45 species from Malesia, 33 from Australasia (excluding New Guinea), 22 from Pacific islands (including New Caledonia, 13 from continental Asia and 1 from Madagascar. Widespread species (*E. floribundus* Blume) and/or species complexes (*E. angustifolius* Blume (Coode 2010), *E. mastersii* King (Coode 1996b) and *E. stipularis* (Coode 2001c)) were represented by multiple samples collected from different geographical localities or representing different morphological variants. Additionally, some taxa were represented by multiple samples collected from different locations.

Because the phylogenetic relationships within the *Elaeocarpus* alliance are incompletely resolved (Crayn *et al*. 2006; Baba 2013), representatives from the latter two genera were included in the study. Representatives of the *Crinodendron*, *Dubouzetia* and *Peripentadenia*, the closest relatives of the *Elaeocarpus* alliance (Crayn *et al*. 2006), were included as outgroups. Altogether 11 outgroup taxa were used: *Crinodendron hookerianum* Gay, *Crinodendron patagua* Molina, *Dubouzetia campanulata* Pancher ex Brongn. & Gris*, Dubouzetia caudiculata* Sprague*, Dubouzetia confusa* Guillaumin & Virot*, Dubouzetia elegans* Brongn. & Gris*, Dubouzetia guillauminii* Virot*, Dubouzetia kairoi* Coode*, Dubouzetia saxatilis* A.R.Bean & Jessup*, Peripentadenia mearsii* (C.T.White) L.S.Sm. and *Peripentadenia phelpsii* B.Hyland & Coode. The phylogenetic trees were rooted on *Peripentadenia* and *Crinodendron*, the early diverging lineages to *Dubouzetia* and the *Elaeocarpus* alliance clade (Crayn *et al*. 2006). Taxa sampled, collection information and GenBank accession numbers for the sequences in the four datasets are listed in Appendix 2.1.

Total genomic DNA was either obtained from the DNA bank at the Australian Tropical Herbarium (ATH) or extracted from silica gel dried leaf tissue using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol.

2.2.2 Selection of DNA markers

A total of eight DNA markers were tested using a sample subset to determine their utility for phylogenetic reconstruction in *Elaeocarpus*. Of these five were plastid (cpDNA: the *psbA-trnH* intergenic spacer, the *trnL* intron and *trnL-trnF* intergenic spacer [hereafter named "*trnL-trnF* region"], the *trnV-ndhC* intergenic spacer, the *trnQ-rps16* intergenic spacer, and the *ndh*F gene) and three were nuclear (nDNA: the Internal Transcribed Spacer [ITS], the chloroplast-expressed glutamine synthetase gene [ncpGS], and the xanthine dehydrogenase gene [*Xdh*]). These DNA markers were selected based on the results of previous studies of the genus *Elaeocarpus* (Maynard 2004; Baba 2013), the families Elaeocarpaceae (Crayn *et al*. 2006), Oxalidaceae (Emshwiller and Doyle 1999) and Cunoniaceae (Pillon *et al.* 2009), and angiosperms (Morton 2011). Up to 14 taxa from the following sample subset were used in the screening of DNA markers: *Aceratium concinnum* (S.Moore) C.T.White, *A. doggrellii* C.T.White, *Crinodendron hookerianum*, *C. patagua*, *Dubouzetia campanulata*, *D. saxatilis*, *Elaeocarpus angustifolius* Blume, *E. clementis* Merr. var. *clementis*, *E. clementis* Merr. var. *clemensiae* (R.Knuth) Coode, *E. cupreus* Merr., *E. euneurus* Stapf ex Ridl., *E. jugahanus* Coode, *E. marginatus* Stapf ex Weibel, *E. multinervosus* R.Knuth, *E. mutabilis* Weibel, *E. obtusus* Blume subsp. *obtusus*, *E. submonoceras* Miq. subsp. *lasionyx* (Stapf ex Ridl.) Weibel, *Sericolea calophylla* (Ridl.) Schltr. subsp. *grossiserrata* Coode and *S. micans* Schltr. var. *micans*. Both distantly and closely related taxa, based on current infrageneric classifications (Coode 1978, 1996a – c, 2001a – b, 2001d – f, 2003, 2010; Coode and Weibel 1994) and previous molecular studies (Maynard 2004; Crayn *et al*. 2006; Baba 2013), were selected. Amplification, sequencing and phylogenetic analysis (using maximum parsimony) of these markers followed the methods described in sections 2.2.3 and 2.2.7 (Table 2.1).

Table 2.1 List of primers and adapters used for amplification and sequencing (Sanger and Illumina MiSeq) of five DNA regions.

DNA region	Primer	Sequence (5' to 3')	Reference
Sanger sequencing			
psbA-trnH spacer	trnH2	CGCGCATGGTGGATTCACAATCC	Tate and Simpson (2003)
	psbAF	GTTATGCATGAACGTAATGCTC	Sang *et al*. (1997)
trnL-trnF region	c	CGAAATCGGTAGACGCTACG	Taberlet *et al*. (1991)
	f	ATTTGAACTGGTGACACGAG	Taberlet *et al*. (1991)
trnV-ndhC spacer	$trnV^{(UAC)}$x2	GTCTACGGTTCGARTCCGTA	Shaw *et al*. (2007)
	ndhC	TATTATTAGAAATGYCCARAAAATATCATATTC	Shaw *et al*. (2007)
ndhF	NDHF-2F	AGGTACACTTTCTCTTTGCGGTATTCC	Kim and Jansen (1995)
	2110R	CCCCCAYATATTTGATACCTTCTC	Kim *et al*. (2001)
ncpGS	GScp687f	GATGCTCACTACAAGGCTTG	Emshwiller and Doyle (1999)
	GScp994r	AATGTGCTCTTTGTGGCGAAG	Emshwiller and Doyle (1999)
ITS	ITS4	TCCTCCGCTTATTGATATGC	White *et al*. (1990)
	ITS5	GGAAGTAAAAGTCGTAACAAGG	White *et al*. (1990)
	17SE	ACGAATTCATGGTCCGGTGAAGTGTTCG	Sun *et al*. (1994)
	26SE	TAGAATTCCCCGGTTCGCTCGCCGTTAC	Sun *et al*. (1994)
	ITS-A	GGAAGGAGAAGTCGTAACAAGG	Blattner (1999)
	ITS-B	CTTTTCCTCCGCTTATTGATATG	Blattner (1999)
	ITS-C	GCAATTCACACCAAGTATCGC	Blattner (1999)
	ITS-D	CTCTCGGCAACGGATATCTCG	Blattner (1999)
Xdh	486F	ACYCCYGGKTTTRTIATGTCIATGTA	Morton (2011)
	1362R	TCWGATATTGGACTAGCWGT	Morton (2011)
	Xdh 18F	GAGCAGGACGCTCAGTTACAACC	Present study
	Xdh 18R	TCGCTGCACAGGCTACAGACC	Present study
Illumina MiSeq Sequencing			
Adapter	Forward (append to 5' end of the forward primer)	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	MiSeq (Illumina Inc. 2013)

Adapter	Reverse (append to 5' end of the reverse primer)	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	MiSeq (Illumina Inc. 2013)
Xdh	Xdh 18F	as above	as above
	Xdh 18R	as above	as above
	Xdh internal 1F	GGT CGA TAC CAC TTA AGC CCA CCA	Present study
	Xdh internal 1R	TGG TGG GCT TAA GTG GTA TCG ACC	Present study
ITS	ITS-A	as above	as above
	ITS-B	as above	as above
	ITS-C	as above	as above
	ITS-D	as above	as above

Out of a total of eight markers screened, five (*psbA-trnH*, *trnL-trnF* region, *trnV-ndhC*, ITS and *Xdh*) showed appropriate levels of variation and were selected for the full study.

2.2.3 PCR amplification

The primers used for PCR amplification are shown in Table 2.1. PCR and purification reactions were performed in an Eppendorf AG 22331 Thermal Cycler (Hamburg, Germany).

PCR amplification of *psbA-trnH* followed the iProof™ High-Fidelity DNA Polymerase kit protocol (BIO-RAD, Hercules, California, USA) with modification; *trnV-ndhC*, *trnL-trnF* region, *trnQ-rps16*, *ndhF* and ITS followed the KAPA Taq DNA Polymerase protocol (Kapa Biosystems, Wilmington, Massachusetts, USA) with modification; *ncpGS* followed the protocols of Emshwiller and Doyle (1999) and Pillon *et al.* (2009) with modification; *Xdh* gene followed Morton's (2011, pers. comm. 2013) protocol with modification. Details are presented in Table 2.2.

Table 2.2 Ingredients and cycling conditions used for PCR amplification of the six plastid and three nuclear markers explored in this study.

Locus	Polymerase (5U/uL)	Buffer (uL)	$MgCl_2$ (uL)	dNTPs (uL)	DMSO (0.4mM solution, uL)	BSA (0.4% solution, uL)	Primers, each (10mM solution, uL)	Template DNA (uL)	Distilled H20 (uL)
psbA-trnH	0.25[1]	4[2]	-	0.4[3]	0.6	-	1	1	12.75
trnV-ndhC	0.15 – 0.2[4]	2.5[5]	1.25 – 2.5[6]	0.4 – 0.5[3]	0.6 – 0.8	0 – 0.25	0.6 – 1	1 – 2	10.25 – 13.5
trnL-trnF									
trnQ-rps16									
ndhF									
ITS									
ncpGS	0.2	2[5]	1.2[7]	0.4[8]	-	0.8	0.8	1	13.6
Xdh	0.2	2[5]	1.2 – 2.4[7]	0.4[3]		0.8	0.6 – 0.8	2	11.4 – 12.8

Locus	Initial denaturation (temperature, length)	Denaturation (cycles, temperature, length)	Annealing (temperature, length)	Extension (temperature, length)	Final extension (temperature, length)	Hold (temperature)
psbA-trnH	98 °C, 45 s	35 cycles, 98 °C, 10 s	64 °C, 30 s	72 °C, 40 s	72 °C, 10 min	10 °C
trnV-ndhC	95 °C, 2 min	35 cycles, 95 °C, 30 s	55 °C, 30 s	72 °C, 1 min	72 °C, 2 min	
trnL-trnF						
trnQ-rps16						
ndhF						
ITS						
ncpGS	94 °C, 2 min	35 cycles, 94 °C, 1 min	59.4 – 62.7 °C, 1 min	72 °C, 1 min 30 s	72 °C, 5 min	
Xdh	96 °C, 1 min	34 cycles, 94 °C, 30 s	50 °C, 30 s	72 °C, 1 min	72 °C, 7 min	

[1] iProof™ DNA Polymerase (BIO-RAD, Hercules, California, USA)
[2] 5X iProof™ HF buffer (BIO-RAD, Hercules, California, USA)
[3] 10mM each dNTP in solution (Kapa Biosystems, Wilmington, Massachusetts, USA)
[4] KAPA Taq DNA Polymerase (Kapa Biosystems, Wilmington, Massachusetts, USA)
[5] Platinum® *Taq* DNA Polymerase (Invitrogen™, Life Technologies, New York, USA)
[6] 50 mM solution (Invitrogen™, Life Technologies, New York, USA)
[7] 25 mM solution (Kapa Biosystems, Wilmington, Massachusetts, USA)
[8] 250 µM each dNTP in solution (Kapa Biosystems, Wilmington, Massachusetts, USA)

PCR products of *psbA-trnH*, *trnL-trnF* region, *trnV-ndhC*, *trnQ-rps16* and *ndhF* regions were purified in a 14 µl mixture containing 10 µl PCR products and 4 µl of a mixture of 4 % Fermentas* Exonuclease I (Exo I) (Thermo Scientific, Waltham, Massachusetts, USA), 20 % Fermentas* Shrimp Alkaline Phosphatase (SAP) (Thermo Scientific, Waltham, Massachusetts, USA) and 76 % of distilled H_2O. The PCR products were purified using the following incubation sequence; 37 ^{o}C for 20 min; 80 ^{o}C for 15 min; hold at 10 ^{o}C. PCR products of *Xdh* were purified with the Wizard® SV Gel and PCR Clean-Up System (Promega Corporation, Madison, Wisconsin, USA) following the manufacturer's protocol.

2.2.4 *Sanger sequencing*

Purified PCR products were sequenced using the BigDye® Terminator ver. 3.1 Cycle Sequencing Kit (Applied Biosystems, New York, USA) following the manufacturer's protocol. Capillary separation of the products was performed on an ABI 3730 x I automated sequencer at the Australian Genome Research Facility (Brisbane, Australia). Chromatograms were examined and edited using ChromasPro version 1.32 (Technelysium Pty. Ltd., South Brisbane, Queensland, Australia). Consensus sequences for each sample were assembled using BioEdit version 7.0.9.0 (Hall 1999).

2.2.5 Next-Generation sequencing

Multiple copies of the *Xdh* gene and ITS region were consistently amplified for most of the samples (Fig. 2.1 & 2.2) using Sanger sequencing technology, despite much troubleshooting of the PCR conditions and mixtures. Therefore, the Illumina MiSeq (Illumina Inc., San Diego, California, USA) Next-Generation Sequencing (NGS) platform was utilised. This platform NGS has the capability to interpret a mixture of templates and in the case of this study, the ability to sequence multiple amplified copies efficiently in a single run. These different copies can then be segregated bioinformatically.

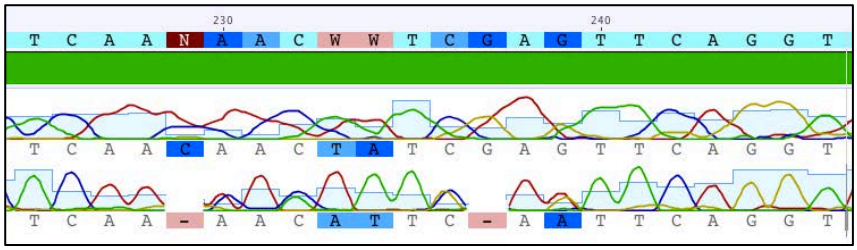


Fig. 2.1 A fragment of the chromatogram of *Elaeocarpus johnsonii* showing divergent gene copies in the reverse (above) and forward (below) sequences of the *Xdh* gene.

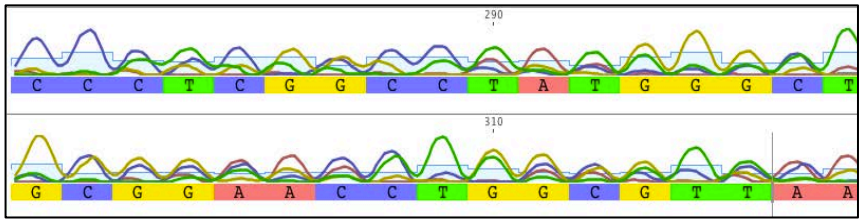


Fig. 2.2 A fragment of the chromatogram of *Elaeocarpus nanus* subsp. *congestifolius* showing divergent gene copies in the forward sequence of the ITS region.

For the Miseq amplicon preparation, amplicons of the samples were prepared using two rounds of PCR. The MiSeq platform is limited to an amplicon read length of approximately 2 × 250 bp (i.e. paired reads, from opposite ends of the fragments). Therefore both *Xdh* (total 690 bp) and ITS (625 bp) were amplified as two smaller fragments with about 50 bp of overlap using the primers as shown in Table 2.1.

In the first round of PCR, *Xdh* and ITS were amplified using the KAPA Taq DNA Polymerase manufacturer's protocol (Kapa Biosystems, Wilmington, Massachusetts, USA) as detailed above. An overhanging adapter was appended to the 5' end of each primer to form fusion primers as suggested in Illumina (2013) (Table 2.1).

The PCR amplification protocol for the first fragment of *Xdh* differed from that described for *Xdh* in section 2.2.3 only in the conditions which were as follows: initial denaturation at 96 ^{o}C for 3 min, 35 cycles of denaturation at 96 ^{o}C for 30 s, annealing at 52 ^{o}C for 30 s, and extension at 72 ^{o}C for 30 s; final extension step of 72 ^{o}C for 5 min; hold at 10 ^{o}C. PCR amplification of the second fragment of *Xdh* used the following

conditions: initial denaturation at 96 ^{o}C for 1 min; 31 – 35 cycles of denaturation at 94 ^{o}C for 30 s, annealing at 54 ^{o}C for 30 s, and extension at 72 ^{o}C for 30 – 40 s; final extension step of 72 ^{o}C for 5 min; hold at 10 ^{o}C.

PCR amplification of the two fragments of ITS was done using the same PCR mastermix and thermal cycler as described for ITS in section 2.2.3 with the following conditions: initial denaturation at 98 ^{o}C for 45 s; 35 cycles of denaturation at 98 ^{o}C for 10 s, annealing at 64 ^{o}C for 30 s, and extension at 72 ^{o}C for 30 – 40 s; final extension at 72 ^{o}C for 10 min; hold at 10 ^{o}C.

PCR products were purified with ExoFastAP (Thermo Scientific, Waltham, Massachusetts, USA) in 6.5 µL mixture of 5 µL PCR products and 1.5 µL of a mixture of 20 U/µL Fermentas* Exonuclease I (Exo I) and 1 U/µL of Fast AP. The PCR products were purified using the following incubation sequence: 37 ^{o}C for 15 min; 85 ^{o}C for 15 min; hold at 10 ^{o}C.

In the second round of PCR, amplicons were prepared using the Nextera® Index Kit (96 Indices, 384 Samples) (product code: FC-121-1012, Illumina Inc., Woodlands, Singapore) following the manufacturer's protocol. Amplicons from each sample were uniquely dual-indexed and terminated with N5 and N7 indexes at each read end to make the amplicons compatible with the MiSeq flow cell and identifiable after sequenced. Purification and quality control of the amplicons, together with library preparation were performed at the Ramaciotti Centre (University of New South Wales, Kensington, New South Wales, Australia).

2.2.6 NGS data cleaning and assembly of Sanger and NGS data

The NGS raw reads were processed using a custom pipeline which combines existing tools and custom scripts (written by Dr. Phillipa Griffin, University of Melbourne). First, raw reads were trimmed using Trimmomatic v 0.30 (Lohse *et al*. 2012): leading and trailing bases with a quality score below 30 were trimmed and then a sliding window trimmer (window size = 10 bp, mean quality threshold = 25) was

applied across the read. Reads under 200 bp in length after trimming were discarded. The following steps were then performed using tools from the Qiime package (Caporaso *et al*. 2010): joining paired-end reads; renaming of reads to reflect the Qiime naming conventions and converting from fastq to fasta format; and lastly a round of read clustering with the pick_otus.py tool using the CD-HIT algorithm (Fu *et al*. 2012) and a minimum similarity of 0.9985, meaning that only reads with at least 99.85% identity were clustered. The amplicon identity of each cluster was then obtained by matching to a custom BLAST database (blast_wrapper.py) that contained one example sequence for each of the amplicons (obtained from previous Sanger sequencing). The clusters were then ranked by the number of reads they contained, and the 12 highest-ranked clusters for each amplicon were retained. Each retained cluster was aligned using MAFFT (Katoh and Frith 2012) and a consensus sequence produced using a custom Python script, recording the total number of reads in the consensus sequence name. The consensus sequences were examined in Geneious® version 6.1.5 (Biomatters Ltd., Auckland, New Zealand). For each individual, PCR chimeras were manually identified as sequences with a low total read number and with obvious recombination sites with respect to the more abundant sequence copies. These were discarded, and sequences with obvious identity (e.g. differing only at one or two homopolymer repeat sites) were combined. Finally, sequences of all remaining amplicon copies (at maximum 4 per individual, corresponding to 2 genomic copies with 2 alleles each) were exported for further multi-individual alignment. An improved version of this pipeline, with additional functionality, is publicly available (https://github.com/griffinp/qiime-amp_reseq).

Sanger sequence data were edited and assembled using BioEdit version 7.0.9.0 (Hall 1999). Finally, assembly of both Sanger and NGS sequence data to form contigs was done using Geneious® version 6.1.5 (Biomatters Ltd., Auckland, New Zealand).

2.2.7 Alignment

Contig sequences were aligned using MAFFT multiple sequence alignment software version 7 (Katoh and Standley 2013) and then refined manually using Geneious® version 6.1.5 (Biomatters Ltd., Auckland, New Zealand).

One to several polyT/polyA/polyTA stretches were found in the *psbA-trnH* spacer (46 bp), *trnL-trnF* region (44 bp) and *trnV-ndhC* spacer (20 bp) regions. These stretches could not be unambiguously aligned and were therefore excluded from analyses, giving final alignments of 1113, 1264 and 835 characters, respectively. In the *psbA-trnH* spacer, alignment of a further 37 characters was somewhat ambiguous, but initial independent Bayesian analyses, with and without this region, resulted in similar topologies and support values (data not shown), therefore this region was included in all analyses. In the *trnV-ndhC* spacer dataset, *E. ferruginiflorus NSW 1754* lacked four bp of TA repeats in a polyTA stretch, whereas *E. polydactylus NSW 170* lacked two bp of A repeats in a polyA stretch. Both of these occurrences were coded as missing data instead of gaps to minimise the effect of possible alignment bias.

Alignment of ITS was done using MAFFT with minimal manual refinement. Up to four copies of ITS were obtained for each sample, and the orthologous copy could not be identified with confidence. Because phylogenetic reconstruction based on paralogous gene copies could generate incorrect estimates of the phylogeny, the ITS data were not included in the phylogenetic analyses. DNA sequences of the protein-coding gene *Xdh* were translated to amino acid sequences using Geneious® version 6.1.5 (Biomatters Ltd., Auckland, New Zealand) to check for internal stop codons or frameshift mutations (Tutar 2012) which may indicate non-functional (possibly paralogous) gene copies. Identification of synonymous and non-synonymous codons was done following Graur (2003). Two paralogous copies of *Xdh* were identified, but only copy 1 was used because its sample size (95.8 % of the total samples sequenced) was larger than copy 2 (57.3 %), minimising missing data. The most abundant alleles were selected and the two fragments, which sequenced independently due to the read length limitation of MiSeq, were partitioned.

Alignment gaps identified as potentially parsimony-informative indels were coded according to the simple method described by Simmons and Ochoterena (2000) using Seqstate version 1.4.1. For the *Xdh* region, gaps were observed between *Elaeocarpus* and outgroups, but not within *Elaeocarpus*. For all other loci gaps were observed within *Elaeocarpus*.

2.2.8 Phylogenetic analyses

Phylogenetic reconstruction was undertaken using maximum parsimony (MP), maximum likelihood (ML) and Bayesian Inference (BI) analyses. Parsimony analyses were conducted on the four DNA regions individually, on a combined plastid (*psbA-trnH* spacer + *trnL-trnF* region + *trnV-ndhC* spacer, hereafter named "cpDNA") matrix, and on a combined plastid and nuclear (*psbA-trnH* spacer + *trnL-trnF* region + *trnV-ndhC* spacer + *Xdh*, hereafter named "cp+nDNA") matrix using PAUP* version 4b10 (Swofford 2003). For the analyses of the combined data matrices, only taxa that had at least two regions sequenced were included. All characters were weighted equally and treated as independent and unordered (Fitch parsimony) and gaps were treated as missing data (Fitch 1971). The most parsimonious trees were obtained using MP ratchet analyses (Nixon 1999) with command files generated using PRAP (Müller 2003). For each of the ten random addition replicates, 200 ratchet iterations were performed. Each iteration comprised 10 rounds of TBR swapping, saving one shortest tree, and the most parsimonious trees were used to compute the consensus trees. Statistical support for the clades was estimated by nonparametric bootstrapping as implemented in PAUP*. Bootstrap support was calculated using 1000 simple stepwise addition pseudoreplicates with TBR branch swapping, and 100 trees saved per replicate. Tree length, consistency indices (CI) and retention indices (RI) were calculated for trees from the separate and combined analyses.

For the ML and BI analyses, MrModeltest version 2.3 (Nylander 2008) was used to determine the appropriate nucleotide substitution model and gamma rate heterogeneity using the Akaike Information Criterion (AIC) (Posada and Crandall 1998; Posada and Buckley 2004; Posada 2006) for each locus. The best-fit model determined

for the *trnL-trnF* region and *trnV-ndhC* spacer was the General Time Reversible model with among site rate variation modelled as a Gamma distribution (GTR + Γ); for the *psbA-trnH* spacer, the Hasegawa, Kishino and Yano model incorporating a parameter (I) to account for the proportion of invariant sites and among site rate variation modelled as a gamma distribution (HKY + I + Γ); and for *Xdh* the Hasegawa, Kishino and Yano model with among site rate variation modelled as a gamma distribution (HKY + Γ).

Maximum-likelihood (ML) trees were generated using the program GARLI version 2.0 (Zwickl 2011) with clade support calculated from 1000 bootstrap pseudoreplicates. Individual locus matrices and the combined cp+nDNA matrix were analysed with the best-fit models applied to the respective partitions. Five separate likelihood runs were performed using different random starting seeds and trees, and the tree with best ML score was compared with the parsimony consensus tree. Bootstrap analysis was conducted in GARLI using 1000 pseudoreplicates and the following parameter settings: streefname = random, genthreshfortopoterm = 10000.

Bayesian inference of phylogeny was performed using MrBayes version 3.2.1 (Ronquist *et al.* 2011) using the Metropolis-coupled Markov chain Monte Carlo (MCMCMC) estimation of posterior probability distributions. For analysis of the cpDNA and cp+nDNA matrices, the alignment was divided into partitions corresponding to loci and the best-fit model assigned to each. For ITS and *Xdh*, one representative of three other Oxalidalean families, i.e. *Averrhoa carambola* L. (Oxalidaceae), *Cephalotus follicularis* Labill. (Cephalotaceae) and *Cunonia capensis* L. (Cunoniaceae) (Crayn *et al*. 2006; Morton 2011) were used as outgroups and the trees rooted on *Averrhoa*. The overall evolutionary rate in the partitioned models was allowed to vary across partitions, while the nucleotide substitution models were allowed to be independent from each partition. Only branch length and topology remained linked between partitions. For each analysis, four independent MCMCMC runs were started from different and randomly chosen trees. Four Metropolis-coupled chains with incremental heating were used. After initial pre-runs, the heating parameter was set to 0.05 to improve mixing behaviour of the chains. The following MCMC analyses were

performed for each analysis: *psbA-trnH* spacer, ITS and *Xdh* datasets: 5 million generations with a tree sampled every 1000 generations; *trnL-trnF* region: 10 million generations, tree sampled every 1000 generations; *trnV-ndhC* spacer: 8 million generations, tree sampled every 1000 generations; cpDNA, and cp+nDNA data matrices: 15 million generations, tree sampled every 2000 generations. Sampling and convergence of the runs to a stationary distribution were considered adequate when the standard deviation of split frequencies between the four independent runs dropped below 0.01, indicating that the tree samples from the different runs were sufficiently similar. Additionally, sampling and convergence of the runs were checked independently and in combination in Tracer version 1.5 (Rambaut and Drummond 2007), where an effective sample size (ESS) of all parameter values greater than 200 was considered adequate. The four separate runs were combined and then the trees sampled prior to the four Metropolis-coupled chains reaching stationarity were discarded as burn-in. Bayesian posterior probabilities (PP) were calculated from the consensus of the remaining sampled trees after the burn-in; PP values greater than or equal to 0.95 were considered strong support.

2.2.9 Seed morphology

All Elaeocarpaceae except *Sericolea* and some *Elaeocarpus* have straight embryos (Coode 2004; Fig. 2.3). *Sericolea* species have very weakly curved embryos, *Elaeocarpus holopetalus* F.Muell. has straight embryos but with a hooked tip (an intermediate state between straight and curved embryos; Coode 1984) and some *Elaeocarpus* have obviously curved embryos. In order to encompass the range of variation in curved embryos this character was scored in two states: embryo straight versus not straight.

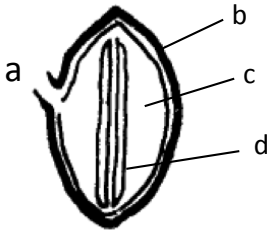


Fig. 2.3 Cross-section of an *Elaeocarpus* seed showing a straight embryo. a, funicle attachment; b, seed coat; c, entire endosperm; d, straight embryo (reproduced with permission from Coode 1984).

The endosperm ornamentation (entire or ruminate, Fig. 2.4) is entire in all Elaeocarpaceae except some *Elaeocarpus*. The shape of the cotyledon (broad or narrow) was not considered suitable for analysis in this study because the interpretation of the shape is ambiguous as it can be strongly influenced by the age of the fruits and the drying process used on herbarium specimens.

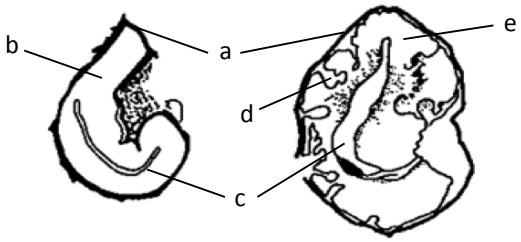


Fig. 2.4 Cross-section of an *Elaeocarpus* seed showing a curved embryo and ruminate endosperm. a, seed coat; b, entire endosperm; c, curved embryo; d, integument; e, ruminate endosperm (reproduced with permission from Coode 1984).

The embryo shape and endosperm condition were traced on the phylogenetic tree to investigate their evolutionary history and usefulness in defining monophyletic groups. Information on the embryo shape and endosperm condition was compiled from the literature or expert opinion (Coode unpublished data) supplemented by up to three observations per species by the author made on herbarium specimens held at CNS and K (Appendix 2.2). Up to three mature fruits were selected from herbarium specimens. Transverse and longitudinal sections were made using a fine-toothed X-

acto X75300 razor saw (Elmer's Products Inc., Westerville, Ohio, USA) and viewed with the unaided eye. The interpretation of states followed Coode (1984). The evolution of the two characters was determined by reconstructing the ancestral states on the combined cp+nDNA Bayesian tree under parsimony and using the all-installed-modules mode implemented in Mesquite version 2.75 (Maddison and Maddison 2011). This method reconstructs the ancestral states at the nodes of the given tree by minimising the number of steps of character state change on the tree. The all-installed-modules mode enables the program to operate through the cooperation of various appropriate modules (determined by the program itself), where each has a different but inter-related function (Mesquite Project Team 1999 – 2014). In addition to these, the following parameter settings were used: reconstruct ancestral states as character history source, stored characters as source of characters (for ancestral state reconstruction) and Fitch parsimony model (all characters weighted equally and treated as independent and unordered).

2.3 Results

A total of 176 samples representing 134 taxa were sequenced for up to four DNA markers. A total of 632 sequences were analysed, 485 of these newly generated in this study and the remainder obtained from GenBank (http://www.ncbi.nlm.nih.gov/genbank/), Baba (2013) or donated by Prof. Ferry Slik (Xishuanbanna Tropical Botanical Garden, China). Full details of the samples studied are provided in Appendix 2.1.

2.3.1 Variability of the DNA markers screened

Of the eight DNA markers screened, *Xdh* showed the highest percentage (14.27 %) of parsimony-informative characters across the dataset. For the *psbA-trnH*, the *trnL-trnF* region, *trnV-ndhC* and ITS datasets the percentage of parsimony-informative characters ranged between 2.80 – 7.88 %. The other markers – *trnQ-rps16*, *ndh*F and ncpGS – showed insufficient variation to be considered further for phylogenetic reconstruction in *Elaeocarpus*. For *trnQ-rps16*, a total of 593 nucleotides were

sequenced of which 0.51 % were potentially parsimony-informative; for *ndhF*, 851 nucleotides were sequenced and 0.24 % were potentially parsimony-informative; and for ncpGS, 1030 nucleotides were sequenced and 0.97 % were potentially parsimony-informative.

Most of the indels observed were found in *psbA-trnH* (84 indels), *trnV-ndhC* (65) and the *trnL-trnF* region (49). In the *Xdh* data matrix, no gaps were required. Indels were not coded for ITS due to the problem of paralogy as described in section 2.3.2. Statistics for the DNA markers are presented in Table 2.3.

Table 2.3 Statistics of six DNA markers – *psbA-trnH* spacer, *trnL-trnF* region, *trnV-ndhC* spacer, *ndhF*, ncpGS and *Xdh* – based on alignments of up to 14 samples representing species of *Aceratium*, *Crinodendron*, *Dubouzetia*, *Elaeocarpus* and *Sericolea*.

Region	Aligned length (bp)	No. potentially parsimony-informative characters (%)
psbA-trnH spacer	1299	52 (4.00)
trnL-trnF region	1285	36 (2.80)
trnV-ndhC spacer	850	67 (7.88)
trnQ-rps16 spacer	593	3 (0.51)
ndhF	851	2 (0.24)
ncpGS	1030	10 (0.97)
Xdh	694	99 (14.27)

2.3.2 Xdh *and ITS*

The sequences produced by the MiSeq platform were generally shorter than those produced by the Sanger method due to the read length limitation of MiSeq and the design of the new primers. This study amplified approximately 690 of the total 4083 nucleotides (Morton 2011) of the *Xdh* region using both Sanger and NGS technologies. In the NGS, the sequencing success rate was 95.8 % and 57.3 % for the first and second fragments of *Xdh*, respectively. The dropout could have occurred

either at the PCR step or at the sequencing step. If it did occur at the latter step, it would likely have been due to unequal pooling of amplicons, which is a common issue with NGS approaches and is difficult to avoid when pooling large numbers of amplicons in one lane on the MiSeq flow cell. Two paralogous gene copies of *Xdh* were identified which differ by 11 non-synonymous codons and 12 synonymous codons. All *Xdh* sequences amplified lacked internal stop codons and thus are possibly functional copies. The Bayesian trees of the two gene copies are generally congruent with the tree generated from the cpDNA data matrix (trees not shown).

For the ITS region approximately 650 nucleotides were sequenced, which comprised the complete ITS-1, 5.8S ribosomal RNA gene and ITS-2, and partial 18S and 26S ribosomal RNA genes. The first and second fragments of ITS have the same NGS sequencing success rate of 97.9 %.

2.3.3 Analysis of individual DNA regions

Analyses of the individual DNA regions each provided low resolution of relationships within the *Elaeocarpus* alliance. Nonetheless, these analyses suggest that *Aceratium* and *Sericolea* are each monophyletic and form a clade with the sampled *Elaeocarpus* species, and that this clade is sister to *Dubouzetia*. Within *Elaeocarpus*, the topology is largely unresolved (*psbA-trnH* Appendix 2.3; *trnL-trnF* region Appendix 2.4; *trnV-ndhC* Appendix 2.5; *Xdh* Appendix 2.6). The statistics for the MP analyses are presented in Table 2.4. ML and Bayesian analyses were also performed for each of the individual DNA regions and the results were similar to those obtained using MP, with the highest support values in the Bayesian analysis, followed by ML (Appendices 2.3 – 2.6).

Table 2.4 Data matrix and tree statistics from the maximum parsimony (MP) analysis (inclusive of outgroups) of the *psbA-trnH* spacer, *trnL-trnF* region, *trnV-ndhC* spacer, *Xdh* gene, combined chloroplast (cpDNA) and combined chloroplast and nuclear (cp+nDNA) data matrices.

Region	No. of aligned positions, excluding ambiguously aligned regions (nucleotides + indels)	No. parsimony-informative characters/% (nucleotides + indels /%)	Consistency index	Retention index	Rescaled consistency index	Tree length	No. missing data (%)
psbA-trnH spacer*	808 (866)	81/10.02 (114/13.16)	0.681	0.869	0.592	301	0 (0)
trnL-trnF region*	1153 (1189)	45/3.90 (60/5.05)	0.803	0.921	0.739	157	0 (0)
trnV-ndhC spacer*	690 (743)	52/7.54 (78/10.50)	0.826	0.943	0.779	172	0 (0)
Xdh gene (fragment 1)**	264 (264)	43/16.28 (nil)	0.835	0.888	0.741	127	0 (0)
Xdh gene (fragment 2)**	339 (339)	47/13.86 (nil)	0.815	0.885	0.722	195	0 (0)
cpDNA data matrix**	3014 (3212)	233/7.73 (246/7.66)	0.718	0.866	0.622	682	32 (3.65)
cp+nDNA data matrix**	3617 (3815)	248/6.86 (320/8.39)	0.715	0.844	0.603	1071	127 (14.50)

* some taxa represented by more than one sample.

** each taxon represented by only one sample.

2.3.4 Analysis of the combined data matrices

In general, the phylogenetic analyses based on the combined cpDNA (Fig. 2.5) and cp+nDNA (Fig. 2.6) data matrices generated similar tree topologies. The latter provided the greatest resolution of relationships between *Elaeocarpus*, *Aceratium* and *Sericolea*, and within *Elaeocarpus*. The MP, ML and Bayesian tree topologies are highly similar and show only minor differences in the degree of support (MP and ML trees not shown).

The final cpDNA data matrix comprised 146 taxa and 3212 characters (246 or 7.66 % of these were potentially parsimony-informative), excluding the ambiguously aligned regions. MP analysis of these data produced 2000 equally most parsimonious trees of 715 steps. The final cp+nDNA data matrix comprised 131 taxa and 3815 characters (320 or 8.39 % of these parsimony-informative), excluding the ambiguously aligned regions. MP analysis of these data produced 368,000 equally most parsimonious trees of 1071 steps (Table 2.4).

In analyses based on the cpDNA data matrix, the *Elaeocarpus* alliance formed a clade (MP bootstrap 100 %, ML bootstrap 98 %, PP 1 – values reported in this order hereinafter, node 1 in Fig 2.5). *Aceratium* (MP bootstrap 98 %, ML bootstrap 98 %, PP 1, node 2) and *Sericolea* (98 %, 92 %, 1.0, node 3) are each monophyletic. All but one of the *Elaeocarpus* samples formed a clade (88 %, 89 %, 1.0, node 4): the position of *E. holopetalus* was unresolved within the *Elaeocarpus* alliance clade. In the analyses based on the cp+nDNA data matrix, the three genera each formed a monophyletic group with strong support (*Sericolea* 100 %, 100 %, 1.0, node 1; *Aceratium* 99 %, 99 %, 1.0, node 2; and *Elaeocarpus* 86 %, 97 %, 1.0, node 3; Fig. 2.6).

Fig. 2.5 Maximum clade credibility tree from Bayesian analysis of the combined plastid regions *psbA-trnH*, *trnL-trnF* and *trnV-ndhC*. Branch support values are to the left of the nodes in the following order: MP bootstrap/ ML bootstrap/ Bayesian posterior probability. Genus abbreviations are *Aceratium* (Ac), *Crinodendron* (C), *Dubouzetia* (D), *Elaeocarpus* (E), *Peripentadenia* (Pe) and *Sericolea* (Se). * Branch collapses in the MP or ML 50 % majority rule consensus tree. The insert shows branch lengths proportional to the amount of nucleotide change and the section of the tree which has been enlarged.

Fig. 2.5 (continued).

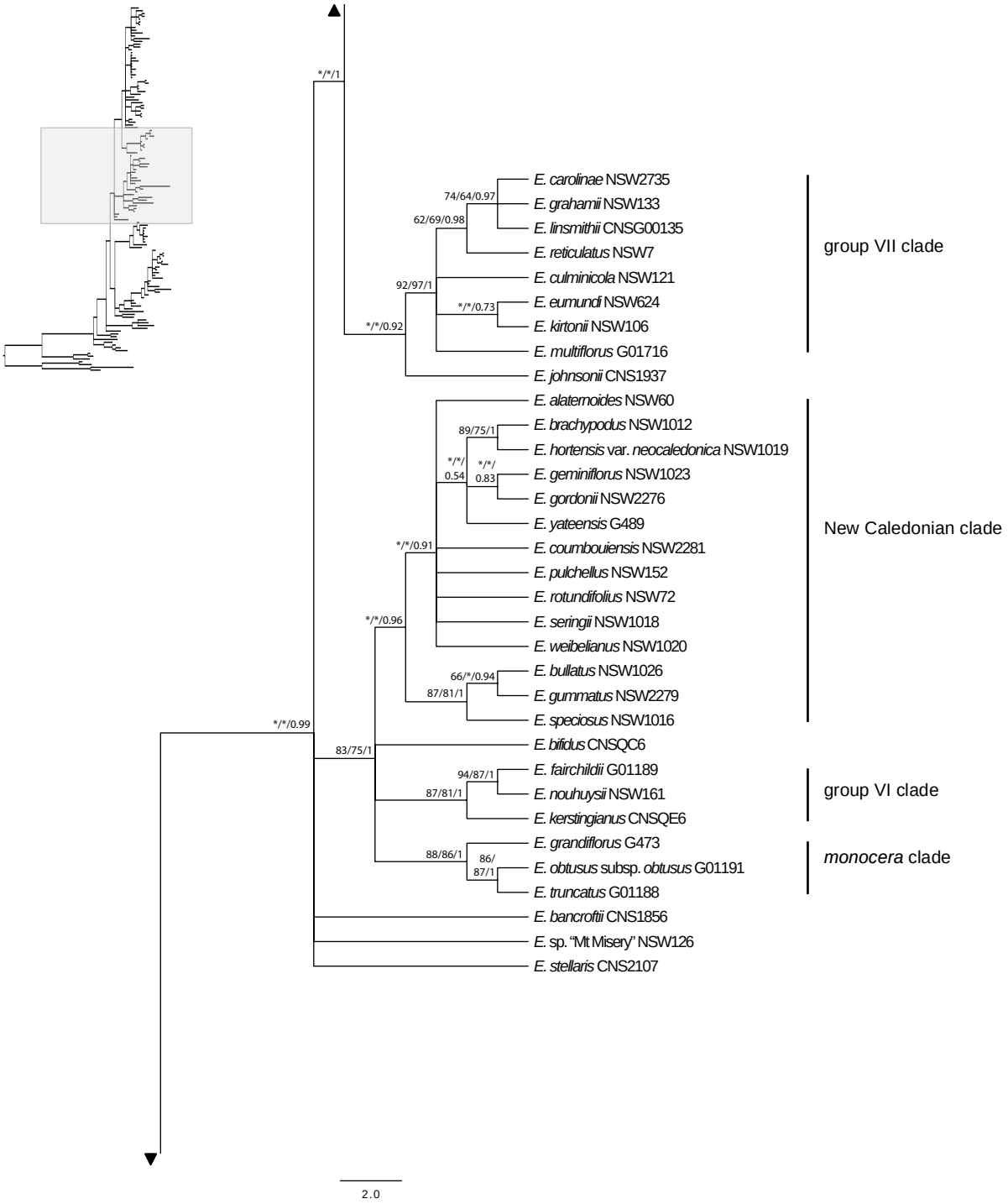


Fig. 2.5 (continued).

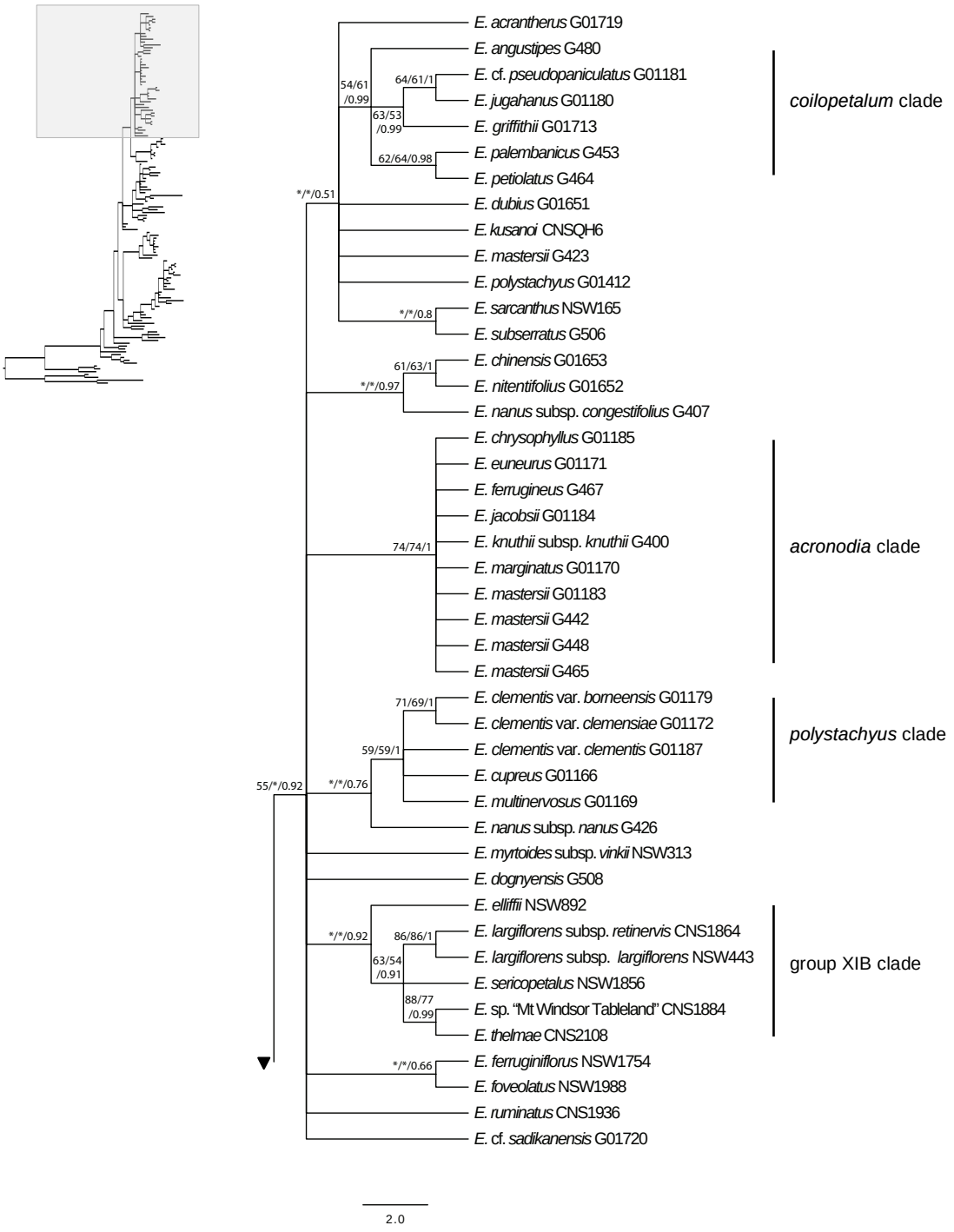


Fig. 2.5 (continued).

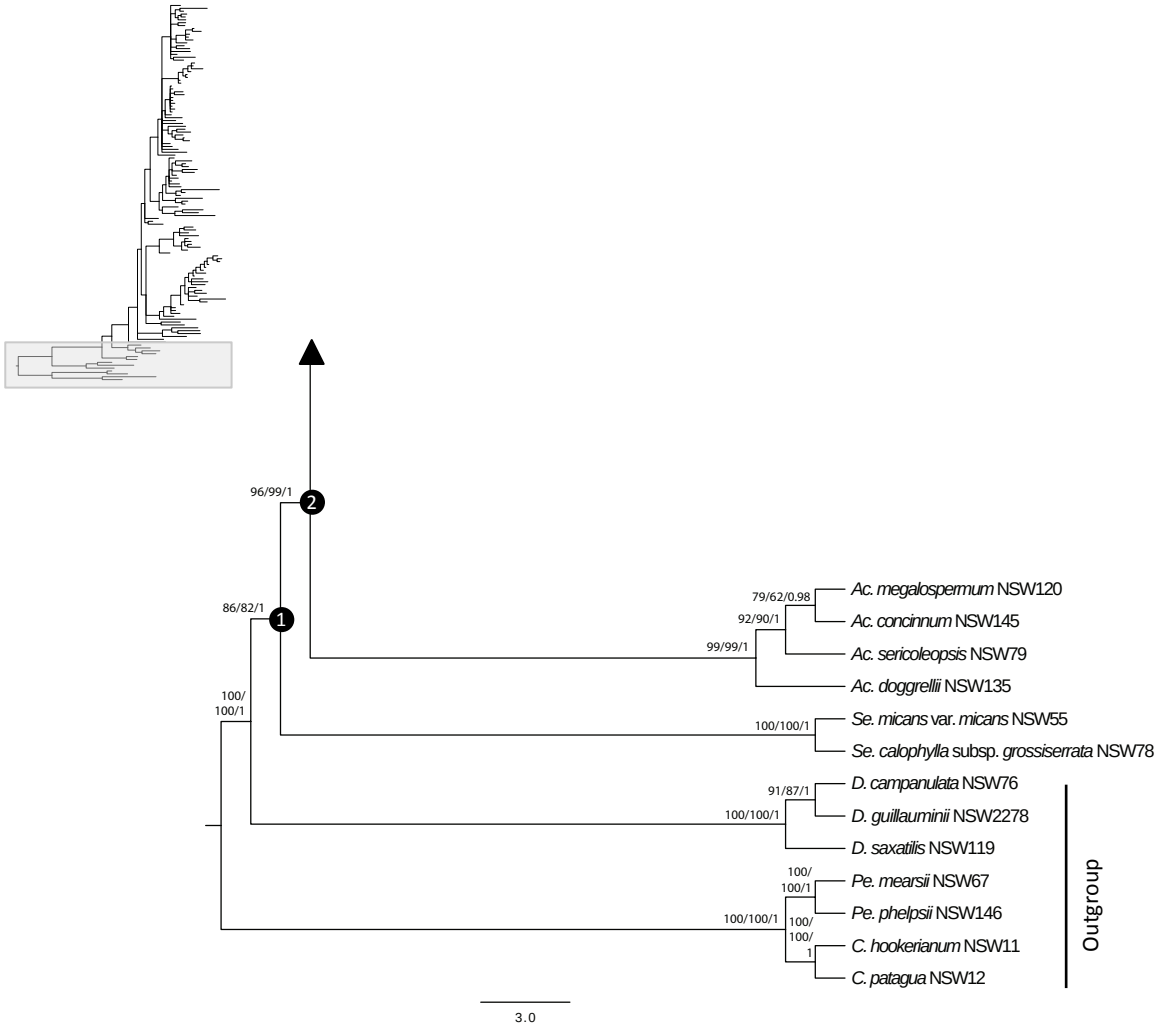


Fig. 2.6 Maximum clade credibility tree from Bayesian analysis of the combined plastid regions *psbA-trnH*, *trnL-trnF* and *trnV-ndhC*, and the low-copy protein coding nuclear gene xanthine dehydrogenase (*Xdh*). Branch support values are to the left of the nodes in the following order: MP bootstrap/ ML bootstrap/ Bayesian posterior probability. Genus abbreviations are *Aceratium* (Ac), *Crinodendron* (C), *Dubouzetia* (D), *Elaeocarpus* (E), *Peripentadenia* (Pe) and *Sericolea* (Se). * Branch collapses in the MP or ML 50 % majority rule consensus tree. The insert shows branch lengths proportional to the amount of nucleotide change and the section of the tree which has been enlarged.

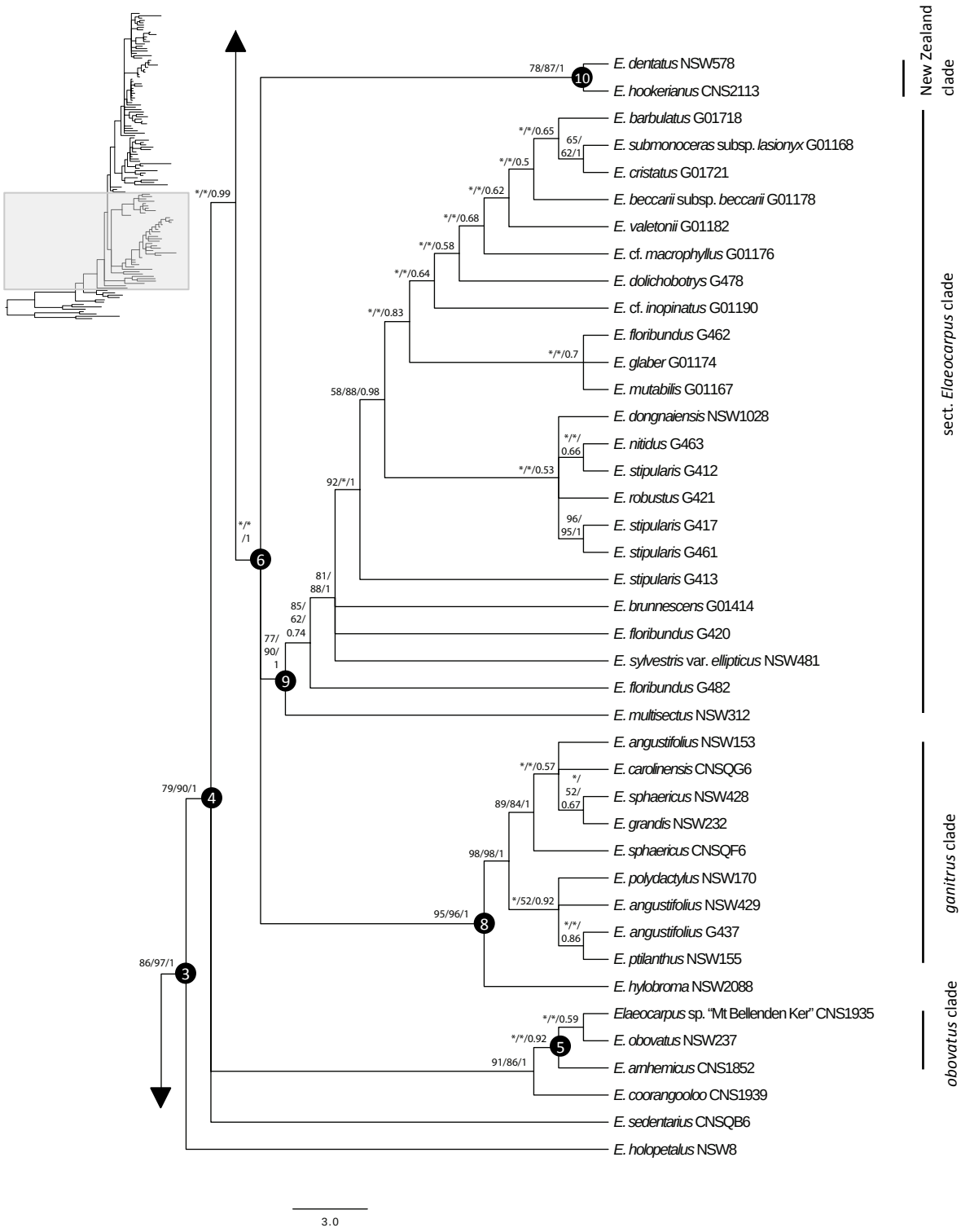


Fig. 2.6 (continued).

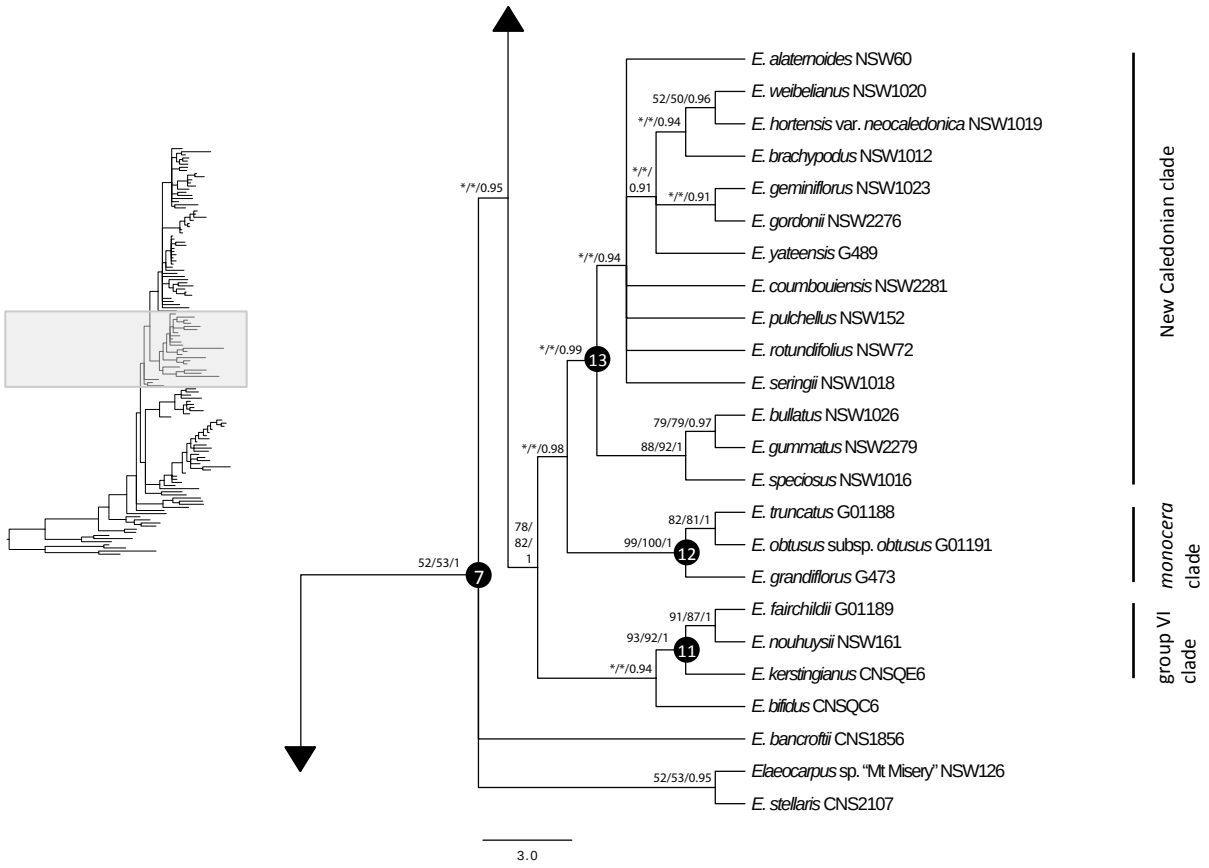


Fig. 2.6 (continued).

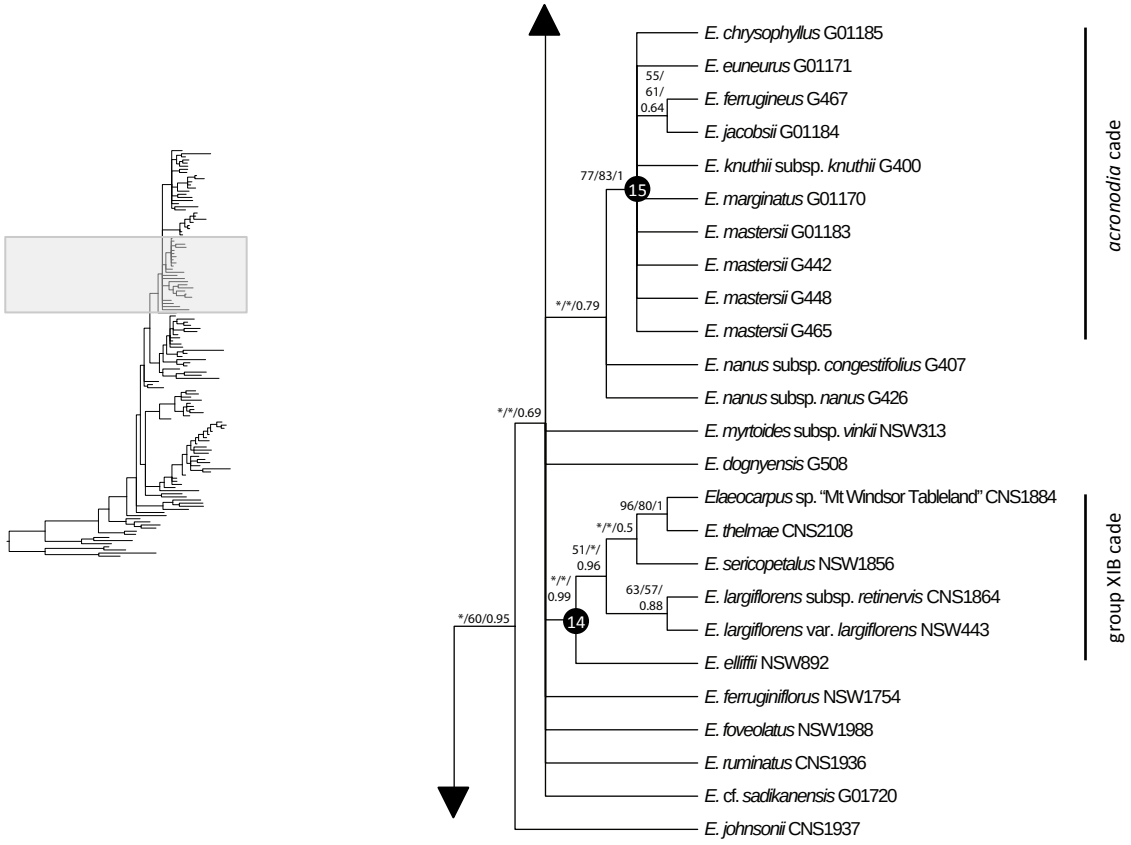


Fig. 2.6 (continued).

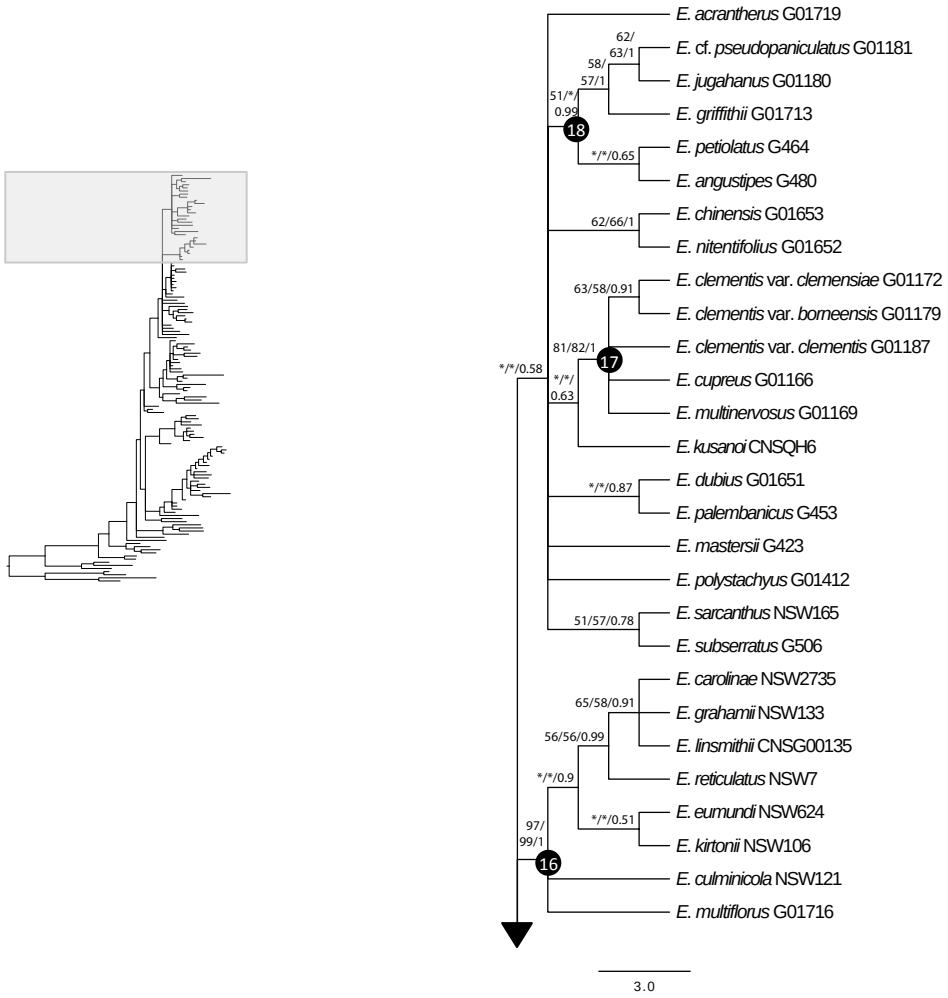


Fig. 2.6 (continued).

Within the *Elaeocarpus* clade, 13 labelled lineages or clades are resolved by the Bayesian analysis of the cp+nDNA data matrix (Fig. 2.6). These are *E. sedentarius*, the *coilopetalum*, *polystachyus*, group VII, *acronodia*, group XI subgroup B (XIB), New Caledonian, *monocera*, group VI, *ganitrus*, section (sect.) *Elaeocarpus*, *obovatus* and New Zealand clades. All of these clades are strongly supported in the Bayesian analysis except for the *obovatus* clade (< 50%, < 50%, 0.92, node 5). In the MP and ML analyses, the tree topologies are largely congruent with the Bayesian tree, the only differences being in the support values. In the MP analysis, all labelled clades received > 50 % bootstrap support except the *obovatus* (node 5), the New Caledonian (node 13), and the group XIB (node 14) clades. In the ML analysis, other than the three labelled clades as in the MP analysis, the *coilopetalum* (node 18) clade also received < 50 % bootstrap support, while the rest of the labelled clades received > 50 % bootstrap support.

Elaeocarpus holopetalus is placed with strong support (86 %, 97 %, 1.0, node 3 in Fig. 2.6) as sister to a clade comprised of the remainder of *Elaeocarpus*. The basal node within this clade is a trichotomy (79 %, 90 %, 1.0, node 4) consisting of *E. sedentarius*, *E. coorangooloo* J.F.Bailey & C.T.White and the *obovatus* clade (< 50 %, < 50 %, 0.92, node 5), and the remainder of *Elaeocarpus*. Within this last clade a total of 12 labelled subclades are recognised within two main lineages. The first of these lineages (< 50 %, < 50 %, 1.0, node 6) comprises three major labelled clades, i.e. the New Zealand (78 %, 87 %, 1.0, node 10), sect. *Elaeocarpus* (77 %, 90 %, 1.0, node 9) and *ganitrus* clades (95 %, 96 %, 1.0, node 8). The second (52 %, 53 %, 1.0, node 7) is larger and comprises eight major labelled clades: the group VI (93 %, 92 %, 1.0, node 11), *monocera* (99 %, 100 %, 1.0, node 12), New Caledonian (< 50 %, < 50 %, 0.99, node 13), group XIB (< 50 %, < 50 %, 0.99, node 14), *acronodia* (77 %, 83 %, 1.0, node 15), group VII (97 %, 99 %, 1.0, node 16), *polystachyus* (81 %, 82 %, 1.0, node 17) and *coilopetalum* (51 %, < 50 %, 0.99, node 18) clades. The relationships of the remaining species are unresolved.

2.3.5 Seed morphology transformation

The most parsimonious reconstruction of the embryo shape and endosperm condition within *Elaeocarpus* required 5 and 13 steps on the phylogeny, respectively (Figs. 2.7, 2.8). The following taxa included in this study have curved embryos: *Sericolea*, *E. holopetalus*, the group VII (including *E. multiflorus* (Turcz.) Fern.-Vill.), the group XIB, *acronodia*, *polystachyus* and *coilopetalum* clades, as well as 16 of the taxa whose phylogenetic relationships are not fully resolved, namely *E. acrantherus* Merr., *E. chinensis* Hook.f. ex Benth., *E. dubius* A.DC., *E. ferruginiflorus* C.T.White, *E. foveolatus* F.Muell., *E. nanus* Corner (including its two subspecies), *E. nitentifolius* Merr. & Chun, *E. palembanicus* (Miq.) Corner, *E. polystachyus* Wall. ex Müll. Berol., *E. ruminatus* F.Muell., *E. sarcanthus* Schltr., *E. subserratus* Baker, *Elaeocarpus* sp. "Mt Windsor Tableland" *sensu* Jessup & GJM 1378 and possibly *E. kusanoi* Koidz. and *E.* cf. *sadikanensis* R.Knuth. The remaining taxa have straight embryos. The parsimony reconstruction indicates that the curved embryos in *Sericolea*, *E. holopetalus* and the remaining *Elaeocarpus* are not homologous, but that the condition shared by members *Elaeocarpus* may have a single origin, with several reversals to the straight condition. Alternatively, two independent origins in sister lineages may have occurred, an equally parsimonious reconstruction.

Ruminate endosperm is found in the group XIB, *acronodia* and *polystachyus* clades, *E. linsmithii*, possibly *E. multiflorus*, *E. petiolatus*, and several taxa whose relationships are incompletely resolved, namely *E. ferruginiflorus*, *E. foveolatus*, *E. nanus* (including its two subspecies), *E. nitentifolius*, *E. polystachyus*, *E. ruminatus*, *Elaeocarpus* sp. "Mt Windsor Tableland" and possibly *E. kusanoi* (Appendix 2.2). The parsimony reconstruction indicates that ruminate endosperm has a complex evolutionary history within *Elaeocarpus* (Fig. 2.8).

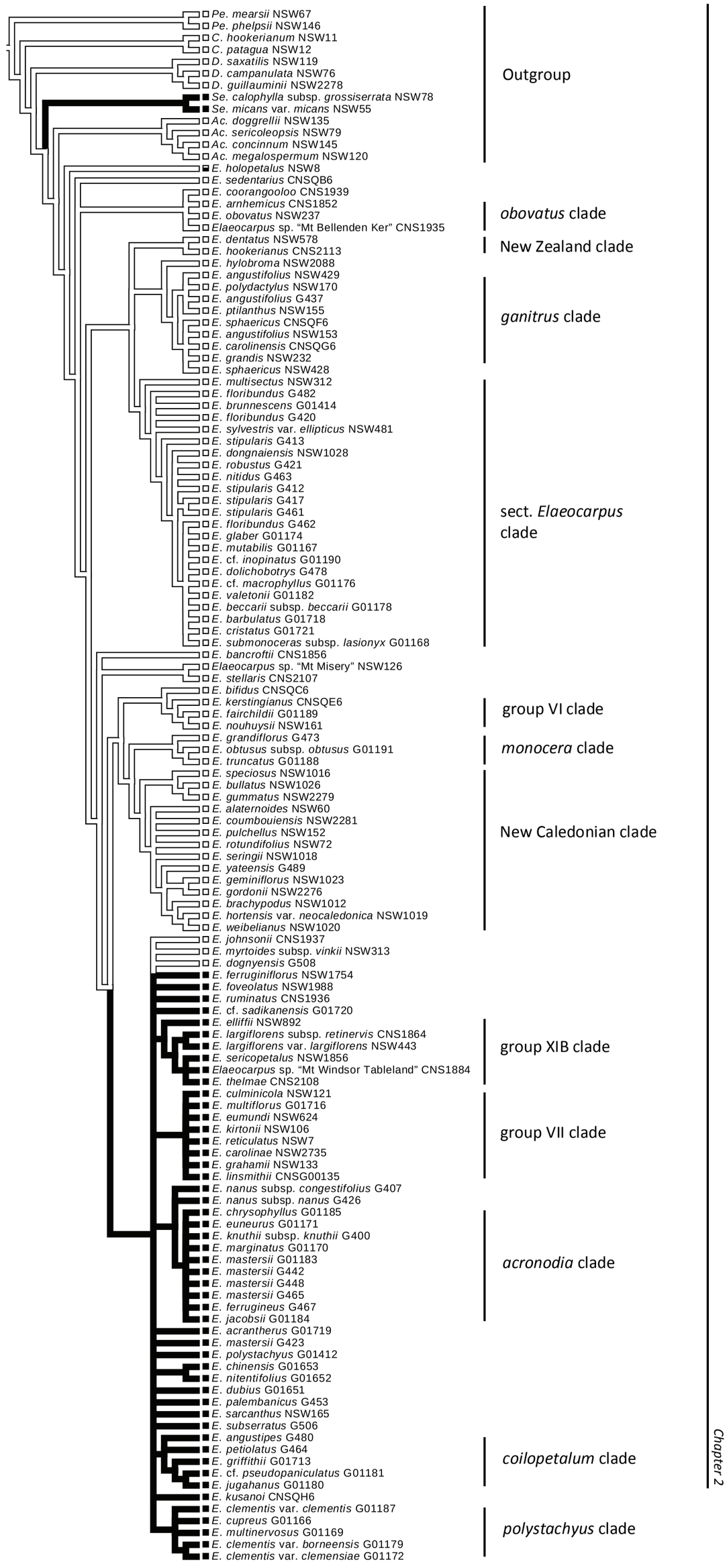


Fig. 2.7 Transformation of the shape of embryo reconstructed using Fitch parsimony and traced on the Bayesian phylogenetic tree of *Elaeocarpus* based on the combined plastid regions *psbA-trnH, trnL-trnF* and *trnV-ndhC*, and the low-copy protein coding nuclear gene – xanthine dehydrogenase (*Xdh*). White branch represents lineage(s) with straight embryo; black branch represents lineage(s) with curved embryo.

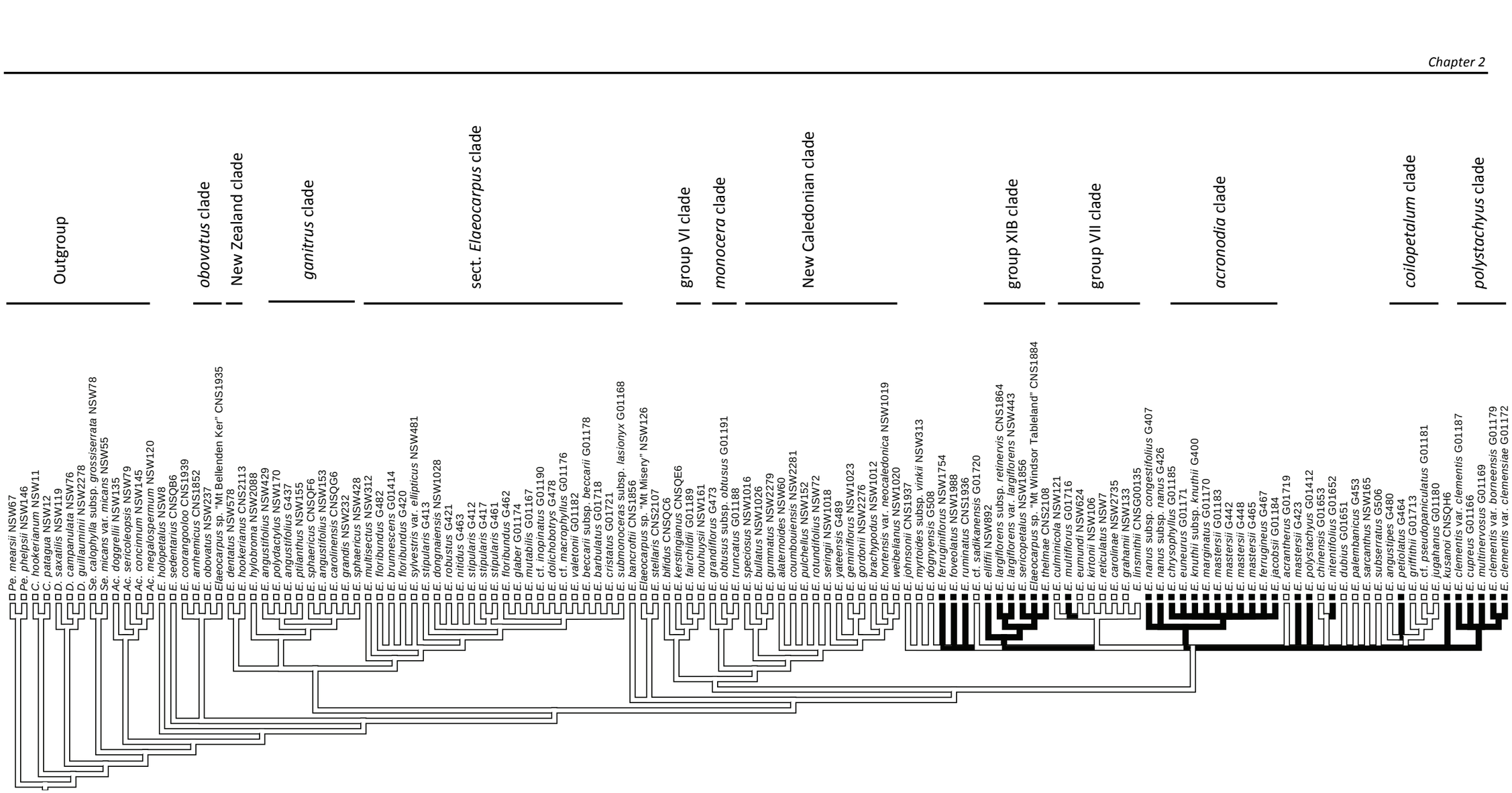


Fig. 2.8 Transformation of the ornamentation of endosperm reconstructed using Fitch parsimony and traced on the Bayesian phylogenetic tree of *Elaeocarpus* based on the combined plastid regions *psbA-trnH*, *trnL-trnF* and *trnV-ndhC*, and the low-copy protein coding nuclear gene – xanthine dehydrogenase (*Xdh*). White branch represents lineage(s) with entire endosperm; black branch represents lineage(s) with ruminate endosperm.

2.4 Discussion

2.4.1 *Phylogenetic utility of* psbA-trnH *and* Xdh *and paralogy of the nuclear DNA markers*

In previous molecular studies, the cpDNA (*trnL-trnF* region and *trnV-ndhC* intergenic spacer) and the nuclear ribosomal (ITS) regions exhibited relatively low sequence divergence within the *Elaeocarpus* alliance (Maynard 2004; Crayn *et al*. 2006; Baba 2013), leaving the phylogenetic relationships and the monophyly of the genus *Elaeocarpus* unresolved. Additionally, Baba (2013) explored three more plastid markers (*matK*, *trnQ-rpl16* intergenic spacer and *rpl32-trnL* intergenic spacer) in her pilot study, however none of them showed sufficient variation to be informative for phylogenetic reconstruction in *Elaeocarpus*. The two DNA markers included here for the first time in a phylogenetic study of *Elaeocarpus* – *psbA-trnH* (plastid gene) and *Xdh* (the low-copy protein coding nuclear gene) – exhibit a higher level of variation than the three previously used regions. When analysed in combination with the plastid markers, these markers provide a greatly improved estimate of the phylogenetic relationships within the *Elaeocarpus* alliance, as well as the main lineages with the genus *Elaeocarpus*.

In most angiosperms only one copy of *Xdh* has been detected. In the cases where two or more copies of the gene were observed, they ranged from 91 % to 99.8 % similarity (Hesberg *et al*. 2004; Morton 2011). In this study, two paralogous copies were found in *Elaeocarpus* and other Elaeocarpaceae (i.e. *Aceratium*, *Crinodendron*, *Dubouzetia* and *Sloanea*), which differ in 11 non-synonymous codons and 12 synonymous codons. Independent Bayesian analysis on the two copies showed that the results were congruent with the phylogenetic trees generated from the cpDNA data, indicating the paralogs have undergone a similar evolutionary history.

Up to four copies of ITS per sample were identified. Which of these copies was functional and orthologous across samples could not be determined with confidence because obvious nucleotide substitutions occurred in the conserved motif (Harpke and

Peterson 2008) of the 5.8S ribosomal RNA gene in all copies, in some cases in different samples of the same species. This phenomenon could be due to *in vitro* selection against amplification of the functional copy(ies).

2.4.2 Utility of Next-generation sequencing mixed templates

DNA sequencing technology is developing rapidly, with many researchers now moving from traditional Sanger sequencing to the cutting edge Next-Generation sequencing (NGS) technologies. Griffin *et al*. (2011) demonstrated that certain NGS approaches are much more effective than Sanger methods for sequencing mixtures of templates or amplicons, allowing different alleles and paralogous gene copies to be identified. At the same time, advances in bioinformatic analytical tools, such as Galaxy (https://usegalaxy.org/) and QIIME (http://qiime.org/), provide effective data processing workflows for the high-throughput data generated from the NGS. Additionally, the MiSeq platform used in this study generates large volumes of data per run (about 15 Gb for 2 regions) at much lower cost per base than the Sanger method.

The 16S metagenomic sequencing library preparation method (Illumina Inc. 2013) was generally successful in this study. This method uses a two-step PCR approach. The first PCR step (i.e. amplicon PCR) amplifies templates out of a DNA sample using fusion of specific primers with overhanging adapters attached. In the second step of PCR (i.e. index PCR), amplicons are barcoded using two indices. The barcoded amplicons are then pooled in a lane on the flow cell of the sequencing machine and sequences from different amplicons are separable based on their unique indices. High quality sequences from various samples were all identifiable in this study. Therefore, the dual-indexing method proved to be a reliable barcoding technique for multiplexing amplicons, although there were differences in the amplicon fragment size and the unavoidable variance in read number associated with low accuracy of quantification and pipetting when pooling amplicons.

2.4.3 *Phylogenetic relationships within the* Elaeocarpus *alliance*

Two major challenges faced in previous studies of the phylogenetic relationships between and within *Elaeocarpus* were inadequate sampling of this highly diverse genus and low sequence variability of the DNA markers used (Maynard 2004; Baba 2013). These studies were mainly focused on the phylogeny of the Australian species, with few representatives from other regions. Furthermore, the selected DNA markers (*trnL-trnF* region, *trnV-ndhC* and ITS) exhibited low sequence variability leading to insufficiently resolved phylogenies, particularly with respect to the phylogenetic relationships within the *Elaeocarpus* alliance (Crayn *et al*. 2006; Baba 2013). Furthermore, few clades within *Elaeocarpus* were resolved. Based on a much-expanded taxonomic sampling and two additional new DNA regions (*psbA-trnH* and *Xdh*), the results of the phylogenetic analyses in this study consistently indicate that *Elaeocarpus*, *Aceratium* and *Sericolea* are closely related, strongly confirming the preliminary results of the earlier studies (Maynard 2004; Crayn *et al*. 2006; Baba 2013).

The inclusion of both *psbA-trnH* and *Xdh* sequence data in the phylogenetic reconstruction of *Elaeocarpus* greatly improves resolution and provides strong evidence that the three genera, i.e. *Elaeocarpus*, *Aceratium* and *Sericolea*, are each monophyletic. The sister group relationship between *Elaeocarpus* and *Aceratium* is resolved for the first time in this study.

2.4.4 *Phylogenetic relationships within* Elaeocarpus

Within the *Elaeocarpus* clade, taxon sampling (especially the number of representatives from different infrageneric groups and biogeographic regions) and the resolution of relationships have greatly improved compared to previous molecular studies (Maynard 2004; Crayn *et al*. 2006; Baba 2013). Shallow relationships within many labelled clades, however, remain uncertain.

The statistical support for some nodes (nodes 6, 7, 10, 9, 13, 14 and 18, Fig. 2.6) received high posterior probabilities (at least 0.95) but low (less than 80 %) bootstrap support in the MP and/or ML analyses. The Markov Chain Monte Carlo simulation in the parametric Bayesian inference is assessing the probability of a clade to be true (Huelsenbeck *et al*. 2002; Huelsenbeck and Rannala 2004; Yang 2006) by sampling highly similar trees from the region of the tree space, and specific clades are more likely to have higher frequencies in the consensus, regardless of the amount of evidence (i.e. characters) in the data matrix supporting them (Alfaro *et al*. 2003; Hillis *et al*. 2005; Garcia-Sandavol 2014). On the contrary, the non-parametric bootstrap in the MP and ML analyses indirectly measures the amount of evidence in the data matrix that supports a clade (Goloboff *et al*. 2003) and will only assign a high frequency when there is a large amount of evidence in the matrix to support a clade and vice versa (Berbee *et al*. 2000; Garcia-Sandoval 2014). Therefore, a Bayesian clade is likely to be true, even when there is little evidence supporting it, but in this case, the bootstrap value will not be high.

The phylogeny estimated in this study based on molecular data is broadly congruent with Coode's system (1978, 1996a – c, 2001a, b, d – f, 2003, 2010; Coode and Weibel 1994). The *ganitrus*, *acronodia*, *coilopetalum and polystachyus* groups, group VII, group XIB, and sect. *Elaeocarpus* are each monophyletic and can be diagnosed based on the morphological characters described in Coode's system (diagnostic morphologies are discussed under each clade). On the other hand, group V subgroup D (group VD) is paraphyletic and the phylogenetic placement of some species is equivocal, namely: *E. acrantherus* Merr., *E. bancroftii* F.Muell., *E. bifidus* Hook. & Arn., *E. myrtoides* A.C.Sm. subsp. *vinkii* Coode R.Knuth, *E. dognyensis* Guillaumin, *E. dubius*, *E. ferruginiflorus*, *E. foveolatus*, *E. kusanoi*, *E. nanus* subsp. *nanus*, *E. polystachyus*, *E. ruminatus*, *E.* cf. *sadikanensis*, *E. sarcanthus*, *E. sedentarius*, *E. stellaris* L.S.Sm., *E. subserratus* and the putative new species *Elaeocarpus* sp. Mt Misery (Webb & Tracey 10905).

Elaeocarpus holopetalus

Elaeocarpus holopetalus represents a distinct lineage sister to the rest of the *Elaeocarpus* species examined here (node 3, Fig 2.6). It is endemic to Victoria and New South Wales, Australia and belongs to the monotypic group X (Coode 1984). The species is found in cool temperate upland rainforests (up to 1500 m altitude) often growing in association with Gondwanan relicts like *Nothofagus* (Coode 1984); whereas other *Elaeocarpus* occurring in southern Australia are either wet sclerophyll, warm temperate or subtropical rainforest species. This restriction to cool temperate rainforest is consistent with an origin for the lineage under the cool and wet conditions that characterised Gondwana. Its differentiation from the rest of the *Elaeocarpus* species is consistent with its possession of the following combination of atypical morphological characters: the petal apex is undivided (versus weakly to highly divided in all other species), the fruits are dark purple or black (versus dull green-blue to bright blue), and the embryo is straight but with a hooked tip, which resembles an intermediate state between the two other embryo types – straight and curved (Weibel 1968, Coode 1984; transformations of embryo shape within the genus is discussed in section 2.4.5).

The remainder of *Elaeocarpus*

Among the remaining *Elaeocarpus*, the basal node (node 4, Fig. 2.6) is a polytomy of *Elaeocarpus sedentarius*, the *E. coorangooloo* + *obovatus* clade and the rest of the *Elaeocarpus*. *Elaeocarpus sedentarius* and the *E. coorangooloo* + *obovatus* clade are early diverging lineages, although relationships between them, and with the rest of the labelled *Elaeocarpus* clades needs to be clarified by additional molecular evidence. The remainder of *Elaeocarpus* is further divided into two main lineages: the first (node 6, Fig. 2.6) comprises the New Zealand, sect. *Elaeocarpus*, and *ganitrus* clades; the second (node 7, Fig. 2.6) comprises the group VI, *monocera*, New Caledonian, group VII, group XIB, *acronodia*, *polystachyus*, and *coilopetalum* clades.

- ***Elaeocarpus sedentarius***

Elaeocarpus sedentarius is a rare species endemic to South-eastern Australia (Maynard *et al*. 2008). It is morphologically distinct from all other Australian *Elaeocarpus* in having leaves that are strongly discolorous (very pale green to glaucous beneath) on the abaxial surface, fruits which are often triangular in cross section, and the outer endocarp covered with a "pseudo-fleshy" layer consisting of dense, radial fibres (Maynard *et al*. 2008). Nonetheless, *E. sedentarius* morphologically closely resembles *Elaeocarpus blepharoceras* Schltr. from New Guinea, particularly in having petal apices with numerous divisions, a "pseudo-fleshy" outer endocarp, and outer bark that is longitudinally furrowed. However, the two species can be differentiated by the size of their fruits, sepals, petals and the anther awns, the indumentum on the sepal and anther awns, and the venation of the leaf margins. Morphological comparisons between the two species were fully described in Maynard *et al*. (2008).

- ***Elaeocarpus coorangooloo* and the *obovatus* clade**

Elaeocarpus coorangooloo J.F.Bailey & C.T.White and the *obovatus* clade are endemic in Australasia. The *obovatus* clade consists of three species, *E. arnhemicus* F.Muell., *E. obovatus* G.Don and the putative new species, *Elaeocarpus* sp. "Mt Bellenden Ker" *sensu* Brass 18336. This clade is part of group V subgroup D (Coode 1984), which is also known as group *fissipetalum* in Malesia (Coode 1978), and comprises about 21 taxa including one putative new taxon (Coode 1984, unpublished data). *Elaeocarpus coorangooloo* from group VI subgroup E (Coode 1984), is shown to be sister to the *obovatus* clade in this study, where all of these four species formed a robust clade (MP bootstrap 91 %, ML bootstrap 86 %, PP 1; Fig. 2.6) (details discussed under the group VI and *monocera* clades).

- **New Zealand clade**

The New Zealand clade consists of *E. dentatus* (J.R. & G.Forst.) Vahl and *E. hookerianus* Raoul. They are both endemic to New Zealand and are the only two *Elaeocarpus* species found there.

The New Zealand clade and the Australian *obovatus* clade are treated as group VD in Coode's system (1984). However, the present study indicates that the group VD is paraphyletic, where the Australian and New Zealand species formed two clades. Morphologically, the New Zealand species differ from the Australian ones in having heterophyllous leaves in the juvenile stage (versus not heterophyllous) and longer petals with fewer divisions at the apex (3 – 6 versus 4 – 16) (Coode 1984). Therefore, segregating the New Zealand and the *obovatus* clades into two groups would better reflect both the morphological dissimilarity and the phylogenetic relationships. Additionally, *E. hookerianus* is hypothesised to be closely related to the Papuasian group V subgroup C (Coode 1984). *Elaeocarpus bakaianus* Coode was the only representative of subgroup C available for this study – while DNA amplification was successful for the *psbA-trnH* region, it was not for the other three regions studied and the resolution of the phylogenetic analyses based on the *psbA-trnH* data matrix alone with respect to the relationships of this taxon was very low (Appendix 2.3). Further research is needed to identify the phylogenetic relationships of this species.

- **sect. *Elaeocarpus* clade**

Section *Elaeocarpus* (= group III of Coode 1978) consists of about 102 taxa (including infraspecific taxa) and is distributed throughout the Malesian region extending to continental Asia (Weibel 1968; Coode unpublished data). The present study included 15 taxa, which formed a well-supported clade (node 9, Fig 2.6). The clade can be identified morphologically based on the combination of the following characters: flowers which are widely open when mature (for the West and Central Malesian species, Coode pers. comm. 2014), petals which are small to medium-sized and with a much-divided apex, disk which is either with discrete lobes or lobes that are

contiguous or fused, ovaries which have 3 (– 5) loculi with two ovules per loculus, mature fruits having a flattened D-shape in cross section due to usually only one fertile loculus developing, and fruit stones which are woody and not flattened and contain one seed with a straight embryo and entire endosperm (Coode 1994, 1996a). It is noteworthy that *E. multisectus* from New Guinea has flowers that open widely at anthesis which is a character for the western Malesian species of this group (Coode pers. comm. 2014).

- ***ganitrus* clade**

The *ganitrus* group, which is also known as group V subgroup A in Papuasia (Coode 1978) and Australasia (Coode 1984), consists of approximately 19 taxa (including infraspecific taxa). The centre of diversity is in the Malesian region with some species extending to Australasia in the south and continental Asia in the north (Weibel 1968; Coode unpublished data). This group is represented by two species in this study, *E. angustifolius* and *E. ptilanthus* Schltr. *Elaeocarpus angustifolius* is a complex, morphologically highly variable across its range and widely distributed throughout Malesia and extends to continental Asia, Australia and Pacific islands (Coode 2010). This study included also two variants that are treated as separate entity here, the Australian *E. grandis* F.Muell. and the widespread *E. sphaericus* K.Schum.

The *ganitrus* clade (node 8, Fig. 2.6) retrieved in the phylogenetic analyses here consist of five taxa, i.e. *E. angustifolius*, *E. grandis, E. sphaericus*, *E. ptilanthus, E. polydactylus* Schltr. and *E. carolinensis* Koidz. The first three taxa are part of the *E. angustifolius* complex treated in Coode (1984, 2010). The next two are New Guinean taxa: *E. ptilanthus* is currently placed in group V (= *ganitrus* group, Coode 2010); whereas *E. polydactylus* is classified in subgroup D in the current infrageneric systems (Coode 1980), and the molecular results show that this taxon is more closely related to species of subgroup A, further evidence that subgroup D (including the *obovatus* and New Zealand clades as discussed above) is paraphyletic. The last taxon, *E. carolinensis* is found on the Caroline Islands and is unplaced in the current infrageneric systems. Notwithstanding, the results of the phylogenetic analyses in this study show that *E.*

ptilanthus, *E. polydactylus* and *E. carolinensis* could possibly be part of the *E. angustifolius* complex, of which the molecular results do not contradict with the morphological evidence. In previous taxonomic treatments, Weibel (in Coode 2010) treated *E. ptilanthus* as a subspecies of *E. sphaericus* (a variant of *E. angustifolius*); while a mountain form of *E. ptilanthus* is placed in group V subgroup A, the same infrageneric group where *E. angustifolius* is assigned and the only differences between two taxa are on their stone size, shape and surface structure, and their mesocarp (Coode 2010). On the other hand, Coode (2010) noted that a montane specimen of *E. polydactylus* is indistinguishable from the widespread *E. angustifolius*. *Elaeocarpus carolinensis* is not well investigated so far, but Coode (2010) suggested to include this taxon in the future taxonomic investigation of the *E. angustifolius* complex (Coode 2010). Therefore, the complex could possibly be larger than what is included in the current treatment (Coode 2010), a question that can be best addressed through thorough sampling of this complex from various biogeographic regions.

Elaeocarpus hylobroma from Australia is placed in group V in the current infrageneric systems, but unassigned at the subgroup level (Baba and Crayn 2012). The species exhibits morphological similarities with *E. carolinae* and *E. coorangooloo* from Australia and *E. tariensis* Weibel from New Guinea. Previous molecular evidence indicated the species is related with members of group V, which included the widespread *E. angustifolius*, the Australian *E. grandis*, the New Guinean *E. polydactylus*, *E. burderi* Coode, *E. dolichostylus* Schltr. and *E. ptilanthus*, although resolution at shallow branches were low (Baba and Crayn 2012; Baba 2013). The results of the molecular phylogenetic of this study are similar to Baba and Crayn (2012) and Baba (2013), where *E. hylobroma* is closely related to the *ganitrus* clade retrieved here, but resolution with the clade is low. Therefore, until a more well-resolved phylogenetic evidence is presented, this study follows Baba and Crayn (2012) and Baba (2013) by placing *E. hylobroma* in group V, but not assigned to subgroup.

The *ganitrus* clade can be identified based on the combination of the following morphological characters: disk weakly developed (slightly inconspicuous), stamens 11 – 80, anthers without awns but with a seta at tip, ovary with 3 – 9 loculi each with 4 –

6 ovules, fruit stone surface variously rugose, usually more than one seed developing, embryo straight, and endosperm entire (Coode 1978, 2010; Baba and Crayn 2012). The inclusion of *E. polydactylus* in the *ganitrus* clade required amendment of the morphological diagnosis as follows: minimum number of loculi 3 (previously 4) (Coode 1978) and the fruit stone surface may be rugose (Coode 2010). Morphology of *E. carolinensis* is not well studied so far.

- **group VI and *monocera* clades**

Group VI consists of approximately 64 taxa (including taxa below species level) primarily distributed within the Malesian region, with a few species extending to continental Asia (Coode unpublished data). The *monocera* group, which is more commonly known in West and Central Malesia, is part of group VI (Coode 1978, 1998, 2001f). This study examined 11 representatives from group VI: *E. bancroftii* (subgroup B), *E. coorangooloo* (subgroup E), *E. cristatus* Coode (subgroup not assigned), *E. fairchildii* Merr. (subgroup D), *E. grandiflorus* Sm. (subgroup not assigned), *E. kerstingianus* Schltr. (subgroup D), *E. nouhuysii* Koord. (subgroup C), *E. obtusus* subsp. *obtusus* (subgroup not assigned), *E. stellaris* L.S.Sm. (subgroup B), *E. truncatus* Weibel (subgroup not assigned) and *E. williamsianus* Guymer (subgroup F). Two well supported clades are formed. The first belongs to the group VI clade and comprises species from East Malesia and areas eastwards, namely *E. fairchildii*, *E. kerstingianus* and *E. nouhuysii* (node 11, Fig. 2.6). The second belongs to the *monocera* clade and comprises species from West Malesia, namely *E. grandiflorus*, *E. obtusus* subsp. *obtusus* and *E. truncatus* (node 12, Fig. 2.6). Members of the group VI clade can be identified morphologically based on the following combination of characters: anthers awned, ovary usually with 2 loculi each containing 6 – 12 ovules, embryo straight and endosperm entire (Coode 1978, unpublished data; Weibel 1968). Members of the *monocera* clade can be identified by: ovary usually with 2 loculi each containing 6 – 10 ovules, embryo straight, and endosperm entire (Coode 2014, unpublished data).

Elaeocarpus cristatus is not part of the *monocera* clade, instead it is nested within the *Elaeocarpus* clade (node 9). Only one flowering specimen of *E. cristatus* was

available for this study (from Borneo), and DNA extraction and PCR amplification were repeated and the same results were obtained. The species is morphologically a better fit in the *monocera* clade than the sect. *Elaeocarpus* clade. Hence, a contradiction between molecular and morphological evidence is observed here. Additional specimens of this species should be collected to confirm the results of this study before a conclusion can be reached regarding its phylogenetic relationships, as well as the monophyly of the *monocera* group.

Elaeocarpus coorangooloo is morphologically similar to the Papuasian subgroup E (*E. meigei* Weibel and *E. hartleyi* Weibel) in having anthers without awns and ovary usually glabrous (Coode 1978). In spite of that, the morphology of *E. coorangooloo* does bear close resemblance to species of the *obovatus* clade, particularly the terminal bud of the shoot without resin, leaves serrate at the margins, stones short (less than 13 mm) with walls thick, tough and with a rugose surface (Coode 1984). The results of this study support the sister relationship between *E. coorangooloo* and the *obovatus* clade; whereas the phylogenetic relationships between *E. coorangooloo*, *E. meigei* and *E. hartleyi* need further investigation by including the two latter species in an expanded phylogenetic analysis. Furthermore, the results of this study also show that the inclusion of *E. coorangooloo* in group VI is not natural and the species should be classified in group V.

The phylogenetic relationships of the other three members of group VI, *E. bancroftii*, *E. stellaris* (subgroup B) and *E. williamsianus* (subgroup F) are unresolved. The first two species are situated in the same big clade (node 7) as the group VI clade; additionally, *E. stellaris* formed a weakly supported clade with the putative new species from Australia, *Elaeocarpus* sp. "Mt Misery" *sensu* Webb & Tracey 10905 (52 %, 53 %, 0.95) suggesting that this entity belongs to group VI. Morphologically it resembles *E. stellaris*, differing mainly in the larger flowers and smaller leaves and fruit. However, because only *trnL-trnF* sequence data was available for the latter this result should be tested against an expanded dataset.

The phylogenetic relationships between the subgroups of group VI are generally not well addressed in this study as the sampling size for each subgroup is small. Despite that, the findings show that group VI, including the *monocera* group, belongs to one of the main lineages within the genus *Elaeocarpus*.

- **New Caledonian clade**

Approximately 29 species of *Elaeocarpus* are recorded from New Caledonia (Tirel 1983), 16 of which were included in this study: *E. alaternoides* Brongn. & Gris, *E. brachypodus* Guillaumin, *E. bullatus* Tirel, *E. coumbouiensis* Guillaumin, *E. dognyensis* Guillaumin, *E. geminiflorus* Brongn. & Gris, *E. gordonii* Tirel, *E. gummatus* Guillaumin, *E. hortensis* Guillaumin var. *neocaledonica* Tirel, *E. ovigerus* Brongn. & Gris, *E. pulchellus* Brongn. & Gris, *E. rotundifolius* Brongn. & Gris, *E. seringii* Montrouz., *E. speciosus* Brongn. & Gris, *E. weibelianus* Tirel and *E. yateensis* Guillaumin. *Elaeocarpus ovigerus* was represented by *trnV-ndhC* sequence data only (amplification of the rest of the DNA markers was unsuccessful) and the results resolve the species in a very weakly supported clade that comprises most of the New Caledonian species (Appendix 2.5). *Elaeocarpus dognyensis* is in an unresolved position but nested in the same big clade (node 7, Fig. 2.6) as the New Caledonian clade (node 13). The remaining species included belong to a weakly supported New Caledonian clade which is characterised by the following combination of morphological characters: flowers 4- or 5-merous, petals divided into 2 – 12 lobes at the apex, ovary usually with 2 loculi (*E. brachypodus* has 2 or 3 loculi) each containing 2 – 10 ovules, embryo straight, and endosperm entire (Tirel 1983, 1984; Coode unpublished). Most of the New Caledonian species included are unplaced in the current infrageneric systems.

- **group VII clade**

Group VII consists of seven species all of which were included in this study: *E. carolinae* B.Hyland & Coode, *E. grahamii* F.Muell., *E. linsmithii* Guymer, *E. reticulatus* Sm., *E. culminicola*, *E. eumundi* F.M.Bailey and *E. kirtonii* F.Muell. ex F.M.Bailey. The centre of diversity is in Australia, with *E. culminicola* Warb. extending through Malesia

probably to the Philippines (Coode 1978, 1984). The seven species formed a robust clade (node 16, Fig. 2.6), however *E. multiflorus* from the *coilopetalum* group (Coode unpublished data) is nested within it. This species is known from the southern Philippines, Sulawesi and Moluccas, and the DNA sample was obtained from a sterile specimen collected from a plant cultivated in the Bogor Botanic Gardens. Confirmation of this unexpected result is required, for which a fertile, unambiguously identified specimen of *E. multiflorus* will be necessary

The group VII clade can be distinguished by the following combination of the morphologies: petals divided into 8 – 28 divisions at the apex, stamens 15 – 40, anthers awned, disk glabrous, ovary glabrous and with 2 or 3 loculi each with 6 – 10 ovules, fruit stone sculptured, not flattened, embryo curved, and endosperm entire or ruminate (*E. linsmithii*), (Coode 1984). The inclusion of *E. multiflorus*, if confirmed would alter the diagnosis as follows: petals as few as 3 divisions at the apex, disk hairy or glabrous, stamens up to 35, ovary sparsely hairy or glabrous, and fruit stone surface with or without longitudinal ridges, rugose or sculptured (Coode 2001d).

- **group XI subgroup B clade**

Group XI subgroup B (XIB) consists of seven recognised taxa endemic to Australia, all of which were included in this study: *E. elliffii* Hyland & Coode, *E. ferruginiflorus*, *E. foveolatus*, *E. largiflorens* C.T.White subsp. *largiflorens*, *E. largiflorens* subsp. *retinervis* Hyland & Coode, *E. sericopetalus* F.Muell. and *E. thelmae* Hyland & Coode (Coode 1984). In addition, a putative new species, *Elaeocarpus* sp. "Mt Windsor Tableland" which is morphologically similar to *E. largiflorens*, was included. Most of the taxa formed a clade (ML bootstrap & ML bootstrap < 50 %, PP 0.99, node 14, Fig. 2.6), except *E. ferruginiflorus* and *E. foveolatus*, which are unresolved. Similarly, the relationships between group XIB and the monotypic group XI subgroup A (*E. ruminatus*) are unresolved. Nonetheless, *E. ferruginiflorus*, *E. foveolatus* and *E. ruminatus* are situated on a polytomy in the same big clade (node 7, Fig. 2.6) as the group XIB clade. Further analysis using more informative DNA markers may shed light on the relationships within and between subgroup B and subgroup A.

The group XIB clade (excluding *E. ferruginiflorus* and *E. foveolatus*) can be identified based on the following combination of morphological character states: petals entire or with up to 9 divisions at the apex, stamens 30 – 70, anthers not awned but sometimes beaked, ovary hairy with 3 loculi each with 5 or 6 ovules, fruit stone not flattened, embryo curved, and endosperm ruminate (Coode 1984).

- ***acronodia* clade**

The *acronodia* group comprises about 28 species and is centred on West Malesia, with a few members extending to continental Asia and the Lesser Sunda Islands (Coode 1996b, unpublished data; Weibel 1968). This study sampled 10 taxa: *E. chrysophyllus* Merr., *E. euneurus*, *E. ferrugineus* (Jack) Steud., *E. jacobsii* Coode, *E. knuthii* Merr. subsp. *knuthii*, *E. marginatus*, *E. mastersii*, *E. nanus* subsp. *congestifolius* (R.Knuth) Coode, *E. nanus* subsp. *nanus*, and *E. nitentifolius*. The *acronodia* clade retrieved in this study comprises seven of the sampled taxa (node 15, Fig. 2.6) – the relationships of *E. nanus* (including its two subspecies) to the *acronodia* clade are weakly supported (MP bootstrap & ML bootstrap < 50 %, PP 0.79), and *E. nitentifolius* from China and one of the specimens of *E. mastersii* King (G 423) are unresolved. All of the representatives of the *acronodia* group included are nevertheless situated on a polytomy of the same big clade (node 7) as the *acronodia* clade.

Notably, the specimen of *E. mastersii* (G 423) was collected from semi-deciduous rainforest in northern Peninsular Malaysia (Pulau Langkawi), a rather different habitat than the rest of the specimens of *E. mastersii* studied here which were from evergreen rainforest. However, the question of whether G 423 represents a different taxonomic entity needs further study using additional molecular data and broader sampling across its distributional and ecological range.

The *acronodia* clade can be identified based on the following combination of morphological character states: flowers usually 4-merous, unisexual or bisexual, petals either undivided or bearing up to 18 small divisions, disk weakly developed, stamens 8

– 12, anthers not awned, ovary 2 locular each with 4 ovules, embryo curved, and endosperm ruminate (Coode 1996b).

- ***polystachyus* clade**

The West Malesian *polystachyus* group or the *E. polystachyus* complex comprises six species (Coode 1996c) four of which were included in this study. *Elaeocarpus clementis* (including three of its varieties: var. *borneensis* (Ridl.) Coode, var. *clemensiae* and var. *clementis*), *E. cupreus* and *E. multinervosus* formed a well-resolved clade (node 17, Fig. 2.6), but the relationships of *E. polystachyus* to the clade were unresolved beyond its situation on a polytomy of the same big clade (node 7) as the *polystachyus* clade. The other two species, *E. integripetalus* Miq. and *E. polyanthus* Ridl., were not sampled. The former is known only from the type specimen, whereas the latter is known only from the Semengoh Arboretum in Kuching (Sarawak, Malaysia) and resampling from the arboretum was unsuccessful because the known trees are no longer extant. A specimen resembling *E. polyanthus* was collected from the Kelabit Highlands in Baram District in the 1970s (Coode 1996c), but resampling from this location was not possible during the course of this study.

The *polystachyus* clade is morphologically distinguishable from other clades based on the following combination of characters: flowers usually 5-merous and bisexual (functionally unisexual flowers present in *E. cupreus* and *E. multinervosus*), disk weakly developed, stamens 35 – 65 and densely arranged in multiple tiers, filaments almost absent, ovary usually 3 locular each with 5 – 9 ovules, embryo curved, and endosperm ruminate (Coode 1996c).

The *polystachyus* group has been compared to the Australian group XIB due to their morphological similarities: sepaloid petals present; stamens numerous, filaments short and densely arranged in multiple tiers, anthers not awned; ovary 2- or 3-locular; embryo curved; and endosperm ruminate (Coode 1984, 1996c). However, two species in the *polystachyus* group have functionally unisexual flowers (see above), whereas only bisexual flowers are present in the group XIB. The present study shows that the

polystachyus and the group XIB clades belong to the same big clade (node 7, Fig. 2.6), but their phylogenetic relationships are unresolved. Expanded analyses incorporating additional and more informative markers are needed.

Coode (1995) hypothesised that the *polystachyus* group is closely related to a group of Sulawesian species: he called this combined group of 10 species the "enlarged *polystachyus* group". These groups share several floral similarities, particularly stamens numerous and densely arranged in multiple tiers, and anthers not awned. Again, however, functionally unisexual flowers are present in two members of the *polystachyus* group, whereas all members of the "enlarged *polystachyus* group" have bisexual flowers only (Coode 1995). Unfortunately, none of the Sulawesian species of the "enlarged *polystachyus* group" were available for this study, therefore tests of the evolution and taxonomic value of the character states upon which a relationship was hypothesised must await additional analyses.

- ***coilopetalum* clade**

The *coilopetalum* group is a species-rich group, consisting of approximately 82 taxa. The group is widely distributed in Malesia, extending to the Pacific islands (Coode unpublished data). A total of 10 samples were analysed: five of them (*E. angustipes* R.Knuth, *E. griffithii* (Wight) A.Gray, *E. jugahanus*, *E. petiolatus* (Jack) Wall. and *E.* cf. *pseudopaniculatus* Corner) formed a well-supported clade (node 18, Fig. 2.6), whereas four (*E. acrantherus*, *E. kusanoi*, *E. palembanicus* and *E.* cf. *sadikanensis*) are not well-resolved, being part of a polytomy of the same big clade (node 7). The last, *E. multiflorus*, is nested in the group VII clade as noted above. Taxa belonging to the *coilopetalum* clade share the following character states: anthers awned, disk "cog-wheel" shaped, ovary 2- or 3-locular with 6 – 10 ovules per loculus, fruit stone not flattened and without obvious ridges, flanges or wings, embryo curved, and endosperm entire or ruminate (Coode 1998, 2001d, unpublished data). The representatives of the *coilopetalum* group studied here are primarily from the West Malesian region. Future studies should expand the biogeographical sampling.

2.4.5 Seed morphology

The results of the parsimony reconstruction indicate that the curved embryo and ruminate endosperm are both homoplasious within *Elaeocarpus*. The curved embryo has arisen at least three times, in *Sericolea*, *E. holopetalus* and in some lineages of the remainder of *Elaeocarpus* clade. Thus the state observed in these three groups is not homologous. Detailed studies of embryo anatomy and ontogeny are required to understand the basis of this trait. Clearly, however, it is of limited use as a taxonomic character across the family, but may serve to diagnose more restricted groups such as clades within *Elaeocarpus*, if shown to be homologous at that level.

Ruminate endosperm has at least two origins within *Elaeocarpus*. As is the case for embryo shape, detailed studies of ontogeny may provide a better understanding of the basis of these non-homologous states. Noticeably, ruminate endosperm is found only in the curved embryo species, but the bases for this apparent association between embryo shape and endosperm ornamentation in *Elaeocarpus* are unknown. While homoplasious at higher taxonomic levels, like embryo shape, ruminate endosperm has some taxonomic value within *Elaeocarpus*.

Endosperm may become ruminate either as a consequence of ingrowths of the integuments (Periasamy 1962) or infolding of the testa (Bayer and Appel 1996). However, investigations on the seed of *Elaeocarpus* are scarce, thus detailed studies of embryo anatomy and ontogeny may provide understanding on the development of endosperm as well as the functional significance of ruminate endosperm in *Elaeocarpus*. Bayer (1996) hypothesised two roles for ruminate endosperm. The first role is defence – ruminate endosperm often contains ethereal oil, which may render the seeds less attractive to predators (Goebel 1933, Bayer 1996). Secondly, the enlargement of the contact area between endosperm and integument may facilitate the supply of nutrients, oxygen and/or water for the development of embryo and storage tissue (Goebel 1933). There is no evidence for or against either of these hypotheses for *Elaeocarpus*.

2.5 Conclusions

A broad sample of *Elaeocarpus* (more than 25 % of the species) representing various infrageneric groups and biogeographical regions was included in this study, greatly improving on the sampling of previous studies. A total of four DNA regions were sequenced, two of them for the first time in *Elaeocarpus*, and phylogenetic analysis of the concatenated dataset has greatly enhanced phylogenetic resolution compared to previous analyses. Relationships within the *Elaeocarpus* alliance are robustly resolved, as is the monophyly of the genus *Elaeocarpus* and main lineages within it. The study also provides the first assessment of evolutionary transformations of seed morphology within *Elaeocarpus* that will be useful for future taxonomic treatments.

While a number of clades are robustly resolved, relationships among the majority of *Elaeocarpus* species are not yet clear due to low sequence divergence. Moreover, the Madagascan, continental Asian and east Malesian (especially New Guinea) species are not well represented in the dataset. The integration of data from additional informative DNA regions and broader taxon sampling promises to shed more light on the phylogeny and evolution of this highly diverse genus.

The results of the molecular phylogenetic analyses presented in this chapter provide the first comprehensive phylogenetic framework for *Elaeocarpus* which will serve as a strong foundation to further address the questions of divergence times among lineages and historical biogeographical questions relating to past migration events (Chapter 3). Despite that the molecular phylogenetic analyses did not fully resolve the question of monophyly of the *polystachyus* group, the combination of molecular and morphological evidence suggests that the group is probably monophyletic, thus providing the basis for a morphometric analysis to assess species limits with the group (Chapter 4).

The following people made valuable contributions to this chapter: Mark Coode for discussions and for sharing his unpublished data and knowledge of Elaeocarpaceae; Andrew Thornhill and Katharina Schulte for advice on the molecular dating analysis and biogeographical reconstructions, respectively; Yumiko Baba for discussions on Australian *Elaeocarpus*; Gary Wilson and Philippa Griffin for critique; and Choon-Kit Tan and Kim-Wai Tee for software and computing support.

Chapter 3: Evolutionary divergence times in *Elaeocarpus* (Elaeocarpaceae): evidence of its origin in Australia with subsequent diversification in the surrounding regions

ABSTRACT

The taxonomy and morphology of *Elaeocarpus* have been well documented, but molecular dating and historical biogeographical studies are lacking. The present study used the uncorrelated lognormal relaxed molecular clock and five internal fossil calibration points to estimate the age of Elaeocarpaceae and diversification times of the main labelled clades within the family with a focus on the genus *Elaeocarpus*. Over 3000 bp of DNA sequence from four regions (three plastid markers – *psbA-trnH*, *trnL-trnF* region and *trnV-ndhC*; and one nuclear – *Xdh*) were sequenced from 154 taxa of Elaeocarpaceae, including 114 *Elaeocarpus* taxa from various biogeographic regions. Ancestral areas at internal nodes within *Elaeocarpus* were determined using Fitch parsimony and Dispersal-Extinction-Cladogenesis. The family Elaeocarpaceae is estimated to have a crown age of about 82.98 Mya, indicating that the family began to diversify in the late Cretaceous. Two sister lineages within the family are identified: the first comprising the *Aristotelia* alliance and *Sloanea* alliance; and the second comprising the remainder of the Elaeocarpaceae, i.e. *Crinodendron*, *Peripentadenia*, *Dubouzetia*, the Tremandraceous genera and the *Elaeocarpus* alliance. *Elaeocarpus* is inferred to have originated in Australia in the Eocene and began to diverge from its sister, *Aceratium*, at about 45.46 Mya. Diversification within *Elaeocarpus* was initiated at about 39.15 Mya, with the split between the *E. holopetalus* and the remainder of the genus. Within the latter, 13 labelled sub-lineages or clades are resolved, many of which are geographically based. Subsequent migration events were northwards as far north as to continental Asia, westwards to Madagascar, and eastwards to New Zealand and the Pacific islands. Several migratory reversals are also postulated.

3.1 Introduction

The Elaeocarpaceae is a large neo- and palaeo-tropical, subtropical and warm temperate family of flowering plants. The family comprises 12 genera (*Aceratium* DC., *Aristotelia* L'Hér., *Crinodendron* Molina, *Dubouzetia* Brongn. & Gris., *Elaeocarpus* L. *Peripentadenia* L.S.Sm., *Platytheca* Steetz, *Sericolea* Schltr., *Sloanea* L., *Tetratheca* Sm., *Tremandra* R.Br. and *Vallea* L.) and *c*. 550 species, and is widely distributed in predominantly tropical forests, but not on mainland Africa (Coode 2004). It is essentially a Gondwanan element, with all of its genera restricted to the Southern hemisphere except *Vallea,* which extends slightly north of the equator in Colombia and Venezuela, *Elaeocarpus* which extends to Japan and southern China, and *Sloanea* which extends to Central America, Malesia and continental Asia (Fig. 3.1) (Coode 2004).

Various fossils have been compared with or assigned to Elaeocarpaceae. These include wood fragments from India (compared with *Elaeocarpus*, Prakash and Dayal 1964) and South America (compared with Elaeocarpaceae, Petriella 1972), leaves from Europe (compared with or assigned to *Sloanea*, Kvaček *et al*. 2001; Kvaček 2002; Sachse 2005; Hably 2007; Erdei and Rakosi 2009), capsules from Europe (compared with or assigned to *Sloanea*, Kvaček *et al*. 2001; Sachse 2005; Erdei and Rakosi 2009; Collinson *et al*. 2010) and North America (compared with or assigned to *Sloanea*, Manchester and Kvaček 2009), fruit stones (or mesocarps) in Australia (compared with or assigned to *Elaeocarpus*, Rozefelds and Christophel 1996a, b; Dettman and Clifford 2000; Rozefelds and Christophel 2002) and seeds (probably referring to the fruit stones) in New Zealand (assigned to *Elaeocarpus*, Burrows 1995, 1997). Of these fossil records, the capsules and the fruit stones can be reliably assigned to the genera *Elaeocarpus* (Dettman and Clifford 2000; Rozefelds and Christophel 2002) and *Sloanea* (Kvaček *et al*. 2001; Manchester and Kvaček 2009), respectively, as they bear close morphological resemblance to extant species. Fossil fruits assigned to *Elaeocarpus* have a 'woody' mesocarp and a surface that is either smooth, ridged or variously sculptured; are often one-seeded at maturity but with 2 – 9 locules, with the fertile locule much expanded and compressed towards the neighbouring sterile ones; and

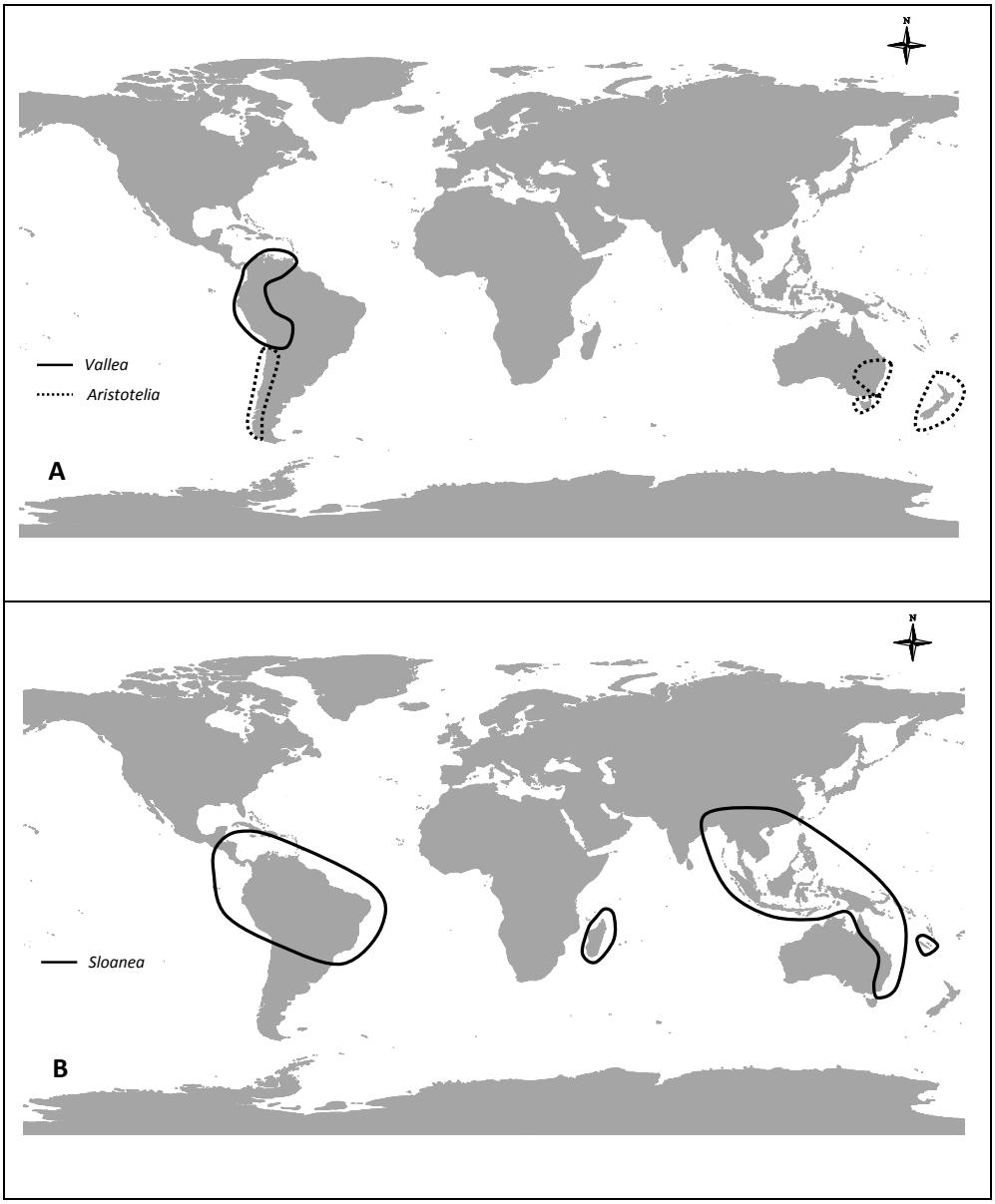


Fig. 3.1 Approximate distribution of Elaeocarpaceae. A, *Aristotelia* alliance. B, *Sloanea* alliance. C, *Crinodendron* alliance. D, *Elaeocarpus* alliance. E, Tremandraceous genera (infrafamilial groupings and distribution follow Coode 2004).

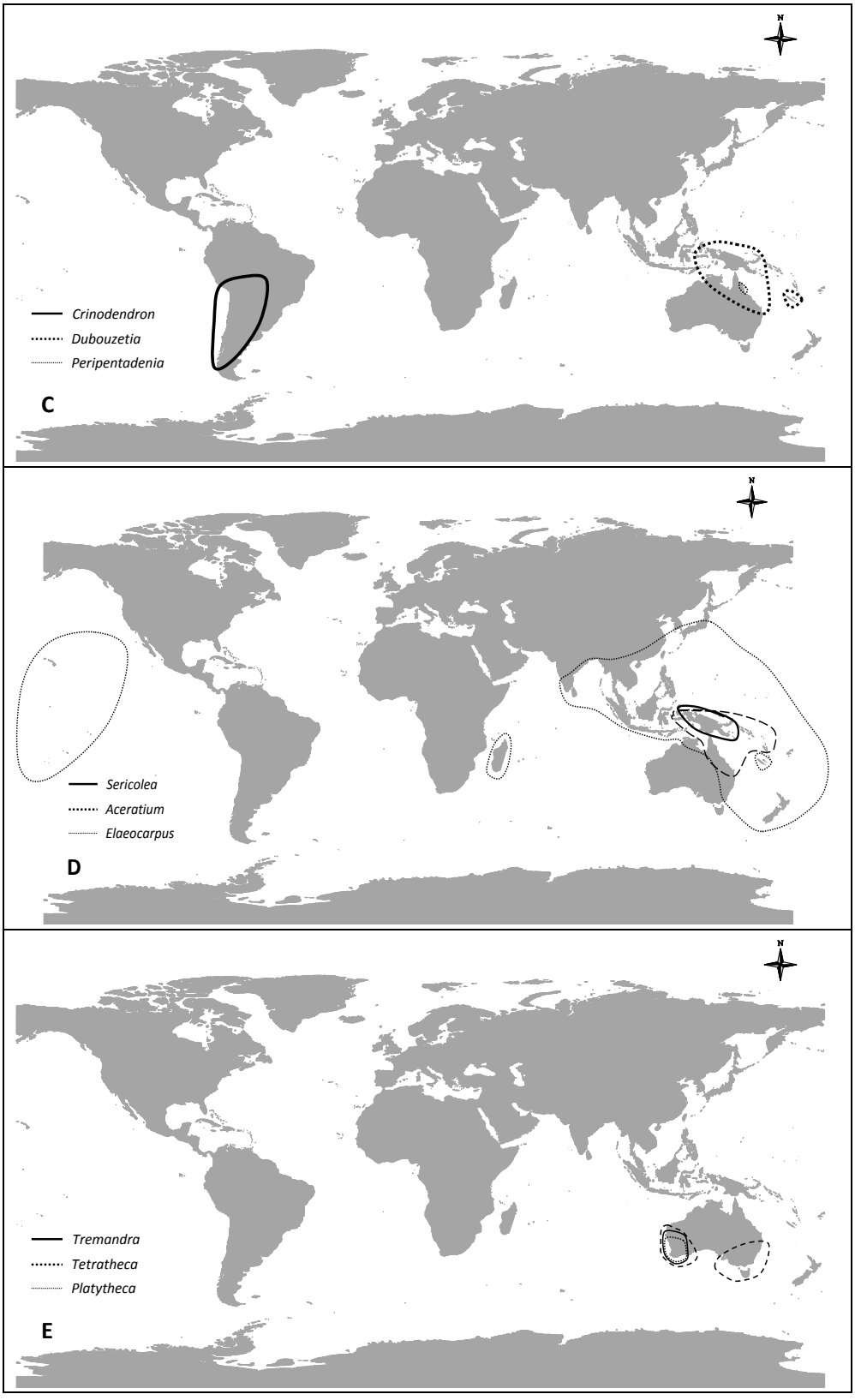


Fig. 3.1 (continued).

have anatropous and pendulous ovules with a ventral raphe (Dettman and Clifford 2000). Fossil fruits assigned to *Sloanea* are loculicidally dehiscent, valved capsules that are often spiny on the outside (Kvaček *et al*. 2001).

The most comprehensive phylogeny of Elaeocarpaceae *sensu lato* (including the former Tremandraceae) so far was reconstructed by Crayn *et al*. (2006) based on molecular data, primarily from the Australasian and the Pacific species. The findings recognise Elaeocarpaceae as a monophyletic family closely related to the Southwestern Australian family Cephalotaceae and the Central and South American family Brunelliaceae. Within the family, all of the genera were resolved as monophyletic, except *Sericolea*, *Aceratium* and *Elaeocarpus* (the *Elaeocarpus* alliance, Coode 2004). The three genera formerly recognised as Tremandraceae were each resolved as monophyletic and were together placed sister to the *Elaeocarpus* alliance (Crayn *et al*. 2006).

Several studies have estimated lineage divergence times for the Elaeocarpaceae based on relaxed-clock analysis of molecular data (Wikström *et al*. 2001; Crayn *et al*. 2006; Heibl and Renner 2012). These studies were based on plastid (*rbcL*, *atpB*, *trnL* intron and *trnL-trnF* spacer), nuclear (ITS and 18S), or a combination of plastid and nuclear DNA sequence data.

The present study focuses on the genus *Elaeocarpus*, which is the most species-rich genus in the family Elaeocarpaceae (Coode 2004). It is included with *Aceratium* and *Sericolea*, which in the *Elaeocarpus* alliance a recently diverged lineage in the family (Crayn *et al*. 2006). *Elaeocarpus* comprises *c*. 350 species of mostly trees and shrubs inhabiting palaeo-tropical, palaeo-subtropical and warm temperate forests, ranging from lowland to montane habitats of Madagascar, Malesia, Asia, Australia, New Zealand and the Pacific islands.

Seed dispersal of some *Elaeocarpus* is known to be facilitated by animals (zoochory), particularly by birds (Crome 1975, 1976; Ridley 1990; Aggarwal 2002;

Rossetto *et al*. 2008). Some *Elaeocarpus* are commonly found near rivers or streams, but dispersal via water (hydrochory) has not been recorded.

The Malesian phytogeographical region harbours the highest species diversity of *Elaeocarpus*. It is separated into three main geographical subregions: West Malesia (including Malay Peninsula, Sumatra, Java, Borneo and the Palawan island), Central Malesia (also known as Wallacea, including Sulawesi, the Lesser Sunda islands, the Moluccas and the Philippines) and East Malesia (New Guinea) (Hall 2001a). New Guinea (*c*. 85 species) and Borneo (*c*. 70 species) represent the two most important centres of species richness and endemism for the genus (Coode 2004).

The historical and evolutionary events and processes that have resulted in this extant diversity have never been investigated for *Elaeocarpus* due largely to the lack of a suitable phylogenetic hypothesis. The only three indirectly relevant studies are by Crayn *et al*. (2006), Niissalo (2011) and Heibl and Renner (2012): the first included only seven (1.75 % of the total) *Elaeocarpus* species, all from Australia; the second included one (0.29 % of the total) unspecified taxon of *Elaeocarpus*; the last included the same set of *Elaeocarpus* species as in Crayn *et al.* (2006). Therefore, the age of *Elaeocarpus* has not been adequately investigated, neither has its origin and diversification history.

3.1.1 Aims

This chapter aims to: (1) infer the origin of *Elaeocarpus*, (2) estimate its divergence time, and (3) explain the current distribution patterns of the main lineages within the genus.

3.2 Materials and methods

3.2.1 Taxon sampling

A total of 154 taxa of Elaeocarpaceae were sampled, including 114 *Elaeocarpus* taxa. The ingroup included all of the *Elaeocarpus* species and the outgroups included

48 species from 11 other genera of Elaeocarpaceae. Five other outgroup taxa representing the three most closely related families – Brunelliaceae, Cephalotaceae, and Cunoniaceae (Crayn *et al*. 2006, Heibl and Renner 2012; Magallón *et al*. 2015) – were included to root the trees: *Brunellia colombiana* Cuatrec. (Brunelliaceae), *Cephalotus follicularis* Labill. (Cephalotaceae), *Ackama rosifolia* A.Cunn. (Cunoniaceae), *Cunonia balansae* Brongn. & Gris (Cunoniaceae) and *Davisonia pruriens* F.Muell. var. *jeryseyana* F.M.Bailey (Cunoniaceae).

3.2.2 DNA sequencing and alignment

The materials and methods for DNA extraction, PCR amplification and sequencing were detailed in Chapter 2. Sequences were assembled and edited in ChromasPro ver. 1.32 (Technelysium Pty. Ltd.). Sequences from each locus were aligned using MAFFT multiple sequence alignment software online version 7 (Katoh and Standley 2013) and then refined manually using Geneious Pro 5.6.5 (Biomatters Ltd.). *Xdh* sequences were checked with amino acid translation in Geneious Pro 5.6.5 to ensure no premature stop codons occurred within the sequence; no pseudogene sequences were found in this region. Datasets were trimmed and ambiguously aligned positions were excluded from the analyses. The combined four-region matrix comprised 3617 characters (see Table 3.1 for data on individual gene regions). Sequences that were downloaded from GenBank or sourced from other unpublished work are presented in Appendix 2.1.

Table 3.1 Length and best-fitting model for data partition analyses. GTR, general time-reversible model; HKY, the Hasegawa, Kishino and Yano model; I, proportion of invariant sites; Γ, gamma-shaped distribution of rates across sites.

	Aligned length (bp)	**Sequence evolutionary model**
psbA-trnH	1029	HKY + I + Γ
trnL-trnF region	1215	GTR + Γ
trnV-ndhC	770	GTR + Γ
Xdh fragment 1	264	HKY + Γ
Xdh fragment 2	339	HKY + Γ

3.2.3 Divergence time estimation

MrModeltest version 2.3 (Nylander 2008) was used to determine the appropriate DNA substitution model and gamma rate heterogeneity using the Akaike Information Criterion (AIC) for each gene region. The concatenated data matrix was partitioned by gene and the best-fitting evolutionary model was applied to each partition (refer Table 3.1).

The tree topology, node ages and substitution rates were simultaneously estimated using a Bayesian MCMC (Markov Chain Monte Carlo) approach implemented in BEAST version 1.8.1 (Drummond and Rambaut 2007, 2008). A Yule speciation tree prior was specified; this prior has been recommended for species-level phylogenies, and assumes a constant rate of speciation per lineage (Drummond and Rambaut 2008). An uncorrelated lognormal distribution relaxed clock (UCLN) model was employed; it allows evolutionary rates to vary along branches according to a lognormal distribution (Drummond *et al*. 2006). The normal probability distribution is considered appropriate for modelling uncertainty in secondary calibration points (Ho 2007; Couvreur *et al*. 2008; Bergh and Linder 2009). Normal distribution priors were therefore applied to calibration points at the root node. Two independent MCMC runs of 70 million generations were performed, with sampling every 5000 generations. The two separate runs were then combined (following the removal of 25 % burn-in) using LogCombiner version 1.8.1 (Drummond and Rambaut 2007, 2008).

Adequate sampling and convergence of the chain to stationary distribution were confirmed by inspection of MCMC samples using Tracer version 1.5 (Rambaut and Drummond 2007). The effective sample size (ESS) values of all parameters were greater than 200, which were considered a sufficient level of sampling. The sampled posterior trees were summarised using TreeAnnotator version 1.8.1 (Drummond and Rambaut 2007, 2008) to generate a maximum clade credibility tree (maximum posterior probabilities) and to calculate the mean ages, 95 % highest posterior density

(HPD) intervals, posterior probabilities (PP) and substitution rates for each node. The tree was visualised with FigTree version 1.4.2 (Rambaut 2014).

3.2.4 Root node calibration

The divergence times of angiosperm lineages were estimated by Magallón *et al*. (2015) using penalized likelihood (Sanderson 2002) and uncorrelated lognormal (UCLN) Bayesian relaxed clock methods (Drummond *et al*. 2006). That analysis used a large dataset of reliable fossil records to calibrate the molecular evolutionary rates and the partitioned sequence dataset comprised sequences of five plastid (*atpB*, *rbcL* and *matK*) and nuclear (18S and 26S nuclear ribosomal DNA) markers. Taxon sampling within the angiosperms included 100 % of orders and 87 % of families that are currently recognised on the Angiosperm Phylogeny Website, version 13 (Stevens 2013, accessed Dec 20, 2014).

The estimated crown age of the clade Elaeocarpaceae + Cunoniaceae + Cephalotaceae + Brunelliaceae is 84.5 Mya (95 % HPD: 90.66 – 80.6 Mya) based on UCLN Bayesian relaxed clock analysis (Magallón *et al*. 2015). This was used as an estimate of the age of the root node of the tree in the present analysis and applied as a secondary calibration point.

3.2.5 Fossil calibrations

Internal calibrations were done using macrofossils. Including a greater number of fossil calibration points can reduce bias and result in more accurate age estimation, assuming that the chosen points are mutually consistent (Hug and Roger 2007). Five macrofossils (four *Elaeocarpus* and one *Sloanea*) were selected for analyses and assignment of these macrofossils to genus was unequivocal because these fossils are morphologically highly similar to certain extant taxa (as discussed in Introduction) and no similar capsules or fruit stones are known elsewhere within angiosperms. Assessment of the phylogenetic placement within these genera was not attempted

because the diversity of fruit morphology within *Elaeocarpus* and *Sloanea* is inadequately characterised for this purpose.

Morphological comparisons between the macrofossils and the fruit stones of the extant taxa were detailed in Dettmann and Clifford (2000). The four *Elaeocarpus* macrofossils used in the internal calibrations were: (i) *Elaeocarpus spackmaniorum* Rozefelds (Oligocene to early Miocene), which resembles the extant taxon *E. angustifolius* Blume complex. The mean age of the fossil (25 Mya) was used as a conservative estimate of the crown age of the *ganitrus* clade; (ii) *E. lynchii* (F.Muell.) Selling (Miocene, mean age 14 Mya, resembles *E. grahamii* F.Muell.) as a conservative estimate of the crown age of the group VII clade; (iii) *E. rozefeldsii* Dettman & Clifford (Oligocene, mean age 28 Mya, resembles *E. stellaris* L.S.Sm.) as a conservative estimate of the crown age of the *E. stellaris* and *Elaeocarpus* sp. "Mt Misery" clade; and (iv) *E. mackayi* (F.Muell.) Kirchheimer (Early Oligocene to Miocene, mean age 20 Mya) resembles species of the clades *ganitrus* and section (sect.) *Elaeocarpus* and represents a conservative estimate of the age of divergence of these two clades. Other known macrofossils of *Elaeocarpus* are either younger and so might bias estimation of the ages of deeper nodes (Bell and Donohue 2005) or cannot be assigned to any of the clades with confidence due to there being very few morphological similarities between the macrofossils and the fruit stones of the extant taxa studied. The *Sloanea* macrofossil, *S. eocenica* (Rasky) Z. Kvaček, Hably & Manchester (Kvaček *et al*. 2001) (early Oligocene, mean age 31 Mya), was assigned as a conservative estimate of the crown age of the *Sloanea* lineage.

The ranges of possible ages of nodes calibrated using macrofossils were modelled as a lognormal distribution (Table 3.2). The ages of the macrofossils were assigned to the nodes of the MRCAs of the crown groups (Table 3.2) by enforcing the monophyly of these clades. In all cases the monophyly of these constrained clades was justified by the results of the phylogenetic analyses in Chapter 2.

Table 3.2 Calibration points and age constraints used in divergence time estimations (MRCA = most recent common ancestor). The macrofossil was assigned to the MRCA of the listed nodes. Geologic time scale following Gradstein and Ogg (2004).

MRCA (Node)	Macrofossil	Geologic epoch (Mya)	Offset (Log (stdev))	References
C1	NA (Secondary calibration using crown age of Oxalidales)	Cretaceous (90.66 – 80.6)	84.5 (2.3)	Magallón *et al*. (2015)
5	*Sloanea eocenica* (Rásky) Z. Kvaček, Hably & Manchester	Early Oligocene (33.9 – 28.1)	28.3 (1.2)	Kvaček *et al*. 2001
30	*Elaeocarpus mackayii* (F.Muell.) Kirchheimer	Early Oligocene – Miocene (33.9 – 5.332)	28.1 (1.0)	Dettman and Clifford (2000)
31	*Elaeocarpus spackmaniorum* Rozefelds	Oligocene – Early Miocene (33.9 – 15.97)	23.0 (1.7)	Dettman and Clifford (2000)
57	*Elaeocarpus rozefeldsii* Dettman & Clifford	Oligocene (33.9 – 23.03)	23.0 (1.7)	Dettman and Clifford (2000)
60	*Elaeocarpus lynchii* (F.Muell.) Selling	Miocene (23.03 – 5.332)	15.0 (1.5)	Dettman and Clifford (2000)

3.2.6 Reconstruction of ancestral areas

The primary focus of the reconstruction of ancestral ranges in this study was to generate and test hypotheses regarding the historical biogeography of the genus *Elaeocarpus*. Therefore, the consensus dated tree resulting from the molecular clock analysis based on the concatenated plastid and nuclear sequence data was pruned using Mesquite version 2.75 (Maddison and Maddison 2011) to retain only *Aceratium* and *Sericolea* as the outgroups, and used as input for the historical biogeography analyses.

Eight main biogeographical areas representing the current distribution of *Elaeocarpus* species and relatives were defined for the ancestral area reconstruction (Fig. 3.2): (A) Madagascar; (B) Continental Asia, including the northern part of Sunda

shelf from the Thailand-Peninsular Malaysia border northwards, and westwards to India and Sri Lanka; (C) West Malesia including the southern part of the Sunda shelf from the Thailand-Peninsular Malaysia border southwards; (D) Central Malesia (or Wallacea); (E) New Guinea and the surrounding islands (e.g. New Britain); (F) Australia; (G) New Zealand; and (H) the Pacific including New Caledonia. The Sunda shelf was separated into two biogeographical areas in this study based on the differences in the species composition between the northern and southern parts. Of 42 taxa sampled from the southern part, only five (*E. angustifolius* complex, *E. nitidus* Jack, *E. petiolatus* (Jack) Wall., *E. robustus* Roxb. and *E. stipularis* Blume) of them extend to the northern part. The northern Sunda shelf was grouped together with Continental Asia. Similarly, the Sahul shelf was segregated into two biogeographical areas, New Guinea and Australia, as these areas show marked differences in their species composition. Of the 30 taxa known from Australia including offshore islands (Australian Plant Census, accessed Jan 15, 2015), only four (*E. angustifolius*, *E. arnhemicus* F.Muell., *E. culminicola* Warb., and *E. miegei* Weibel [this last species was not sampled in this study]) are also found in New Guinea. Some of these biogeographical areas, i.e. Continental Asia (area B) and Central Malesia (area D), have different palaeogeographical origins, however the delimitation of these areas with disparate origins is less critical in this analysis as they are only occupied by *Elaeocarpus* later in geological time. The *Elaeocarpus* species sampled in this study were mostly from West Malesia, with very few representatives sampled from Central Malesia and New Guinea in comparison to the total species occurring in these areas. Therefore, to minimise the effect on the results of this bias in the taxon sampling, the finer scale biogeographic regions used in some previous studies of Malesian biogeography (Thomas *et al*. 2012; Grudinski *et al*. 2014; Van Welzen *et al*. 2014) were not used here. Species distribution data were extracted from Tirel (1983), Tang and Phengklai (2007), Maynard *et al*. (2008), Baba and Crayn (2012), Global Biodiversity Information Facility (2014, accessed Dec 23, 2014), Coode (1978, 1984, 1996a, c, 1998, 2001c) as well as unpublished data (Coode pers. comm. 2014) (Appendix 3.1).

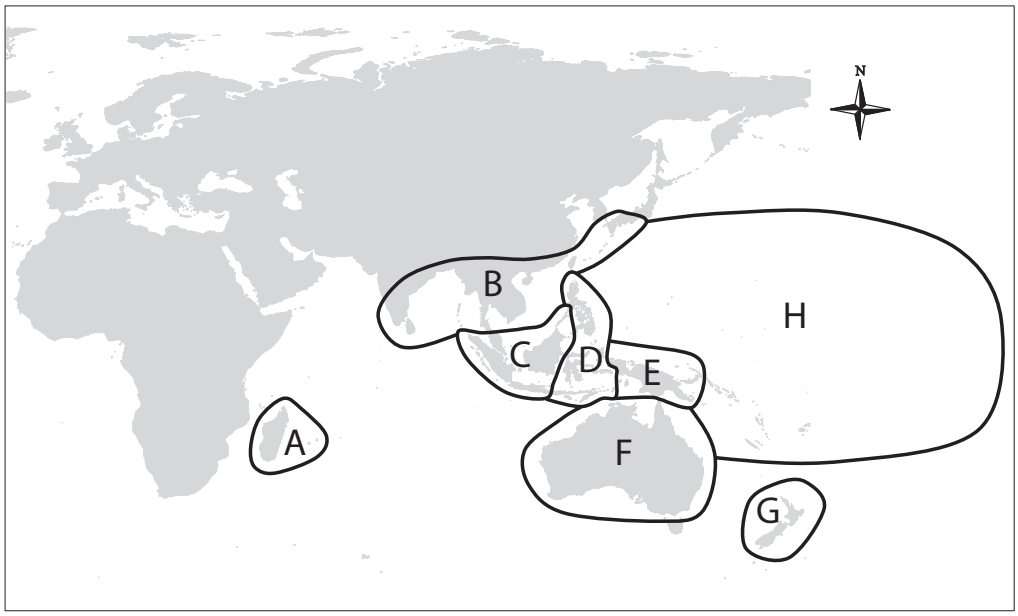


Fig. 3.2 Eight main biogeographical areas representing the current distribution of species of *Elaeocarpus*, *Aceratium* and *Sericolea* defined for the ancestral area reconstruction. A, Madagascar; B, Continental Asia, including the northern part of the Sunda shelf from the Thailand-Peninsular Malaysia border northwards, and westwards to India and Sri Lanka; C, West Malesia including the southern part of the Sunda shelf from the Thailand-Peninsular Malaysia border southwards; D, Central Malesia (or Wallacea); E, New Guinea and the surrounding islands (e.g. New Britain); F, Australia; G, New Zealand; and H, the Pacific including New Caledonia.

The ancestral biogeographical areas occupied by *Elaeocarpus* were inferred using two methods: (1) Fitch parsimony implemented in the program Mesquite version 2.75 (Maddison and Maddison 2010), and (2) the model-based method Dispersal-Extinction-Cladogenesis (DEC) implemented in the program Lagrange version 20130526 (Ree *et al*. 2005; Ree and Smith 2008). The parsimony method optimises the ancestral state as that requiring the minimum number of changes on a given phylogeny. The DEC method allows ancestral state reconstruction for multiple discrete areas and identifies dispersal, vicariance or local extinction as stochastic events through time within a likelihood framework (Ree *et al*. 2005; Ree and Smith 2008). This method evaluates all possible scenarios of the observed occurrence of taxa based on the given phylogenetic relationships and the geographical and geological histories

of the areas (Ree *et al*. 2005; Ree and Smith 2008). Two likelihood values are calculated: the global likelihood for the entire phylogenetic tree and the likelihood for each internal node of the tree. All dispersal paths were considered to have the same rate and the baseline rate of dispersal and local extinction were estimated.

3.3 Results

3.3.1 Data matrix

The concatenated five-locus data matrix (*psbA-trnH*, *trnL-trnF* region, *trnV-ndhC*, *Xdh* fragment 1 and *Xdh* fragment 2) comprised 3617 characters from 159 taxa of the family Elaeocarpaceae and five other taxa from the order Oxalidales. The aligned length of the data partitions corresponding to individual regions, and their best-fitting models, are presented in Table 3.1.

3.3.2 Divergence time estimation

The molecular clock analysis estimated the crown age of family Elaeocarpaceae at *c*. 82.98 Mya (Fig. 3.3 & Table 3.3: node 1; 95 % highest posterior probability density (HPD): 88.41 – 75.09 Mya, Bayesian posterior probability (PP) = 1), indicating that the family began to diversify in the late Cretaceous. The mean rate of evolution is 2.141 x 10^{-4} substitutions per site per million years (HPD: 1.687 x 10^{-4} – 2.604 x 10^{-4}). The "birth rate" (i.e. speciation rate) indicated by the Yule prior is 5.745 x 10^{-2} (HPD: 4.683 x 10^{-2} – 6.793 x 10^{-2}) and the coefficient of variation is 0.275 (HPD: 0.1.491 x 10^{-5} – 0.588).

The first diversification event detected within Elaeocarpaceae gave rise to two lineages: the first comprising the *Aristotelia* alliance and the *Sloanea* alliance; and the second comprising the remainder of the Elaeocarpaceae, namely *Crinodendron*, *Peripentadenia*, *Dubouzetia*, the Tremandraceous genera, and the *Elaeocarpus* alliance (naming follows Coode 2004). In the first lineage, the *Aristotelia* and *Sloanea* alliances diverged around 58.71 Mya (node 2; HPD: 78.26 – 40.26 Mya, PP = 0.99) and within

the *Aristotelia* alliance, *Vallea* diverged from *Aristotelia* around 37.75 Mya (node 3; HPD: 58.81 – 18.95 Mya, PP = 0.98). In the second lineage, *Crinodendron* and *Peripentadenia* form a clade which diverged from the rest around 76.51 Mya (node 6; HPD: 88.41 – 75.09 Mya, PP = 1), and the two genera diverged from each other around 59.64 Mya (node 7; HPD: 75.28 – 41.92 Mya, PP = 0.99). *Dubouzetia* and the Tremandraceous genera + *Elaeocarpus* alliance clade diverged around 64.88 Mya (node 10; HPD: 73.53 – 55.88 Mya, PP = 1). The Tremandraceous alliance and the *Elaeocarpus* alliance diverged from each other around 57.57 Mya (node 12; HPD: 66.77 – 48.16 Mya, PP = 1). Within the Tremandraceous alliance, *Tremandra* and *Platytheca* + *Tetratheca* diverged around 49.55 Mya (node 13; HPD: 59.23 – 39.55 Mya; PP = 1); and *Platytheca* and *Tetratheca* diverged from each other around 45.18 Mya (node 14; HPD: 55.25 – 35.68 Mya, PP < 0.95) although this node lacks strong support. Within the *Elaeocarpus* alliance *Sericolea* and *Aceratium* + *Elaeocarpus* diverged around 48.88 Mya (node 17; HPD: 57.85 – 40.12 Mya, PP < 0.95), and *Aceratium* and *Elaeocarpus* diverged around 45.46 Mya (node 19; HPD: 53.63 – 37.22 Mya, PP < 0.95).

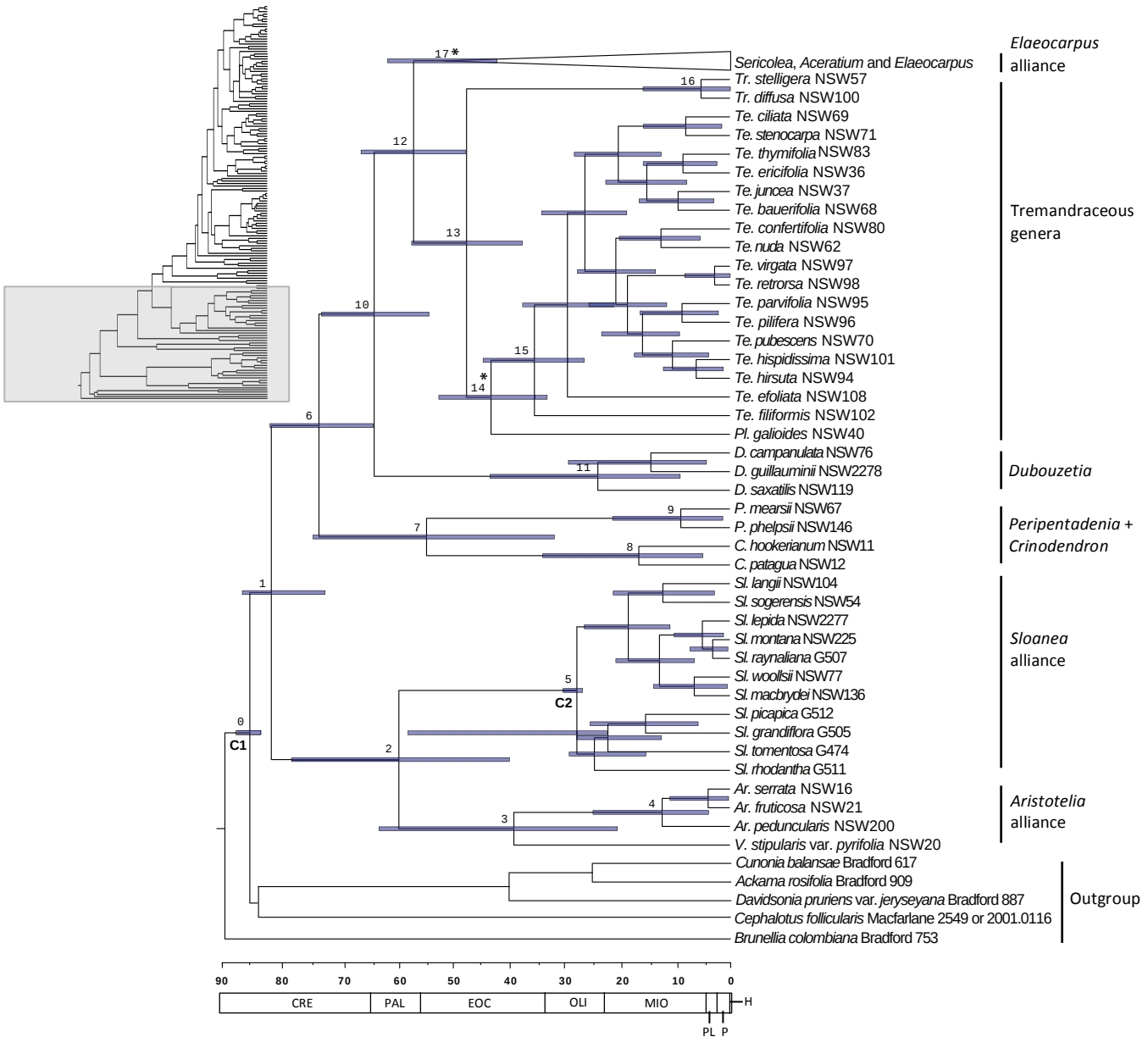


Fig. 3.3 Chronogram of Elaeocarpaceae: maximum clade credibility tree from the BEAST analysis. Posterior estimates of divergence times were inferred using partitioned analyses based on four combined DNA regions, a UCLN model, and six calibration points, i.e. one secondary (labelled C1, below the node) and five macrofossils (four *Elaeocarpus* and one *Sloanea*) (labelled C2 – C6, below the nodes). Nodes are posterior mean ages (Mya), with blue node bars representing the 95 % HPD intervals (see Table 3.3 for details). Bayesian PP < 0.95 are indicated by asterisks above branches. All infrafamilial groups of Coode (2004) retrieved, i.e. *Aristotelia* alliance, *Sloanea* alliance, *Elaeocarpus* alliance and Tremandraceous genera, except the *Crinodendron* alliance, where *Dubouzetia*, *Peripentadenia* and *Crinodendron* did not formed a clade. Genus abbreviations are *Aristotelia* (Ar), *Crinodendron* (C), *Dubouzetia* (D), *Peripentadenia* (Pe), *Platytheca* (Pl), *Tetratheca* (Te), *Tremandra* (Tr), *Sloanea* (Sl) and *Vallea* (V). DNA extraction numbers were given to the right of taxon names. Geological time scale abbreviations: Cretaceous (CRE); Paleocene (PAL); Eocene (EOC); Oligocene (OLI); Miocene (MIO); Pliocene (PL); Pleistocene (P); Holocene (H). Insert showing the location of the enlarged section of the phylogenetic tree.

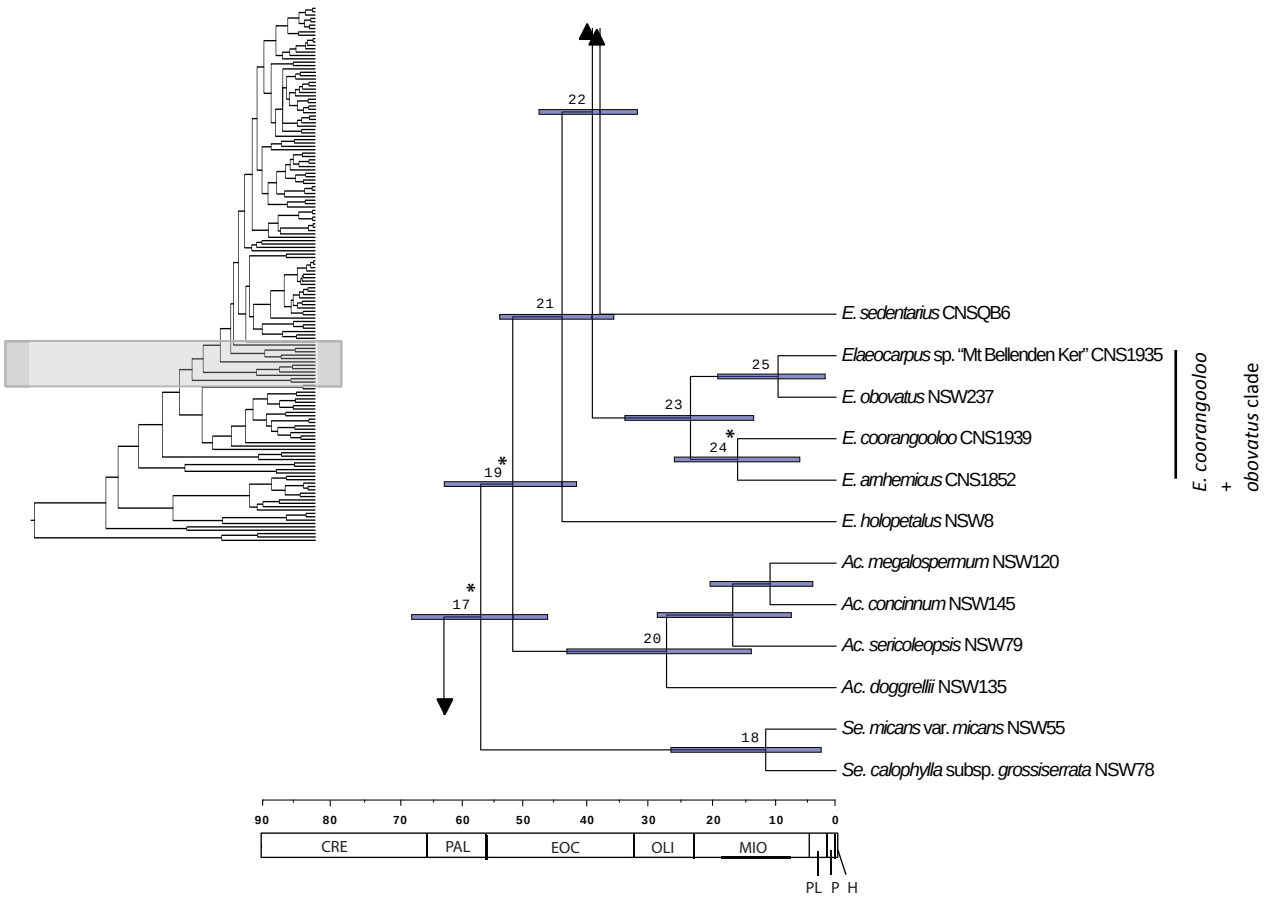


Fig. 3.3 (continued).

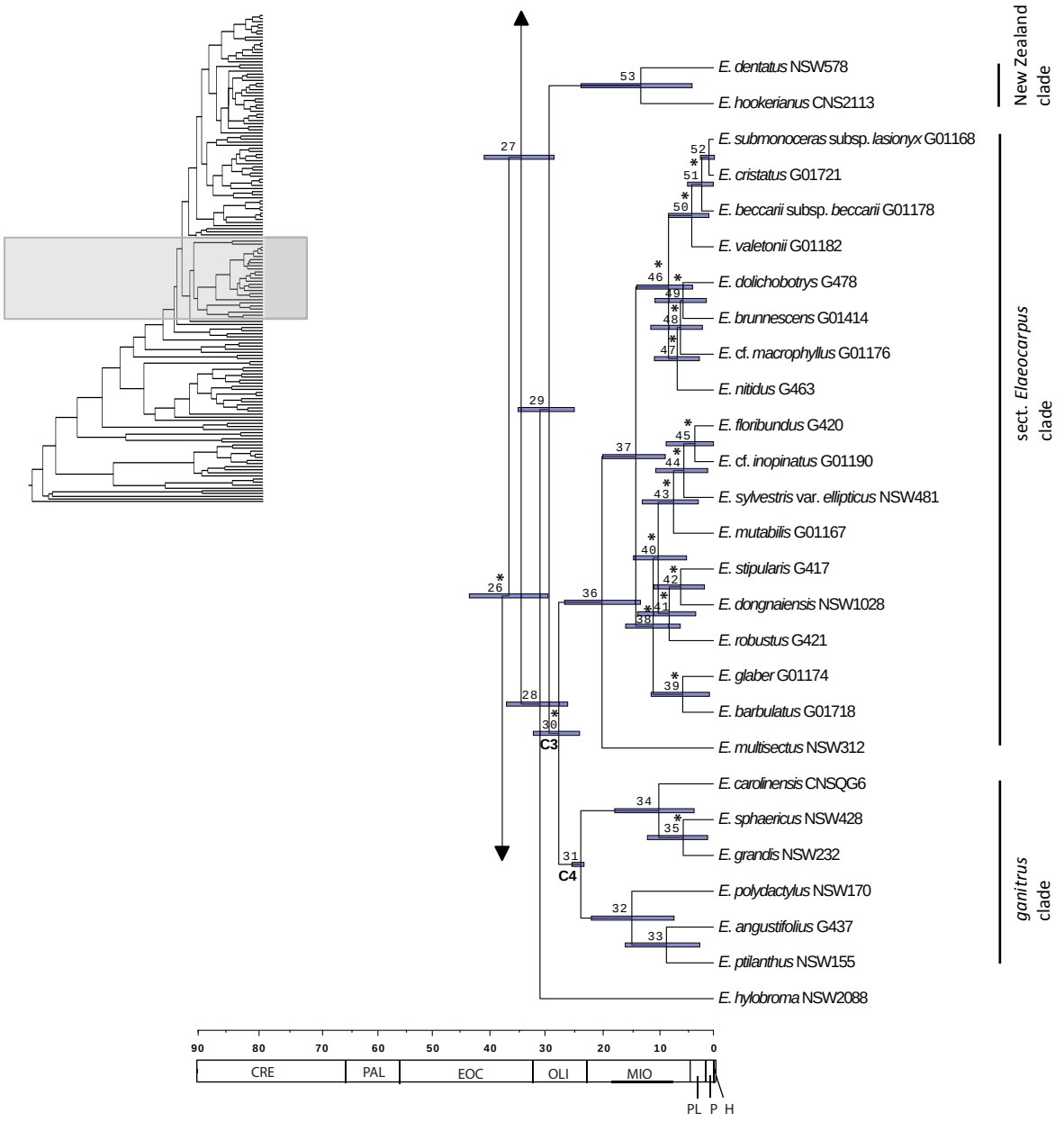


Fig. 3.3 (continued).

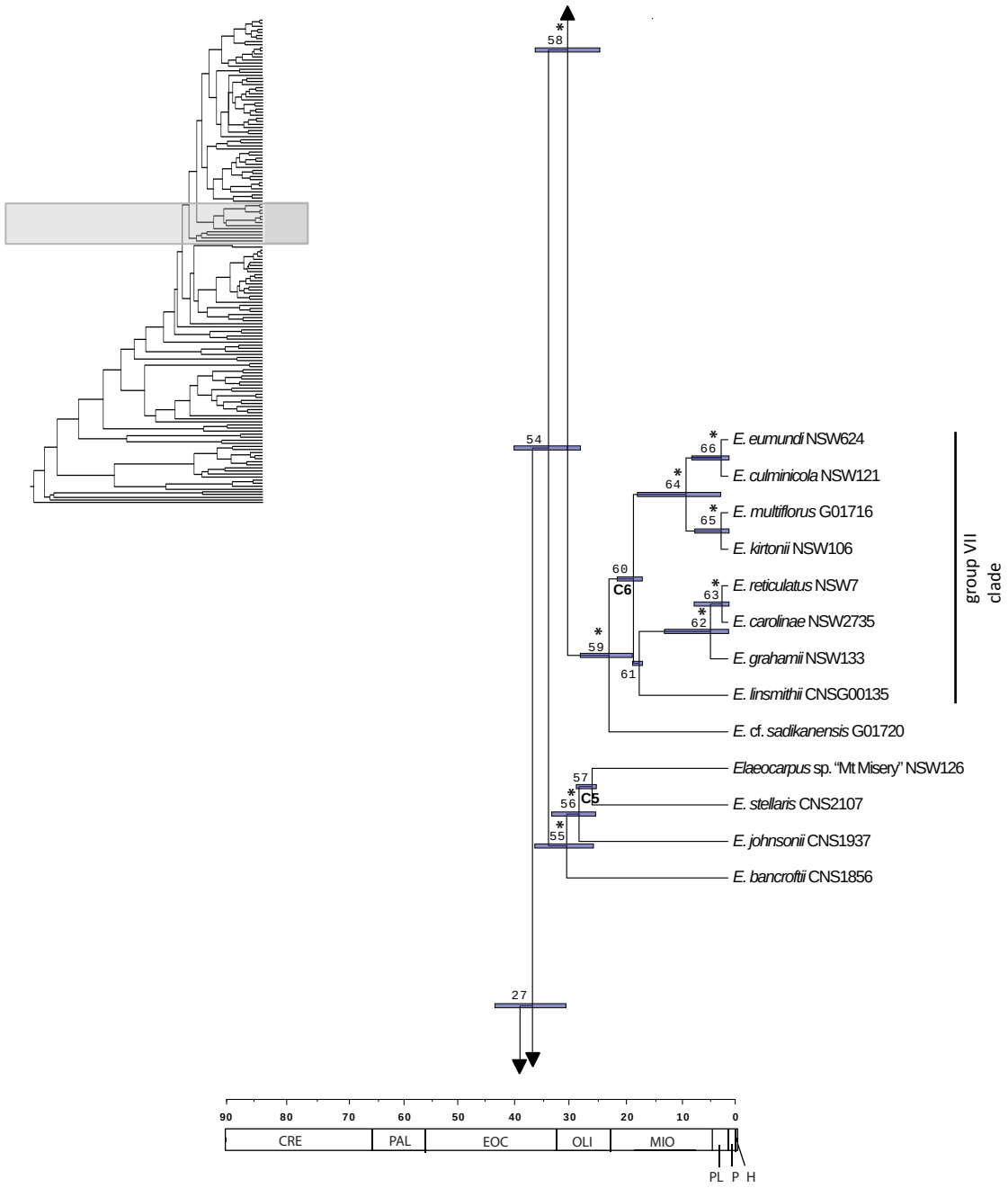


Fig. 3.3 (continued).

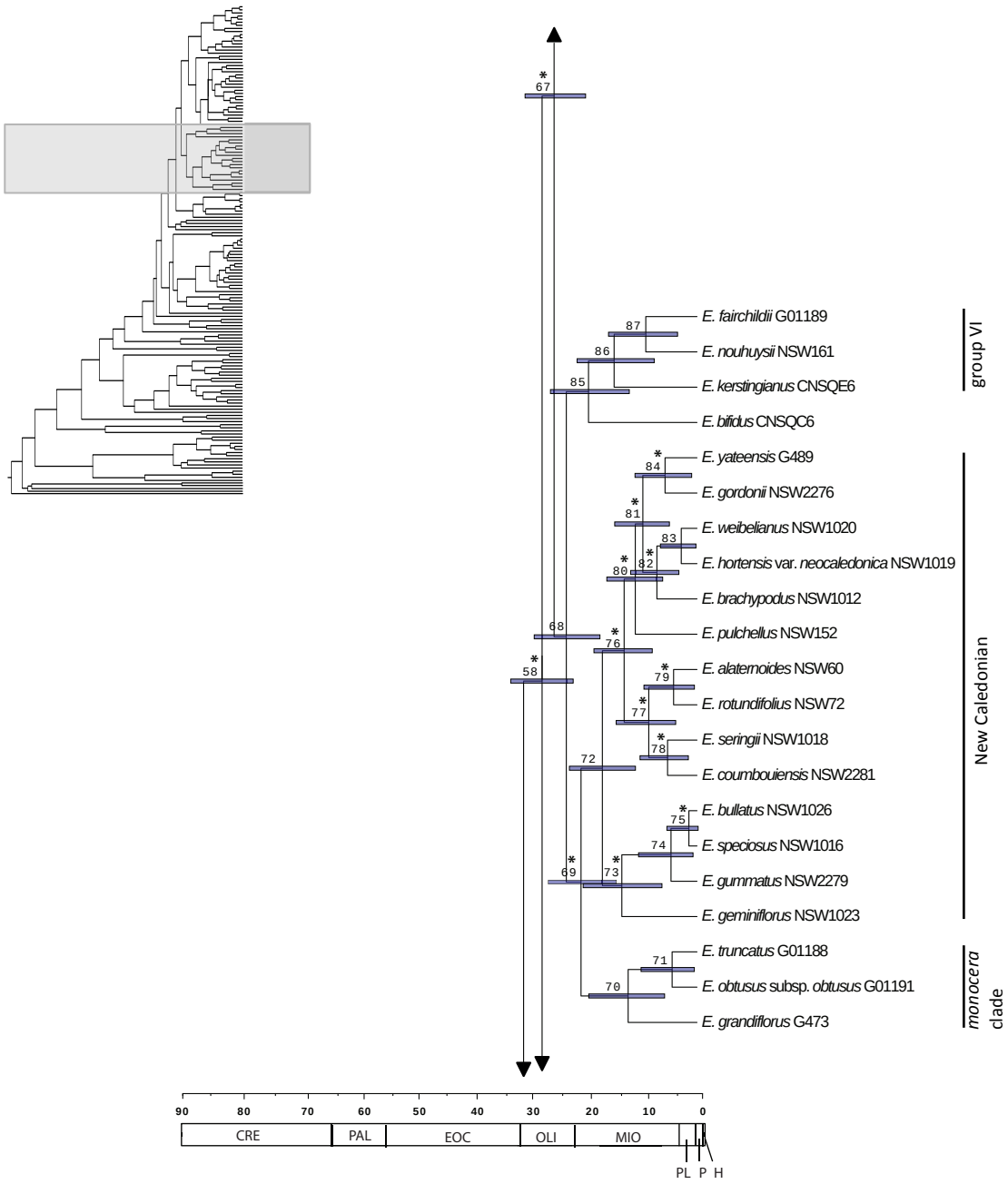


Fig. 3.3 (continued).

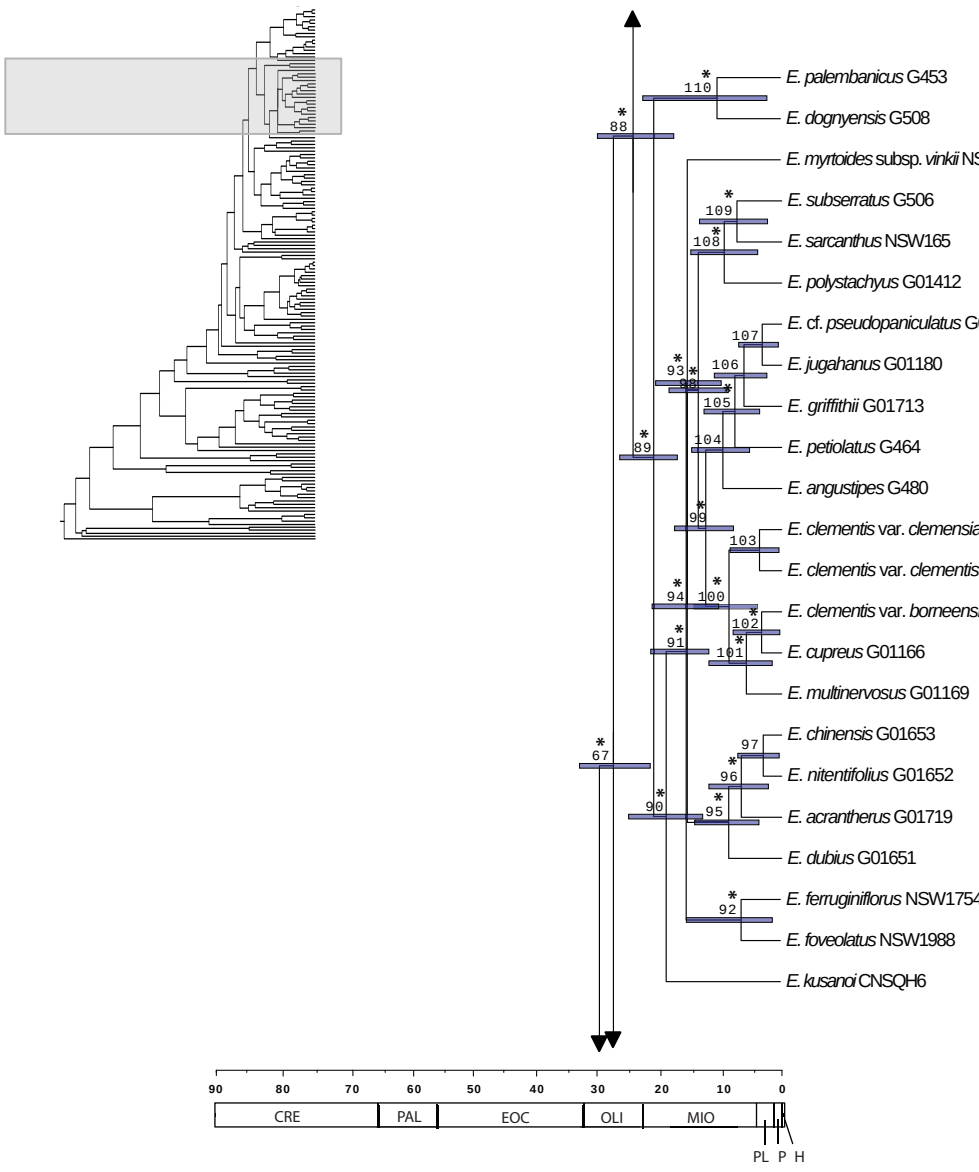


Fig. 3.3 (continued).

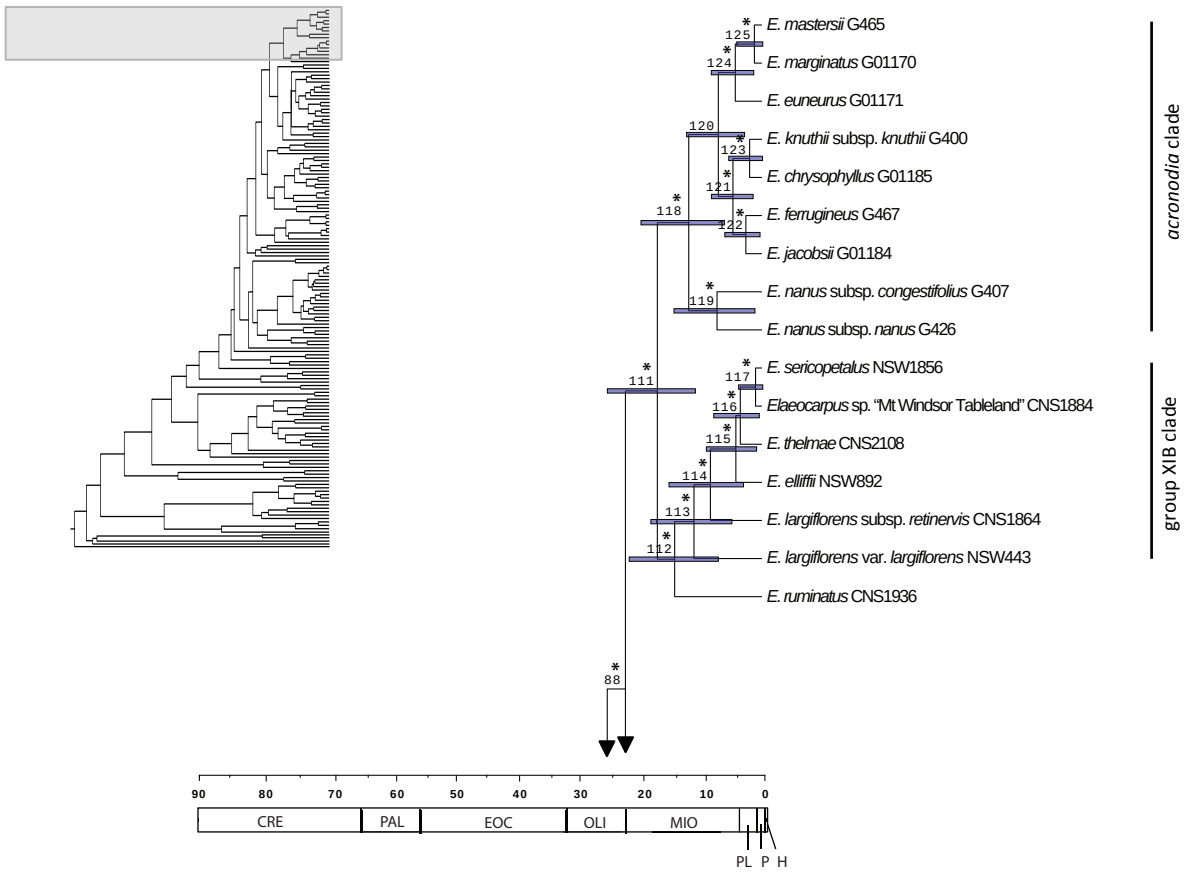


Fig. 3.3 (continued).

Table 3.3 Divergence time estimates (Mya) and ancestral area reconstructions for *Elaeocarpus*.

Divergence time analyses were conducted on the combined data matrix using BEAST. Ancestral area reconstructions were done using the DEC (Dispersal-Extinction-Cladogenesis) model in Lagrange. Node numbers refer to Figure 3.1 & 3.2. Node ages and 95 % HPD intervals of divergence times are in millions of years before the present. Bayesian PP are indicated. LAGRANGE reconstructions are shown with relative probabilities. Biogeographical areas are indicated by capital letters (as defined in Figure 3.2). For the DEC reconstructions, the split format is [left|right], where "left" (the upper branch) and "right" (the lower branch) are the ranges inherited by each descendant branch. Biogeographical scenarios are shown only when the split is within 2 log-likelihood units of the maximum for each node in LAGRANGE. Previous estimated divergence times were: [a] extracted from Niissalo (2011) and [b] extracted from Heibl and Renner (2012), in which the MCMC method implemented in BEAST was used; or [c] extracted from Crayn *et al*. (2006), in which the Langley-Fitch method implemented in r8s was used. Abbreviations and symbols: millions of years before the present (Mya); < less than; ^ estimated from the node topology; - not estimated.

Node	BEAST estimates			DEC ancestral area reconstruction			Previous BEAST estimates[a/b] (Mya)	Previous r8s estimates[c] (Mya)
	Node age (Mya)	95 % HPD (Mya)	Bayesian (PP)	Split	–ln*L*	Relative probabilities		
0	85.82	89.52 – 81.81	1	-	-	-	- / -	-
1	82.98	88.41 – 75.09	1	-	-	-	77^ / 59^	118 ± 8
2	58.71	78.26 – 40.26	0.99	-	-	-	63 (fixed) / -	89 ± 6
3	37.75	58.81 – 18.95	0.98	-	-	-	39^ / 25^	56 ± 4
4	14.56	28.15 – 4.85	1	-	-	-	18^ / -	27 ± 2
5	29.22	32.60 – 28.03	1	-	-	-	38^ / -	29 ± 2
6	76.51	88.41 – 75.09	1	-	-	-	66.5^ / 45^	103 ± 7
7	59.64	75.28 – 41.92	0.99	-	-	-	53^ / 37^	91 ± 6
8	20.40	33.95 – 7.83	1	-	-	-	25^ / -	60 ± 4
9	10.75	22.00 – 2.90	1	-	-	-	10^ / -	19 ± 2
10	64.88	73.53 – 55.88	1	-	-	-	- / 39^	-
11	25.96	40.91 – 13.71	1	-	-	-	38^ / -	40 ± 3
12	57.57	66.77– 48.16	1	-	-	-	52^ / 33^	63^
13	49.55	59.23 – 39.55	1	-	-	-	- / 24^	-
14	45.18	55.25 – 35.68	< 0.95	-	-	-	- / 16^	-
15	37.29	46.43 – 27.98	1	-	-	-	26^ / -	19 ± 1
16	6.90	23.64 – 0.15	1	-	-	-	- / -	-

17	48.88	57.85 – 40.12	< 0.95	F|E	223.6	0.98	38^ / -	33 ± 2
18	10.83	20.17 – 3.74	1	E|E	223.5	1	8^ / -	12 ± 1
19	45.46	53.63 – 37.22	< 0.95	F|F	223.6	0.98	- / -	-
20	24.00	34.30 – 14.47	1	F|F	223.5	1	13^ / -	18 ± 1
21: *Elaeocarpus* crown	39.15	46.45 – 32.41	1	F|F	223.6	0.98	30 (fixed) (including *Elaeocarpus* sp. "Rocky Creek" / -	30 (fixed) (excluding *Elaeocarpus* sp. "Rocky Creek"
22	34.72	40.55 – 29.11	1	F|F	223.6	0.97	- / -	-
23: crown of the *E. coorangooloo* + *obovatus* clade	18.27	25.97 – 11.19	0.99	F|F	223.6	0.97	- / -	-
24	13.38	20.85 – 6.92	< 0.95	F|F	223.6	0.91	- / -	-
25	8.97	16.57 – 2.74	0.96	F|F	223.5	1	- / -	-
26	34.72	40.55 – 29.11	< 0.95	F|F	223.6	0.96	- / -	-
27	31.61	36.65 – 27.41	0.99	F|F	223.7	0.84	- / -	-
28	29.27	33.70 – 25.38	1	F|F EG|F	224.8 225.3	0.28 0.18	- / -	-
29	28.14	32.30 – 24.38	1	G|E G|EF G|F	224.7 225.1 225.2	0.32 0.22 0.18	- / -	-
30	27.45	31.93 – 23.78	< 0.95	E|EF E|E	224.7 224.7	0.32 0.30	- / -	-
31: crown of the *ganitrus* clade	23.47	24.71 – 23.00	1	E|E EF|E	224.7 225.3	0.30 0.16	- / -	-
32	14.06	20.24 – 7.42	1	E|E	224.3	0.48	- / -	-
33	8.87	14.89 – 3.67	1	E|E	225.1	0.20	- / -	-
34	7.80	13.94 – 2.93	1	H|BCDEF	225.7	0.12	- / -	-
35	4.14	8.61 – 0.88	< 0.95	BCDEF|F	224.7	0.30	- / -	-
36: crown of the sect. *Elaeocarpus* clade	19.77	25.40 – 13.95	1	C|E	223.6	0.91	- / -	-
37	12.28	17.17 – 7.95	1	C|C	223.7	0.87	- / -	-
38	9.63	13.89 – 5.95	< 0.95	C|C	223.7	0.83	- / -	-
39	5.62	11.18 – 0.90	< 0.95	C|C	223.7	0.85	- / -	-

				CD|C	225.5	0.14		
40	9.94	14.29 – 4.94	< 0.95	C|C BC|B	223.9 225.5	0.73 0.14	- / -	-
41	7.34	12.03 – 3.11	< 0.95	C|C BC|B B|B	224.6 224.8 225.4	0.36 0.28 0.15	- / -	-
42	5.41	9.71 – 1.97	< 0.95	BC|B B|B	224.3 224.5	0.45 0.40	- / -	-
43	7.23	12.74 – 2.87	< 0.95	C|C C|BC	224 224.7	0.63 0.32	- / -	-
44	5.42	10.36 – 1.22	< 0.95	C|B	223.7	0.89	- / -	-
45	3.48	8.54 – 0.19	< 0.95	C|C CD|C	223.8 225.2	0.81 0.18	- / -	-
46	8.12	13.73 – 3.88	< 0.95	C|C	223.6	0.91	- / -	-
47	6.59	10.61 – 2.68	< 0.95	C|C	223.7	0.84	- / -	-
48	6.02	11.24 – 2.15	< 0.95	C|C	223.6	0.99	- / -	-
49	5.56	10.50 – 1.46	< 0.95	C|C	223.6	0.97	- / -	-
50	3.03	5.79 – 0.83	< 0.95	C|C	223.6	0.97	- / -	-
51	1.73	3.54 – 0.28	< 0.95	C|C	223.6	0.93	- / -	-
52	0.80	1.83 – 0.01	1	C|C CD|C	223.9 224.9	0.73 0.26	- / -	-
53: crown of the New Zealand clade	13.13	21.90 – 5.25	1	G|G	223.5	1	- / -	-
54	28.96	33.55 – 24.67	1	F|F	223.7	0.85	- / -	-
55	26.61	30.60 – 23.40	< 0.95	F|F	223.5	1	- / -	-
56	26.00	30.74 – 23.11	< 0.95	F|F	223.5	1	- / -	-
57	23.60	25.49 – 23.00	1	F|F	223.5	1	- / -	-
58	25.29	29.89 – 20.99	< 0.95	C|CF F|F	224 225.5	0.65 0.14	- / -	-
59	20.51	24.54 – 16.95	< 0.95	F|C	223.7	0.82	- / -	-
60: crown of the group VII clade	16.31	18.64 – 15.05	1	F|F	223.8	0.75	- / -	-
61	15.48	16.46 – 15.00	1	F|F	223.5	1	- / -	-

62	3.22	11.2 – 0.07	< 0.95	F|F	223.5	1	- / -	-
63	1.22	6.05 – 0	< 0.95	F|F	223.5	1	- / -	-
64	7.62	15.42 – 1.83	< 0.95	F|F	225	0.22	- / -	-
65	1.36	5.72 – 0	< 0.95	BDE|F	223.8	0.74	- / -	-
66	1.49	5.88 – 0	< 0.95	F|DEF F|EF F|DF	224.1 225.4 225.4	0.57 0.15 0.15	- / -	-
67	23.61	27.87 – 19.30	< 0.95	C|C C|CF	224.1 225.5	0.59 0.14	- / -	-
68	22.21	27.05 – 17.29	1	H|CH	223.7	0.82	- / -	-
69	20.00	25.22 – 15.00	< 0.95	H|C	223.6	0.90	- / -	-
70: crown of the *monocera* clade	12.57	18.55 – 6.76	1	C|C	223.8	0.76	- / -	-
71	4.07	8.39 – 0.80	1	C|C C|CD	223.9 224.8	0.69 0.29	- / -	-
72: crown of the New Caledonia clade	16.92	22.37 – 11.69	0.98	H|H	223.5	1	- / -	-
73	13.68	20.61 – 6.48	< 0.95	H|H	223.5	1	- / -	-
74	4.21	8.68 – 0.85	1	H|H	223.5	1	- / -	-
75	1.65	5.60 – 0	< 0.95	H|H	223.5	1	- / -	-
76	12.55	17.32 – 8.10	< 0.95	H|H	223.5	1	- / -	-
77	8.85	14.69 – 4.02	< 0.95	H|H	223.5	1	- / -	-
78	5.48	10.41 – 1.75	< 0.95	H|H	223.5	1	- / -	-
79	4.40	9.71 – 0.67	< 0.95	H|H	223.5	1	- / -	-
80	11.27	16.37 – 6.36	< 0.95	H|H	223.5	1	- / -	-
81	9.63	14.10 – 5.55	< 0.95	H|H	223.5	1	- / -	-
82	7.24	11.51 – 3.74	< 0.95	H|H	223.5	1	- / -	-
83	2.83	6.04 – 0.27	0.97	H|H	223.5	1	- / -	-
84	6.23	11.08 – 1.40	< 0.95	H|H	223.7	0.87	- / -	-
85	19.23	24.56 – 13.88	0.98	E|H	223.5	1	- / -	-
86: crown of the group VI clade	15.89	21.26 – 10.74	1	E|E	223.5	1	- / -	-
87	10.86	16.64 – 5.24	0.99	E|E	223.5	1	- / -	-
88	22.81	28.29 – 16.57	< 0.95	C|C	224	0.60	- / -	-

				CF|C	225.2	0.18		
89	19.93	24.75 – 14.90	< 0.95	C|C C|CE	224.2 225.3	0.53 0.16	- / -	-
90	15.39	20.10 – 10.76	< 0.95	CE|E CEF|E	224.1 224.6	0.54 0.34	- / -	-
91	13.47	17.61 – 9.83	< 0.95	CE|F	223.7	0.85	- / -	-
92	6.23	14.63 – 1.44	< 0.95	F|F	223.5	1	- / -	-
93	12.70	17.37 – 8.90	< 0.95	E|CE E|C	224.3 224.5	0.48 0.36	- / -	-
94	12.44	16.50 – 8.58	< 0.95	CE|C C|C	224.3 224.5	0.48 0.36	- / -	-
95	7.69	11.77 – 3.96	< 0.95	BC|B	223.6	0.97	- / -	-
96	5.83	9.60 – 2.13	< 0.95	B|C	223.5	1	- / -	-
97	3.12	6.45 – 0.54	1	B|B	223.5	1	- / -	-
98	12.8	17.29 – 8.20	< 0.95	CE|C C|C	224.3 224.5	0.45 0.38	- / -	-
99	11.65	16.44 – 7.40	< 0.95	C|C	223.6	0.99	- / -	-
100: crown of the *polystachyus* clade	7.02	11.03 – 3.35	< 0.95	C|C	223.5	1	- / -	-
101	4.90	9.43 – 1.54	< 0.95	C|C	223.5	1	- / -	-
102	2.85	6.34 – 0.34	< 0.95	C|C	223.5	1	- / -	-
103	3.31	6.65 – 0.56	0.99	C|C	223.5	1	- / -	-
104: crown of the *coilopetalum* clade	8.22	11.93 – 4.97	1	C|C	223.6	0.97	- / -	-
105	7.18	11.92 – 3.38	< 0.95	C|C	223.7	0.90	- / -	-
106	5.57	8.73 – 2.28	0.99	C|C	223.5	1	- / -	-
107	3.19	6.17 – 0.68	0.99	C|C	223.5	1	- / -	-
108	8.82	13.95 – 3.68	< 0.95	AE|C C|E	224 224.9	0.63 0.25	- / -	-
109	6.45	10.12 – 1.33	< 0.95	A|E	223.5	1	- / -	-
110	10.39	19.87 –2.81	< 0.95	C|H	223.5	1	- / -	-
111	15.81	21.43 – 9.74	< 0.95	C|F	223.5	1	- / -	-

112	12.89	18.88 – 7.65	< 0.95	F|F	223.5	1	- / -	-
113: crown of the group XIB clade	10.36	15.58 – 5.49	< 0.95	F|F	223.5	1	- / -	-
114	7.31	12.26 – 3.31	< 0.95	F|F	223.5	1	- / -	-
115	4.47	9.36 – 1.03	< 0.95	F|F	223.5	1	- / -	-
116	3.24	6.76 – 0.73	< 0.95	F|F	223.5	1	- / -	-
117	1.15	3.48 – 0.01	< 0.95	F|F	223.5	1	- / -	-
118: crown of the *acronodia* clade	10.52	16.53 – 5.61	< 0.95	C|C	223.5	1	- / -	-
119	7.6	14.74 – 1.30	< 0.95	C|C	223.5	1	- / -	-
120	5.92	9.96 – 2.78	1	C|C	223.5	1	- / -	-
121	4.96	8.53 – 1.61	< 0.95	C|C	223.5	1	- / -	-
122	2.47	5.21 – 0.41	< 0.95	C|C	223.5	1	- / -	-
123	2.18	5.65 – 0.04	< 0.95	C|C	223.5	1	- / -	-
124	4.56	8.55 – 1.50	< 0.95	C|C	223.5	1	- / -	-
125	1.24	3.89 – 0	< 0.95	C|C	223.5	1	- / -	-

Within Elaeocarpaceae, two genera are represented by a single sample thus only stem ages are available: *Vallea* arose at *c*. 37.75 Mya (Fig. 3.3 & Table 3.3: node 3; HPD: 58.81 – 18.95 Mya, PP = 0.98) and *Platytheca* at *c*. 45.18 Mya (node 14; HPD: 55.25 – 35.68 Mya, PP < 0.95). For the remaining genera estimated crown ages are as follows: *Aristotelia c*. 14.56 Mya (node 4; HPD: 28.15 – 4.85 Mya, PP = 1), *Sloanea c*. 29.22 Mya (node 5; HPD: 32.60 – 28.03 Mya, PP = 1), *Crinodendron c*. 20.40 Mya (node 8; HPD: 33.95 – 7.83 Mya, PP = 1), *Peripentadenia c*. 10.75 Mya (node 9; HPD: 22.00 – 2.90 Mya, PP = 1), *Dubouzetia c*. 25.96 Mya, (node 11; HPD: 40.91 – 13.71 Mya, PP = 1), *Tetratheca c*. 37.29 Mya (node 15; HPD: 46.43 – 27.98 Mya, PP = 1), *Tremandra c*. 6.90 Mya (node 16; HPD: 23.64 – 0.15 Mya, PP = 1), *Sericolea c*. 10.83 Mya (node 18; HPD: 20.17 – 3.74 Mya, PP = 1), *Aceratium c*. 24.00 Mya (node 20; HPD: 34.30 – 14.47 Mya, PP = 1) and *Elaeocarpus c*. 39.15 Mya (node 21; HPD: 46.45 – 32.41 Mya, PP = 1). Each of these clades received strong statistical support (PP = 0.98 – 1).

The molecular clock analysis estimated that diversification within *Elaeocarpus* was initiated at *c*. 39.15 Mya (Fig. 3.3 & Table 3.3: node 21; HPD: 46.45 – 32.41 Mya, PP = 1) with the split between the *E. holopetalus* lineage, and the remainder of the genus. Within the latter, 13 labelled sub-lineages or clades are resolved in this study: *E. coorangooloo* + *obovatus* (node 23) (*E. coorangooloo* is shown to be sister to the *obovatus* clade, but based on morphological evidence, the former is hypothesised to be more closely related to the Papusian *E. meigei* Weibel and *E. hartleyi* Weibel; details discussed in Chapter 2, section 2.4.4), *E. sedentarius* (node 26), the *ganitrus* (node 31), sect. *Elaeocarpus* (node 36), New Zealand (node 53), group VII (node 60), *monocera* (node 70), New Caledonian (node 72), group VI (node 86), *polystachyus* (node 100), *coilopetalum* (node 104), group XI subgroup B (node 113) and *acronodia* (node 118) clades (naming follows Chapter 2). Among these, the *polystachyus* clade was estimated to have the youngest crown age, *c*. 7.02 Mya in the Miocene (HPD: 11.03 – 3.35 Mya). The node ages, HPD intervals and Bayesian PP value of all nodes within *Elaeocarpus* are indicated in Fig. 3.3 and summarised in Table 3.3.

3.3.3 *Historical biogeography of* Elaeocarpus

The Fitch parsimony reconstruction required 22 steps to explain the present-day distribution of *Elaeocarpus*. The dispersal and local extinction rate of the DEC reconstruction are 0.003872 and 4.285 x 10^{-9} events per million of years, respectively. The inferred ancestral distributions from the Fitch parsimony and DEC reconstructions are shown on the maximum clade credibility chronogram generated from BEAST (Fig. 3.4), whereas the results of the DEC are summarised in Table 3.3.

Both Fitch parsimony and DEC reconstructions placed the most likely ancestral area for *Elaeocarpus* on Australia (area F) (Fig. 3.4: node 21). Diversification of *Elaeocarpus* gave rise to 14 labelled lineages or clades, with some tending to be geographically based. Within Australia, diversification of the genus gave rise to the *E. holopetalus* lineage and the clades *E. coorangooloo* + *obovatus* (node 23), group VII (node 60), *E. ruminatus* (node 112) and group XI subgroup B (XIB, node 113). Diversification in New Guinea (area E) gave rise to the *ganitrus* (node 28) and group VI (node 86) clades; diversification in West Malesia gave rise to the sect. *Elaeocarpus* (node 36), the *monocera* (node 70), the *polystachyus* (node 100), the *coilopetalum* (node 104) and the *acronodia* (node 120) clades; diversification in New Zealand occurred in the New Zealand clade (node 53); and diversification in the Pacific islands, particularly in New Caledonia, occurred in the New Caledonian clade (node 72). Radiations in New Guinea and West Malesia have made the two areas the modern-day centres of diversity for the genus. It is noteworthy, however, that some nodes on the backbone and terminal branches lack statistical support, which may explain the minor incongruence between the results of Fitch parsimony and DEC.

Within the *E. coorangooloo* + *obovatus*, sect. *Elaeocarpus*, group VII, *monocera* and *coilopetalum* clades, northwards dispersal from Australia to New Guinea, New Guinea to West Malesia, and West Malesia to the Continental Asia is inferred. On the other hand, southwards and/or eastwards dispersal from Australia to New Zealand, and West Malesia to Central Malesia is also inferred, in the case of the two New Zealand endemics (*E. dentatus* (J.R. & G.Forst.) Vahl and *E. hookerianus* Raoul (Fig. 3.4:

node 53) and seven primarily Malesian species, *E. glaber* Blume (node 39), *E. stipularis* (node 42), *E. brunnescens* R.Knuth (node 49), *E. submonoceras* Miq. subsp. *lasionyx* (Stapf ex Ridl.) Weibel (node 52), *E. grandiflorus* Sm. (node 70), *E. obtusus* Blume subsp. *obtusus* (node 71) and *E. petiolatus* (node 105), respectively. The ancestral area of *E. subserratus* Baker (terminal node 109), the sole Madagascan species, is not inferred because its phylogenetic placement lacks statistical support.

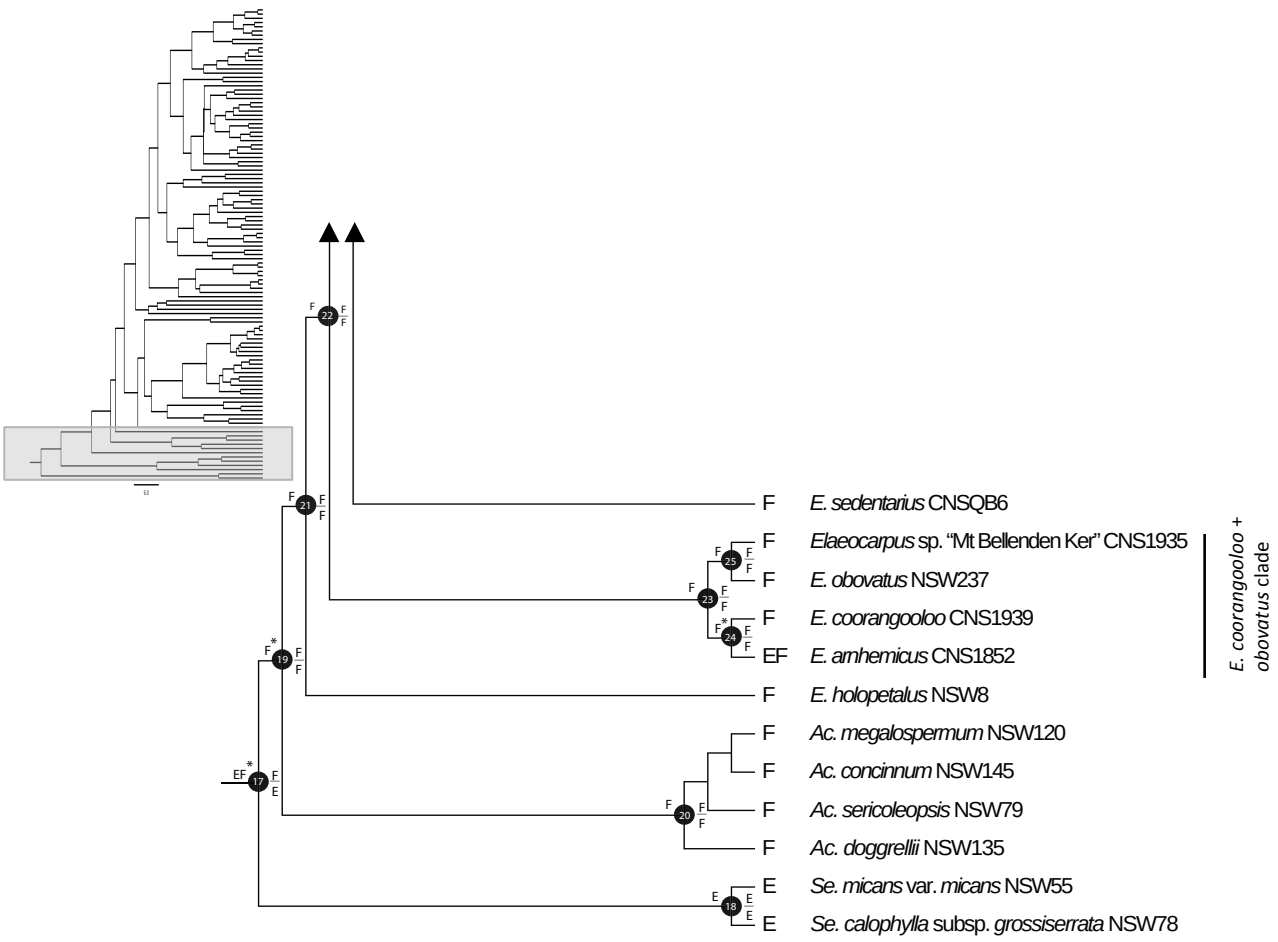


Fig. 3.4 Ancestral area reconstruction for *Elaeocarpus*. Letters on the left of the nodes represent Fitch parsimony results, and letters on the right are Dispersal-Extinction-Cladogenesis results (see Materials and methods). For the latter, all possible ancestral range subdivision/inheritance scenarios are presented in Table 3.3. Letters to the left of the taxa names represent current distribution (as defined in Fig 3.2).

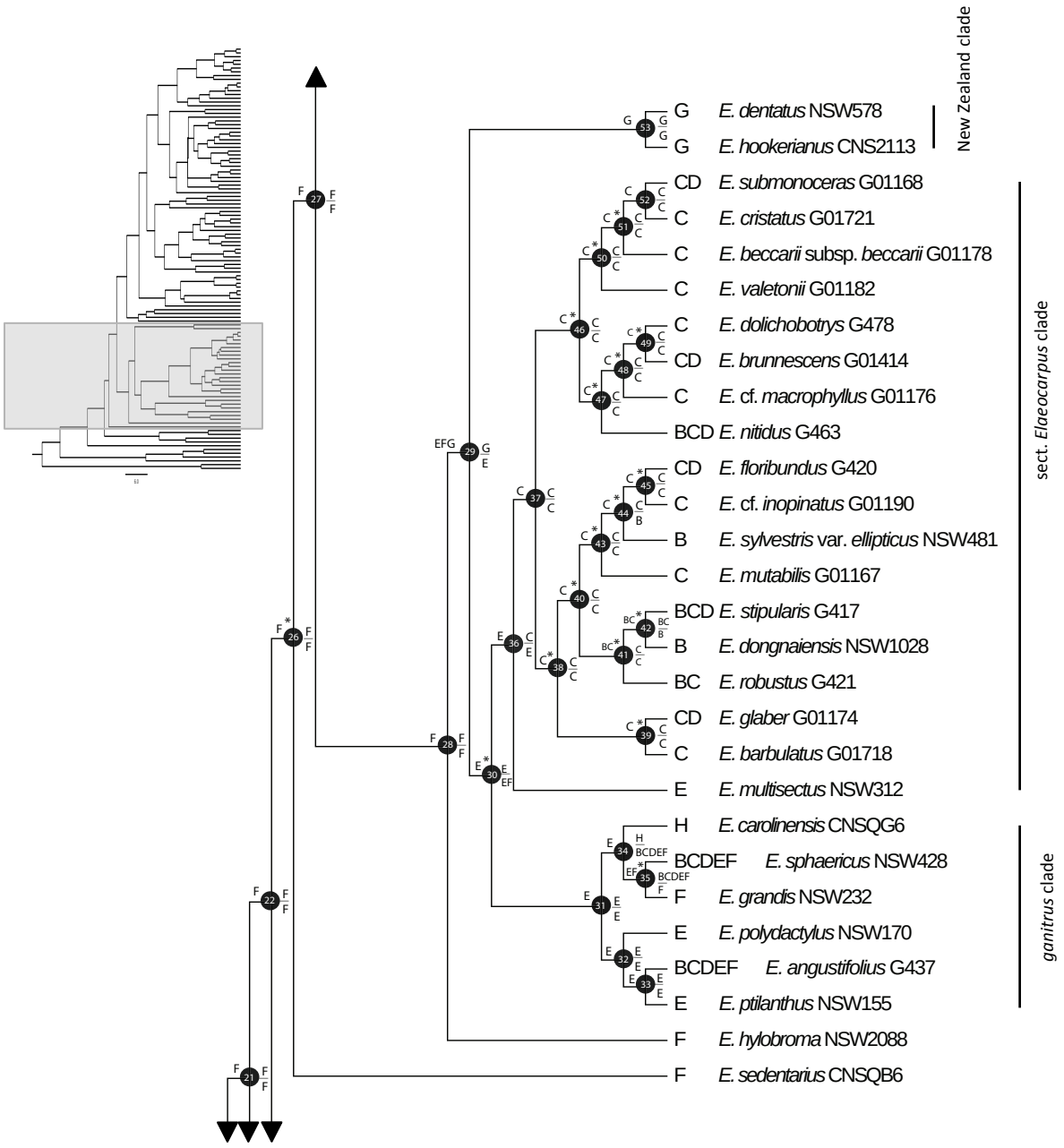


Fig. 3.4 (continued).

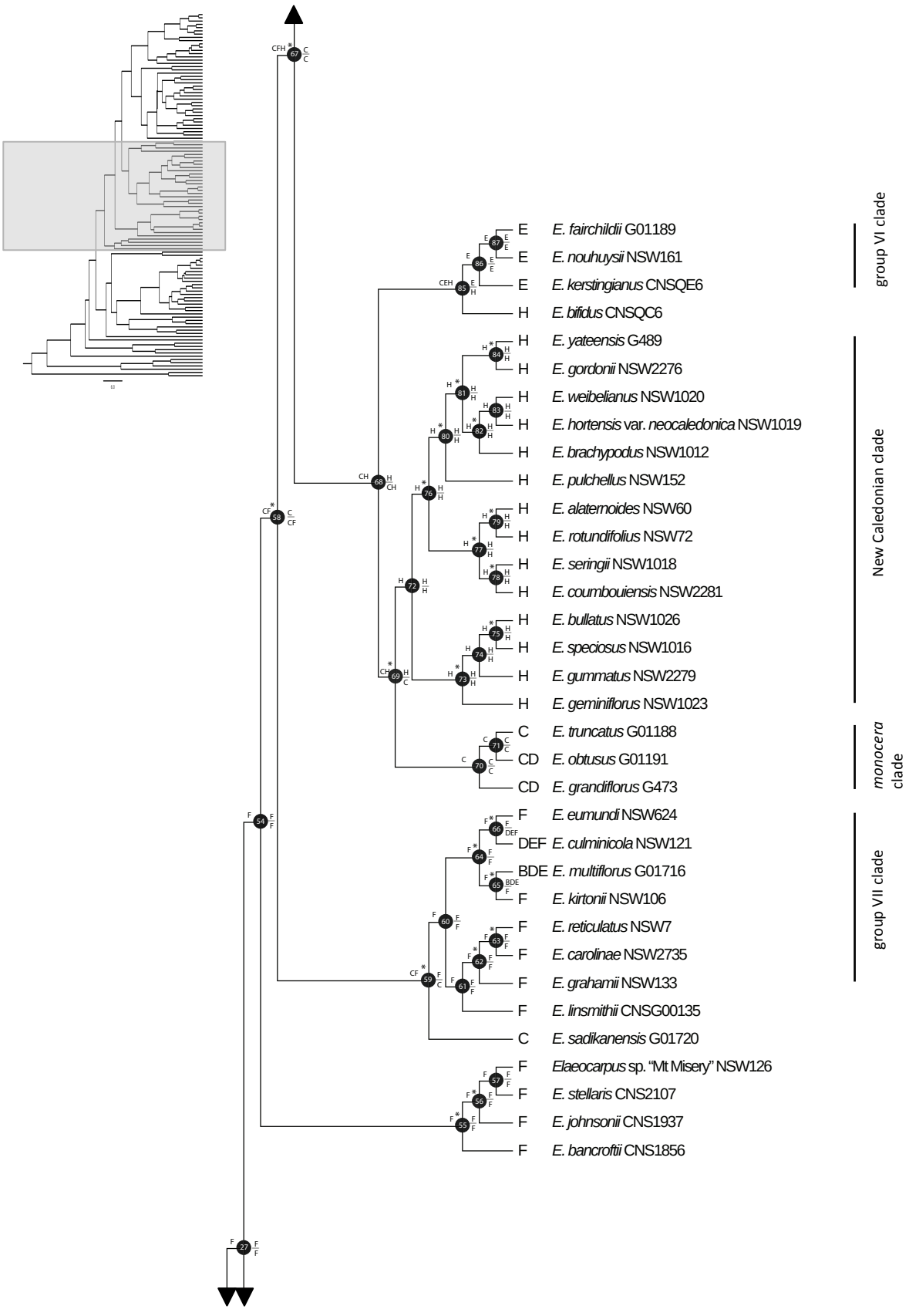


Fig. 3.4 (continued).

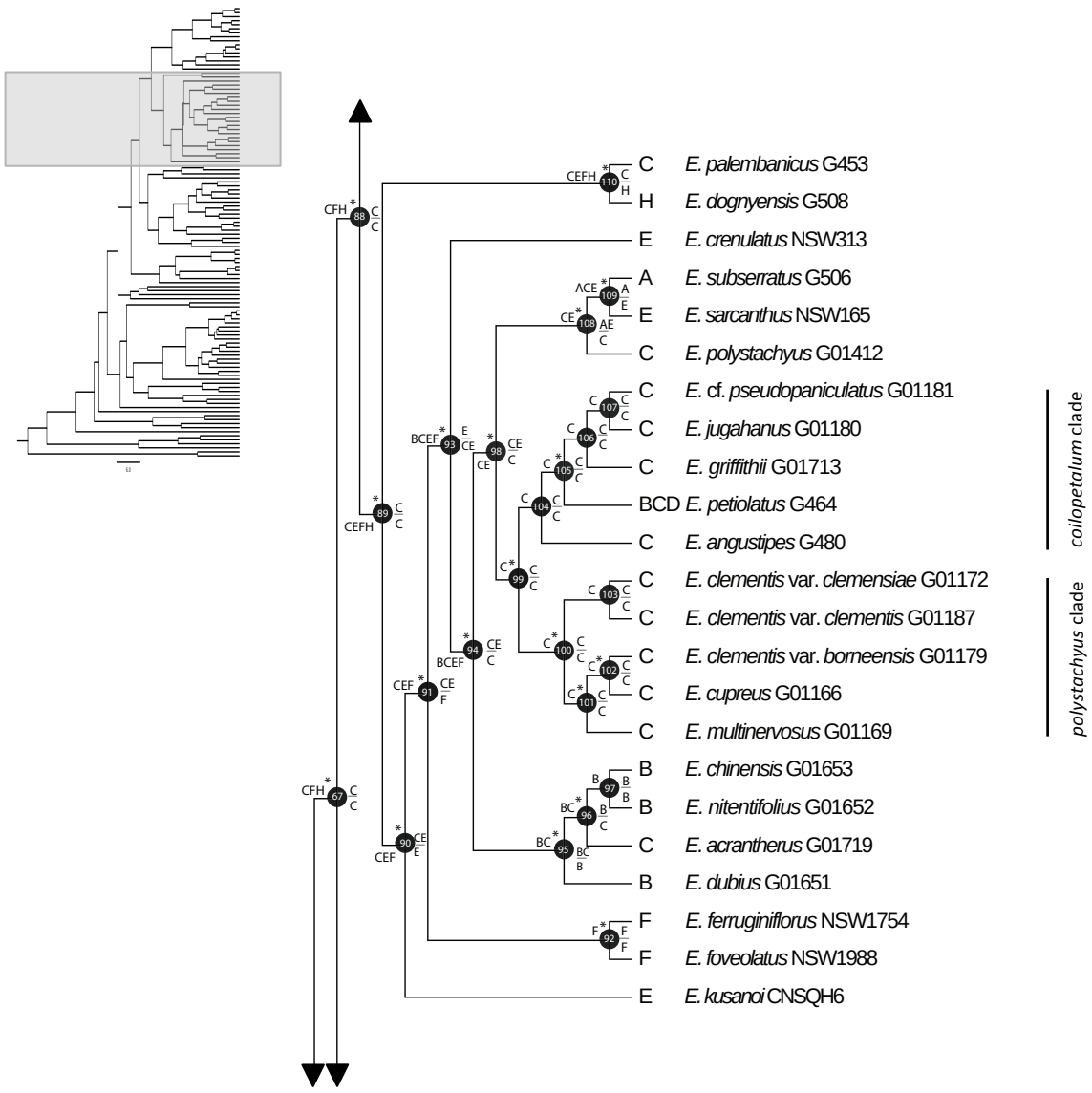


Fig. 3.4 (continued).

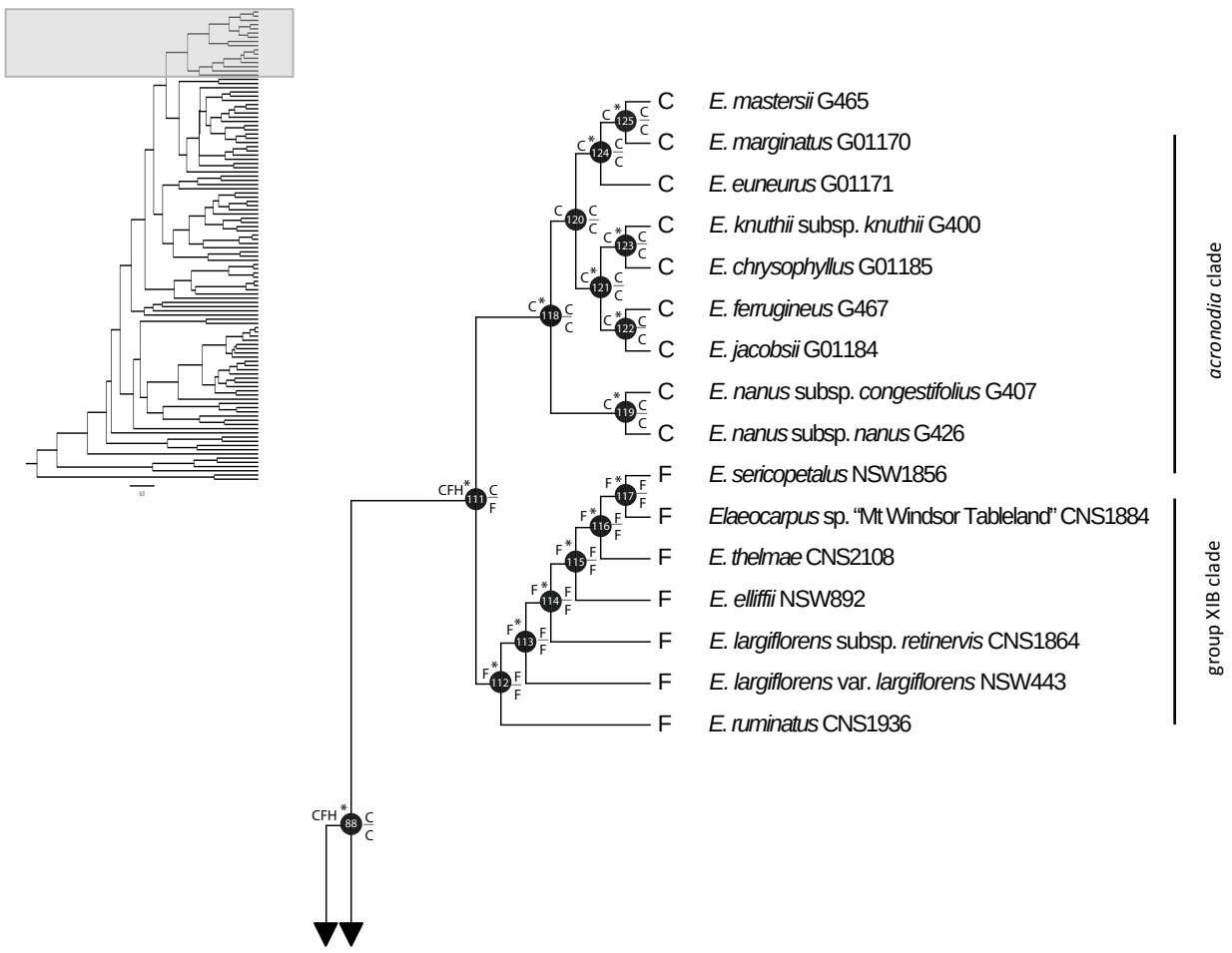


Fig. 3.4 (continued).

3.4 Discussion

3.4.1 Divergence times of the main lineages in Elaeocarpaceae

The molecular clock analysis resulted in a robust phylogenetic tree of the Elaeocarpaceae (Fig. 3.3). The estimated crown ages of Elaeocarpaceae are consistent with the age estimation of Magallón *et al.* (2015, *c*. 81.63 Mya, 95 % HPD: 79.62 Mya – 85.2 Mya), but much older than the estimates of Wikström *et al*. (2001) and Heibl and Renner (2012, estimated from the chronogram of the order Oxalidales), and younger than Crayn *et al*. (2006). Wikström *et al*. (2001) estimated the age of the stem group Elaeocarpaceae around 66 – 64 Mya and the crown group 59 – 57 Mya, whereas Heibl and Renner (2012) estimated 59 Mya for the crown group. In the latter study, however, the genus *Petenaea* (which belongs to a different order, Huertales but was originally described in Elaeocarpaceae) was included in the dataset, which was otherwise restricted to members of Oxalidales. Its placement in that analysis on a long branch within Elaeocarpaceae and sister to *Sloanea* is certainly spurious and likely to have distorted the divergence date estimates.

Incorrect phylogenetic assignment of a fossil taxon used as a calibration point could lead to erroneous estimates of divergence times (Marshall 1990). Crayn *et al*. (2006) estimated the crown group age of Elaeocarpaceae at over 100 million years. That analysis used the divergence dates of Wikström *et al*. (2001) as an external calibration point together with an internal, fossil-based calibration, where the internal calibration point constrained the crown age of *Elaeocarpus* to 30 Mya based on the fossil records of *Elaeocarpus*. Unfortunately, the relationships within the *Elaeocarpus* alliance were not well resolved in that study. The age constraint on the crown of *Elaeocarpus* by including most, but not all, *Elaeocarpus* taxa could possibly underestimate the age estimation within the genus.

Clade age estimates can be sensitive to different calibration methods, particularly the choice of fossils and their supposed taxonomic affinity, as shown in other studies (Bell *et al*. 2005, 2010; Hug and Roger 2007; Sauquet *et al*. 2012; Tripp *et*

al. 2014). Fossil records are incomplete and inevitably laid down after the origin of the lineage they represent, thus divergence times inferred based on fossil calibrations are almost certainly underestimates. A recent empirical study by Sauquet *et al*. (2012) explored using a case study the sensitivity of dating analyses to fossil constraints. In their study, the crown group of Fagaceae was estimated to be 64.2 – 103.6 Mya in one of the scenarios, but the oldest fossil, *Fagus langevinii* Manchester & Dilhoff, attributable to an internal lineage of the family, could only justify a minimum age constraint of 47 Mya for the crown group of Fagaceae.

The estimates of the crown ages of Elaeocarpaceae and major labelled lineages within it are comparable with the previous estimate of Niissalo (2011) (see Table 3.3), but generally younger than the estimates of Crayn *et al.* (2006). For instance, the crown age of the family Elaeocarpaceae (Table 3.3: node 1; 82.98 Mya, HPD: 88.41 – 75.09 Mya, PP = 1), *Aristotelia* alliance (node 3; 37.75 Mya, HPD: 58.81 – 18.95 Mya, PP = 0.98), *Aristotelia* (node 4; 14.56 Mya, HPD: 28.15 – 4.85 Mya, PP = 1), *Crinodendron* (node 8; 20.40 Mya, HPD: 33.95 – 7.83 Mya, PP = 1) and *Peripentadenia* (node 9; 10.75 Mya, HPD: 22.00 – 2.90 Mya, PP = 1) are relatively consistent (between 0.75 – 5.98 Mya) with the time estimated in Niissalo (2011), but between 8.25 – 39.60 Mya younger than the estimation of Crayn *et al*. (2006). The crown age of the Tremandraceous genera (node 13; 49.55 Mya, HPD: 59.23 – 39.55 Mya, PP = 1) and *Sericolea* (Fig. 3.3 & Table 3.3: node 18; 10.83 Mya, HPD: 20.17 – 3.74 Mya, PP = 1) are, nevertheless, consistent with the results of both Crayn *et al*. (2006) and Niissalo (2011).

Noticeably, the estimated crown ages of *Dubouzetia* (node 11; 25.96 Mya, HPD: 40.91 – 13.71 Mya, PP = 1), *Aceratium* (node 20; 24.00 Mya, HPD: 34.30 – 14.47 Mya, PP = 1) and *Elaeocarpus* (node 21; 39.15 Mya, HPD: 46.45 – 32.41 Mya, PP = 1) are markedly different from the estimates of Crayn *et al*. (2006) and Niissalo (2011). The former estimated the crown age of *Dubouzetia* and *Aceratium* at *c*. 40 Mya and *c*. 18 Mya, respectively; and the latter at *c*. 38 Mya and *c*. 13 Mya, respectively; while the age of *Elaeocarpus* was fixed at 30 Mya in both of the studies based on the age of the macrofossils assigned to *Elaeocarpus* (Dettman and Clifford 2000), used as a

conservative estimate of the minimum crown age of the genus. On the other hand, estimation of the divergence time for the family Elaeocarpaceae in Heibl and Renner (2012) is far less comprehensive than Crayn *et al*. (2006) and Niissalo (2011). In the study of Heibl and Renner (2012), one unspecified taxon from each of the 12 genera of the family were included and used as outgroup for the estimation of the divergence times of the Chilean *Oxalis* using BEAST. Therefore, only the stem ages of these lineages were given and they were between 12 – 31 Mya younger than the results of the molecular clock analysis in the present study. These marked differences may be due to a range of factors including sampling differences between studies, choice of markers, phylogenetic resolution, and/or accuracy of the fossil calibrations of the molecular evolutionary rates. It is likely that of all the studies undertaken to date the present one provides the most accurate estimate of divergence dates in Elaeocarpaceae because:

(1) it contains by far the densest sampling of lineages within the family, and especially within *Elaeocarpus*. About 28 % of Elaeocarpaceae and 32.57 % of *Elaeocarpus* are studied here compared with about 9.09 % of Elaeocarpaceae and 1.75 % of *Elaeocarpus* in Crayn *et al*. (2006), about 11.81 % of Elaeocarpaceae and 1.75 % of *Elaeocarpus* in Niissalo (2011) and about 2.18 % of Elaeocarpaceae and 0.29 % of *Elaeocarpus* in Heibl and Renner (2012);

(2) it provides the greatest resolution of relationships. *Aceratium* and *Elaeocarpus* are each robustly supported as monophyletic but were not in the studies of Crayn *et al*. (2006) and Niissalo (2011);

(3) a larger number of macrofossils was used in this study and the phylogenetic placement of these was more thoroughly analysed. These fossils were able to be assigned with confidence to shallower nodes (nearer to the terminal branch) than in previous studies, in part because of the far greater number and diversity of lineages on the trees as a consequence of the much denser taxon sampling. Crayn *et al*. (2006) and Niissalo (2011) conservatively placed all *Elaeocarpus* fossils at the crown node of genus.

It has long been postulated that the family Elaeocarpaceae originated in the southern hemisphere (Raven and Axelrod 1974). The earliest divergence within the

Elaeocarpaceae appears to have occurred in the late Cretaceous at *c*. 85.82 Mya (Fig. 3.3 & Table 3.3: node 0; HPD: 89.52 – 81.81 Mya, PP = 1), which is broadly coincident with the time when the western (Africa and South America) and eastern (Australia, Antarctica, Madagascar and India) parts of Gondwana were separating (a process which was almost complete at about 90 – 85 Mya; Ali and Aitchison 2008).

Within the family, the *Aristotelia* alliance + *Sloanea* alliance clade is sister to the rest of the Elaeocarpaceae, consistent with the results of previous studies (Crayn *et al*. 2006; Niissalo 2011). However, the age of the split between this clade and other Elaeocarpaceae (*c*. 82.98 Mya, Fig. 3.3 & Table 3.3: node 1; HPD: 88.41 – 75.09 Mya, PP = 1) is slightly older than the age estimated (77 Mya) by Niissalo (2011), but much younger than that estimated by Crayn *et al*. (2006) (118 ± 8 Mya). The genera of this clade are currently distributed predominantly on Gondwanan landmasses: *Sloanea* occurs in the tropical forests in Central and South America, in Australia and New Caledonia, and extends northwards to Continental Asia and Madagascar, but it is absent in mainland Africa (Coode 2004); *Vallea* occurs in South America from Peru and Bolivia to slightly north of equator into Colombia and Venezuela (Coode 2004); and *Aristotelia* occurs in Chile, New Zealand and Australia (Coode 2004). The MRCA of the clade is estimated at *c*. 58.71 Mya (Fig. 3.3: node 2; HPD: 78.26 – 40.26 Mya, PP = 0.99) and the ancestors may have dispersed between western and eastern Gondwana from *c*. 82.98 Mya (Fig. 3.3: node 1; HPD: 88.41 – 75.09 Mya, PP = 1) to *c*. 58.71 Mya (Fig. 3.3: node 2; HPD: 78.26 – 40.26 Mya, PP = 0.99). Results of the present study agree with Coode's (1985) phylogenetic reconstruction that *Vallea* and *Aristotelia* are sisters. On the other hand, Coode (1985) considered the *Sloanea* alliance as an isolated group in the family, but the results of this molecular analysis show that the *Sloanea* alliance is sister to the *Aristotelia* alliance, which is consistent with the findings of Crayn *et al*. (2006).

Crinodendron is resolved in this study as sister to *Peripentadenia*. This clade is sister to a clade containing *Dubouzetia*, the Tremandraceous genera, and the *Elaeocarpus* alliance; the split between the two clades is estimated at *c*. 76.51 Mya (Fig. 3.3: node 6; HPD: 88.41 – 75.09 Mya, PP = 1), origin of the South America and

Australia, from *c*. 76.51 Mya (Fig. 3.3: node 6; HPD: 88.41 – 75.09 Mya, PP = 1) to *c*. 59.64 Mya (Fig. 3.3: node 7; HPD: 75.28 – 41.92 Mya, PP = 0.99), and the disjunction between these two genera is possibly through vicariance. *Crinodendron* and *Peripentadenia* are currently confined to the southern hemisphere but in geographically disparate regions: the former occurs in South America and the latter in northeastern Australia (Coode 2004). These two genera diverged *c*. 59.64 Mya (Fig. 3.3: node 7; HPD: 75.28 – 41.92 Mya, PP = 0.99), when Australia and South America were joined to Antarctica. I hypothesise that the ancestor(s) of the Crinodendron + Peripentadenia clade may have been widespread across east Gondwana from *c*. 76.51 Mya (Fig. 3.3: node 6; HPD: 88.41 – 75.09 Mya, PP = 1) to *c*. 59.64 Mya (Fig. 3.3: node 7; HPD: 75.28 – 41.92 Mya, PP = 0.99) and the disjunction between these two genera is a consequence of vicariance. *Dubouzetia* is resolved as sister to the Tremandraceous genera + *Elaeocarpus* alliance clade and the two clades split from each other at *c*. 64.88 Mya (Fig. 3.3: node 10; HPD: 73.53 – 55.88 Mya, PP = 1). The relationships between *Crinodendron*, *Dubouzetia* and *Peripentadenia* inferred by the present analysis are different from those of Coode (1987). In the latter, *Peripentadenia*, *Dubouzetia* and *Crinodendron* formed a clade – the *Crinodendron* alliance – based on the shared derived fruit type (i.e. dehiscent and thin-walled). Within this clade, *Dubouzetia* and *Crinodendron* are sisters based on the number of stamens (35 – 45 per flower), and possession of anthers with single apical pore or short slit and connective not separating the thecae. Comprehensive studies on the transformation of these morphological characters (fruit type, stamen number and anther structure and dehiscence) are lacking, thus the question as to whether they are synapomorphic for clades within the family has not yet been addressed.

The Tremandraceous genera are three small genera endemic to Australia – *Platytheca*, *Tetratheca* and *Tremandra* – which together contain *c*. 13.5 % (54 species) of the total Elaeocarpaceae species. The three genera formed a well-supported clade and the split between this clade and the *Elaeocarpus* alliance is estimated at *c*. 57.57 Mya (Fig. 3.3: node 12; HPD: 66.77 – 48.16 Mya, PP = 1). The MRCA of the clade is dated at *c*. 49.55 Mya (Fig. 3.3: node 13; HPD: 59.23 – 39.55 Mya, PP = 1). These genera are mostly weak woody shrubs or shrublets that are found in heaths,

sclerophyll woodlands and dry forests, with a few restricted to the semi-arid zones of western Australia. The findings of this study are similar to those of Crayn *et al*. (2006) and Niissalo (2011) in inferring that the clade and its sister (the *Elaeocarpus* alliance) diverged in the Palaeocene. Crayn *et al*. (2006) speculated that members of this clade probably colonised drier and/or more seasonal habitats when Australian rainforests contracted in the Palaeocene, and then adapted and radiated in these environments.

The *Elaeocarpus* alliance began to diverge between *c*. 57.57 Mya (Fig. 3.3: node 12; HPD: 66.77 – 48.16 Mya, PP = 1) and *c*. 48.88 Mya (node 17; HPD: 57.85 – 40.12 Mya, PP < 0.95). This is the largest clade in the family, comprising three genera – *Sericolea*, *Aceratium* and *Elaeocarpus* – which together contain over 70 % of all Elaeocarpaceae species. *Sericolea* is confined to New Guinea, whereas the latter two are distributed predominantly in the palaeo-tropical forests of the northern and southern hemisphere (but not mainland Africa). *Aceratium* occurs in Australia, New Guinea, Moluccas, Solomon Islands and New Hebrides (Coode 2004, pers. comm. 2014), and *Elaeocarpus* occurs in Australia, Malesia, Asia, Madagascar, New Zealand and the Pacific islands as far west as Hawaii (Coode 2004). The early diverging lineages of this clade are essentially Sahulian. The Sahul and Sunda shelves collided at *c*. 25 Mya (late Oligocene; Hall 2009) resulting in the accretion of crustal fragments and emergence of land to form the islands of Wallacea (Hall 2012), and uplift of mountains in West Malesia (Sumatra and Java), Wallacea (Sulawesi) and New Guinea (Ufford and Cloos 2005; Hall 2009). At the same time (late Oligocene-Miocene), Borneo experienced active mountain formation and deformation, which opened up new niches, probably permitting establishment and promoting speciation of the biota in these regions (Hall and Nichols 2002; Crayn *et al*. 2015, de Bruyn *et al*. 2014). Therefore, the lineages now present on Sunda may have migrated north- and westwards from Sahul, a process likely facilitated by the consequences of the Sahul-Sunda collision (changing configurations of land and sea, and orogeny) and changes in sea level. Together, these effects may have encouraged subsequent diversification on the Sunda shelf, particularly mountain-building in New Guinea and Borneo that opened up extensive new niches and dynamic habitats available for later colonisation and speciation. Higher speciation rates of montane plants are evident in some plant

groups, such as the South American Bromeliaceae (Givnish *et al*. 2014), the globally distributed Ericaceae (Schwery *et al*. 2014) and the neo- and palaeo-tropical and -subtropical *Impatiens* L. (Balsaminaceae) (Janssens *et al*. 2009), compared with their non-montane relatives. Global climatic oscillations took place in the late Palaeocene to early Eocene and peaked around the early Eocene Climatic Optimum (52 – 50 Mya) (Zachos *et al*. 2001). Morley (2003) postulated that these climatic changes promoted northward migration of tropical plants from the southern to the northern hemisphere. However, this series of climatic oscillations occurred earlier than the estimated late Oligocene to Miocene divergence of the Sunda lineages (node 36 and 58 onwards) in the present study. Thus global climatic changes probably played a lesser role in facilitating northward migration and speciation than did the geological processes.

3.4.2 Divergence times and historical biogeography of Elaeocarpus

The topology of the maximum clade credibility chronogram of *Elaeocarpus* is largely consistent with the topology of the phylogenetic analysis based on the same concatenated data matrix of the five DNA regions (*psbA-trnH* spacer, *trnL-trnF* region, *trnV-ndhC* spacer, *Xdh1* and *Xdh2*) (Chapter 2, Fig. 2.6). Nodes that are well supported (bootstrap value in the Maximum Parsimony (MPBS), Maximum Likelihood (MLBS), PP ≥ 0.95) in the phylogenetic analyses, but less well supported (PP < 0.95) in the molecular clock analysis are: (1) the split between *Sericolea* and *Aceratium* + *Elaeocarpus* (Fig. 3.3: node 17; Chapter 2: Fig. 2.4, node 1, MPBS = 86 %, MLBS = 82 % and PP = 1, respectively); and (2) the split between *Aceratium* and *Elaeocarpus* (Fig. 3.3: node 19; Chapter 2: Fig. 2.4, node 2, MPBS = 96 %, MLBS = 99 %, PP = 1). In addition, the position of some terminal taxa (i.e. *E. acrantherus* Merr., *E. bancroftii* F.Muell., *E. barbulatus* R.Knuth, *E. brunnescens*, *E. coorangooloo*, *E. culminicola*, *E. dognyensis* Guillaumin, *E. dubuis* A.DC., *E. eumundi* F.M.Bailey, *E. geminiflorus* Brongn. & Gris, *E. hylobroma* Y.Baba & Crayn, *E.* cf. *inopinatus* Coode, *E. johnsonii* F.Muell. ex C.T.White, *E. kirtonii* F.Muell. ex F.M.Bailey, *E. kusanoi* Koidz., *E. nitidus*, *E. palembanicus* (Miq.) Corner, *E. polystachyus* Wall. ex Müll.Berol., *E. reticulatus* Sm., *E. sadikanensis* R.Knuth, *E. sedentarius*, *E. sylvestris* (Lour.) Poir. var. *ellipticus* (Thunb. ex Murray) Hara) in the molecular clock analysis in BEAST is different from the

phylogenetic analysis in MrBayes in Chapter 2. The phylogenetic positions of all these taxa should be regarded as unresolved. Notwithstanding, the taxon composition of the labelled main clades is largely the same, thus estimation of divergence times and the ancestral area reconstructions of these lineages were generally not affected.

3.4.2.1 Eocene origin and diversification in Australia

Divergence time analysis using BEAST suggests that *Elaeocarpus* diverged from its sister *Aceratium* at *c*. 45.46 Mya (Fig. 3.3: node 19; HPD: 53.63 – 37.22 Mya, PP < 0.95) in the Eocene. The ancestral area reconstructions using Fitch parsimony and DEC suggest that *Elaeocarpus* originated in Australia (Figure 3.4 & Table 3.3: node 21) and its crown group diversification began at *c*. 39.15 Mya (Fig. 3.3: node 21; HPD: 46.45 – 32.41 Mya, PP = 1) and gave rise to two lineages (*E. holopetalus*, and the remainder of the genus). *Elaeocarpus holopetalus* is confined to moist montane forest in southeastern Australia, and is often found in association with *Nothofagus* Blume (Coode 1984; Crayn pers. comm. 2014), a lineage which is generally considered a Gondwanan relict (Cook and Crisp 2005). Most of the remaining species of *Elaeocarpus* are found in tropical rainforests (Coode 1984). The split between *E. holopetalus* and the remainder of *Elaeocarpus* coincides with the time when Australia was at high latitudes and megathermal rainforests flourished, especially at the north and as coastal vegetation, extending from north Queensland to southeastern Australia (Christophel *et al*. 1987; Hill 2004). It seems likely that the *Elaeocarpus* (excluding *E. holopetalus*) radiated in these warm and wet environments. A similar scenario has been hypothesised for *Nothofagus*, i.e. a radiation coincident with a change from microthermal to megathermal climates (Cook and Crisp 2005).

Among the remainder of *Elaeocarpus*, biogeographical reconstructions using Fitch parsimony and DEC resolved an Australian origin for the following lineages: *E. coorangooloo* + *obovatus*, *E. sedentarius*, *E. hylobroma*, group VII and group XIB (Fig. 3.4: nodes 21, 23, 26, 60 and 113). The *E. coorangooloo* + *obovatus* clade diverged from its sister at *c*. 34.72 Mya (Fig. 3.3 & Table 3.3: node 22; HPD: 40.55 – 29.11 Mya, PP = 1) diversified from *c*. 18.27 Mya (node 23). This clade consists of four species, *E.*

coorangooloo, *E. arnhemicus*, *E. obovatus* and the putative new species, *Elaeocarpus* sp. "Mt Bellenden Ker". All except *E. arnhemicus* are endemic to Australia, and *E. arnhemicus* extends from northeastern Australia as far west as Java.

Elaeocarpus sedentarius represents an isolated lineage which diverged from the remainder of *Elaeocarpus* at *c*. 34.72 Mya (Fig. 3.3 & Table 3.3: node 26; HPD: 40.55 – 29.11 Mya, PP < 0.95). This is a narrowly distributed species occuring in the Koonyum and Nightcap Ranges in northeastern New South Wales, Australia (Maynard *et al*. 2008). Compared to other Australian species, *E. sedentarius* is morphologically distinctive and is similar to a New Guinean species, *E. blepharoceras* Schltr. These species share the following characteristics: a highly divided petal apex, a "pseudo-fleshy" outer endocarp consisting of densely arranged radial fibres, and longitudinally-furrowed outer bark (Maynard *et al*. 2008). No DNA material of *E. blepharoceras* was available for this study, but this putative relationship merits investigation with molecular evidence.

A topological discrepancy regarding the placement of *E. hylobroma* is observed between the BEAST (Fig. 3.4) and Mr Bayes analyses (Fig. 2.6). In the BEAST analysis, *E. hylobroma* (Fig. 3.3: node 28; PP = 1) is sister to the New Zealand + sect. *Elaeocarpus* + *ganitrus* clade (node 29; PP = 1). The BEAST analysis suggests that *E. hylobroma* is most closely related to the *ganitrus*, sect. *Elaeocarpus* and New Zealand clades with the MRCA of these dated at *c*. 29.27 Mya (Fig. 3.3 & Table 3.3: node 28; HPD: 33.70 – 25.38 Mya; PP = 1). However, in the MrBayes analyses, *E. hylobroma* is shown to be sister to the *ganitrus* clade only (Fig. 2.4: node 10; MPBS = 95 %, MLBS = 96 %, PP = 1). Baba and Crayn (2013) placed *E. hylobroma* in group V based on preliminary molecular evidence and the shared possession of four ovules per locule. *E. hylobroma* differs from members of the *ganitrus* clade which have 4 ovules per locule (discussed in Chapter 2, section 2.4.4). On the other hand, the phylogenetic relationships between *E. hylobroma*, the sect. *Elaeocarpus* clade and the New Zealand clade are also supported morphologically: *E. hylobroma* and the sect. *Elaeocarpus* clade both have a disk that is lobed and a 3-locular ovary; whereas *E. hylobroma* and the New Zealand clade both have 4 ovules per locule and a rugose fruit stone surface. *Elaeocarpus*

hylobroma shares some morphological similarities with each of the three clades, therefore rejection of either the results of molecular clock analysis in BEAST or the phylogenetic analysis in MrBayes would be premature. It may be best to re-investigate the phylogenetic position of *E. hylobroma* and its divergence time using more informative DNA markers and increased taxon sampling in putatively related clades. This uncertainty does not have a major effect on the biogeographical interpretation because the species in clades New Zealand, *ganitrus* (except *E. angustifolius* and *E. sphaericus* which are widespread in Australia, Malesia and Continental Asia) and the early diverging lineage in the sect. *Elaeocarpus* (i.e. *E. multisectus* Schltr.) are all distributed in geographical regions that were formerly part of Gondwana.

Species in the group VII and group XIB clades are predominantly Australian, with the exception of two species from the group VII clade: *E. culminicola* and *E. multiflorus* (Turcz.) Fern.-Vill. The first (*E. culminicola*) occurs in northeastern Australia, New Guinea and possibly extends to the Philippines; whereas *E. multiflorus* occurs in New Guinea and northward to Central Malesia (Philippines, Sulawesi, Flores, Ambon, Buru) and possibly Continental Asia (Taiwan) but is absent from Australia (Coode 1978, 1984, pers. comm. 2014). The group VII clade originated at *c*. 16.31 Mya (Fig. 3.3 & Table 3.3: node 60; HPD: 18.64 – 15.05 Mya, PP = 1), and the group XIB clade slightly later at *c*. 10.36 Mya (node 113; HPD: 15.58 – 5.49 Mya, PP < 0.95). *Elaeocarpus culminicola* is estimated to have an origin in Australia (Fig. 3.4 & Table 3.3: node 66) and diverged at *c*. 1.49 Mya (Fig. 3.3 & Table 3.3: node 66; HPD: 5.88 – 0 Mya, PP < 0.95) in the Pleistocene. On the other hand, the identity of the DNA sample of *E. multiflorus* needs further investigation as it was obtained from a sterile specimen (see detailed discussions in Chapter 2, section 2.4.4).

3.4.2.2 Migration out of Australia

In the Oligocene, while diversification of *Elaeocarpus* continued in Australia, the genus appears to have begun to migrate out of Australia to the surrounding regions. The results of Fitch parsimony and DEC analyses indicate that multiple such events occurred, beginning *c*. 29.27 Mya (Table 3.3: node 28; HPD: 33.70 – 25.38 Mya,

PP = 1), and coinciding with the approach and collision of the Sahul and Sunda shelves from the late Oligocene to the late Miocene (around 25 Mya – 10 Mya) (Ufford and Cloos 2005; Hall 2009). Further dispersals from the West Malesia region to Continental Asia were evident in the clades *ganitrus*, sect. *Elaeocarpus*, group VII and *coilopetalum*, followed by recent dispersal of the *E. subserratus* lineage (node 109; 6.45 Mya, HPD: 10.12 – 1.33 Mya, PP < 0.95) to Madagascar. Reverse dispersals from the West Malesia to Central and East Malesia were also observed in the clades sect. *Elaeocarpus*, *monocera* and *coilopetalum*.

Within the Sahul Shelf, radiation of *Elaeocarpus* from *c*. 29.27 Mya (Fig. 3.3 & Table 3.3: node 28; HPD: 33.70 – 25.38 Mya, PP = 1) gave rise to three main lineages: *ganitrus*, sect. *Elaeocarpus* and group VI. The *ganitrus* clade further dispersed to Caroline Islands (i.e. E. *carolinensis* Koidz.; Fig. 3.3 & Table 3.3: node 34; *c*. 7.80 Mya; HPD: 13.94 – 2.93 Mya, PP = 1). The *ganitrus* clade is composed of *E. carolinensis*, *E. ptilanthus* Schltr., *E. polydactylus* Schltr., *E. angustifolius*, *E. grandis* F.Muell. and *E. sphaericus* K.Schum., and the MRCA of this clade is estimated at *c*. 23.47 Mya (Fig. 3.3 & Table 3.3: node 31; HPD: 24.71 – 23.00 Mya, PP = 1). *Elaeocarpus carolinensis* is found in Caroline Islands, the next two are probably endemic to New Guinea, whereas the last two were treated as synonyms under *E. angustifolius* (Coode 1984), which is widespread in Continental Asia, Malesia and Australia. The phylogenetic relationships of the taxa in this clade, particularly the limits of the *E. angustifolius* complex were discussed in Chapter 2, section 2.4.4. The BEAST results suggest that the *ganitrus* clade arose in the Oligocene, between *c*. 27.45 Mya (Fig. 3.3 & Table 3.3: node 30; HPD: 31.93 – 23.78 Mya, PP < 0.95) and *c*. 23.47 Mya (node 31; HPD: 24.71 – 23.00 Mya, PP = 1), and Fitch parsimony and DEC analyses resolve New Guinea as the most likely ancestral area (Fig. 3.4: node 31). Fossil fruit stones – *E. spackmaniorum* – that closely resemble those of the *E. angustifolius* complex are known from early Oligocene to Miocene deposits in Queensland, Australia (Dettman and Clifford 2000). If it is accepted that these fossils belong to the *ganitrus* clade, then the age is consistent with the BEAST-derived age of the group. It is noteworthy that the inferred age of the *E. angustifolius* complex in Australia is relatively young, between *c*. 7.80 Mya (Fig. 3.3 & Table 3.3: node 34; HPD: 13.94 – 2.93 Mya, PP = 1) and *c*. 4.14 Mya (node 35; HPD:

8.61 – 0.88 Mya, PP < 0.95). This suggests extinction of the *ganitrus* clade in the southern part of Sahul (the island of Australia), perhaps due to the contraction of the rainforests and climatic oscillations in the late Cenozoic (Byrne et al. 2011), followed by recolonisation from the north, probably New Guinea, relatively recent. Phylogeographical studies of Australian members of the *E. angustifolius* complex showed low genetic diversity and low phylogeographical structure suggest a recent range expansion, which is consistent with a recent arrival scenario (Rossetto *et al*. 2004, 2007). The alternative hypothesis – persistence of populations on Australia through the contraction of the rainforests – is not supported by either the molecular dating, nor the phylogeographic (Rossetto *et al*. 2004, 2007) evidence.

The crown age of the sect. *Elaeocarpus* clade is estimated at *c*. 19.77 Mya (Fig. 3.3 & Table 3.3: node 36; HPD: 25.40 – 13.95 Mya, PP = 1) with the MRCA in New Guinea. Subsequent migration into West Malesia began between *c*. 19.77 Mya (node 36; HPD: 25.40 – 13.95 Mya, PP = 1) and *c*. 12.28 Mya (node 37; HPD: 17.17 – 7.95 Mya, PP = 1), probably enabled by the Miocene collision between the Sunda and Sahul shelves (Hall 2002) and dispersal of fruits by frugivorous birds (Crome 1975, 1976). Further migration into Continental Asia is evident in five species, viz.: *E. robustus* (node 41; 7.34 Mya, HPD: 12.03 – 3.11 Mya, PP < 0.95), *E. dongnaiensis* Pierre and *E. stipularis* (split at node 42; 5.41 Mya, HPD: 9.71 – 1.97 Mya, PP < 0.95), *E. sylvestris* var. *ellipticus* (node 44; 5.42 Mya, HPD: 10.36 – 1.22 Mya, PP < 0.95) and *E. nitidus* (node 47; 6.59 Mya, HPD: 10.61 – 2.68 Mya, PP < 0.95), although these nodes are not statistically supported. On the other hand, possible reverse dispersals from West Malesia to Central Malesia are shown in six species, *E. glaber* (node 39; 5.62 Mya, HPD: 11.18 – 0.90 Mya, PP < 0.95), *E. stipularis* (node 42; 5.41 Mya, HPD: 9.71 – 1.97 Mya, PP < 0.95), *E. floribundus* Blume (node 45; 3.48 Mya, HPD: 8.54 – 0.19 Mya, PP < 0.95), *E. nitidus* (node 47; 6.59 Mya, HPD: 10.61 – 2.68 Mya, PP < 0.95), *E. brunnescens* (node 49; 5.56 Mya, HPD: 10.50 – 1.46 Mya, PP < 0.95) and *E. submonoceras* subsp. *lasionyx* (node 52; 0.80, HPD: 1.83 – 0.01 Mya, PP = 1). It is noteworthy that *E. multisectus* from eastern Malesia resembles the western Malesian species of the sect. *Elaeocarpus* clade, whose flowers open widely at anthesis. Relationships and divergence dates between the eastern and western Malesian species of the sect.

Elaeocarpus clade will become clearer when a larger number of the eastern species are included.

The group VI clade consists of three essentially New Guinean species in this study, *E. fairchildii* Merr., *E. kerstingianus* Schltr. and *E. nouhuysii* Koord. The MRCA is estimated at *c*. 15.89 Mya (Fig. 3.3 & Table 3.3: node 86; HPD: 21.26 – 10.74 Mya, PP = 1). The historical biogeography reconstructions using Fitch parsimony suggested West Malesia, New Guinea and the Pacific islands as ancestral areas (node 85, 19.23 Mya, HPD: 24.56 – 13.88 Mya, PP = 0.98; and node 68, 22.21 Mya, HPD: 27.05 – 17.29 Mya, PP = 1) with eastwards dispersal from West Malesia or westwards from the Pacific islands back into New Guinea across oceanic and mountain barriers. Interestingly, the backbone nodes defining these relationships (node 67 and 58, both PP < 0.95) are not statistically supported. Node 54 (PP = 1) is, however, well supported and Australia is inferred as the ancestral area by both of the Fitch parsimony and DEC methods. Dispersal from Australia to New Guinea seems more likely because Australia and New Guinea were connected intermittently in the Neogene, finally separating about 8,000 – 6,500 years ago when sea levels rose globally and flooded the lowlands, creating the Papuan Peninsula (Ufford and Cloos 2005).

3.4.2.3 New Guinea

New Guinea exhibits the highest species diversity of *Elaeocarpus* with *c*. 113 taxa known, about 90 % of which are endemic (Coode pers. comm. 2014). The formation of the Central Range began about 8 Mya and resulted a *c*. 1300 km-long mountainous backbone extending east from the Bird's Head to the Papuan Peninsula (Ufford and Cloos 2005). The uplift of mountains with some peaks over 5000 m altitude (Hill and Hall 2003) greatly influences the temperatures and rainfalls at different altitudinal range (Beebe and Cooper 2002), leading to the creation of dynamic habitats and new niches available for later colonisation, followed by radiation and speciation. Apart from the Central Range orogeny, climatic oscillations between dry-cooling and moist-warming phases from middle Miocene until late Pliocene (Zachos *et al*. 2001; Morley 2002), followed by several cycles of glacial-interglacial

episodes during Pleistocene (from *c*. 1.8 Mya) (Poulin *et al*. 2002; Johnson 2004), could have also encouraged speciation of *Elaeocarpus* in New Guinea. Notwithstanding, more than one mechanism of species radiation and diversification, such as vicariant allopatric divergence associated with the geologic uplifts, hybridisation and introgression, morphological convergence and innovations, biotic interactions and polyploidy (summarised in Wen *et al*. 2014), may have played a role. Species radiation due to the effects of orogeny was studied in *Begonia* (Begoniaceae), where mountain formations and deformations in Sulawesi and New Guinean were hypothesised to have created topographical heterogeneity and suitable microhabitats and acted as the main drivers in the radiation of *Begonia* in these regions (Thomas *et al*. 2012). However evidence for the roles of these mechanisms on the radiation of *Elaeocarpus* is lacking to date. Moreover, very few New Guinean representatives were included in this study; therefore, conclusions on the relationships and diversification of this genus in New Guinea are premature at this stage.

3.4.2.4 New Zealand

Both Fitch parsimony and DEC analyses indicate that New Zealand is the most likely ancestral area for the New Zealand clade (Fig. 3.3: node 53). Only two *Elaeocarpus* species are known from New Zealand, *E. dentatus* and *E. hookerianus* and both were sampled here. The age of this node is estimated at *c*. 13.13 Mya (HPD: 21.90 – 5.25 Mya, PP = 1). New Zealand was separated from the Australia-Antarctica landmass around 130 – 85 Mya (Hall 2009). Later, the "Oligocene drowning" event that took placed at *c*. 36 – 24 Mya (Cooper and Cooper 1995), where either a substantial or the whole landmass of the present day New Zealand was submerged and led to two main categories of the present day biota of New Zealand: the palaeoendemic lineages of putatively Gondwanan character (Nelson 1975; Cooper *et al*. 2001; Ericson *et al*. 2002) and the neoendemics that colonised islands by long-distance dispersal (Pole 1994; McGlone 2005; Waters and Craw 2006; McDowall 2008; Biffin *et al*. 2010). Considering that both of these geological events happened long before the New Zealand lineage and its sister diverged, *Elaeocarpus* in New Zealand is likely neoendemic, the result of long distance dispersal, rather than vicariance.

Neoendemicity or post-Gondwanan arrival into New Zealand is seen in other organisms, such as plants – *Myosotis* L. (Boraginaceae, Winkworth *et al*. 2005), *Nothofagus* (Nothofagaceae, Cook and Crisp 2005), tribe Styphelieae (Ericaceae, Puente-Lelievre *et al*. 2013); birds – Callaeatidae (Shepherd and Lambert 2007), *Hemiphaga* Bonaparte (Columbidae, Pereira *et al*. 2007); reptiles – *Oligosoma* Girard (Scincidae, Chapple *et al*. 2009); fish – *Neochanna* Günther (Galaxiidae, Waters and McDowall 2005); insects – *Kikihia* Dugdale (Cicadidae, Arensburger *et al*. 2004).

3.4.2.5 West Malesia

West Malesia harbours almost a quarter of all species of *Elaeocarpus*, most of them occurring in Borneo. The earliest suggested date for the colonisation of West Malesia is estimated at *c*. 25.29 Mya (Fig. 3.3 & Table 3.3: node 58; HPD: 29.89 – 20.99 Mya, PP < 0.95) in the Oligocene and gave rise to the clades *monocera*, *polystachyus* and *acronodia*. The MRCA of the *monocera* clade is estimated at *c*. 12.57 Mya (Fig. 3.3 & Table 3.3: node 70; HPD: 18.55 – 6.76 Mya, PP = 1). Clades *polystachyus* and *acronodia* are with MRCA estimates of *c*. 7.02 Mya and *c*. 10.52 Mya, respectively (node 100; HPD: 11.03 – 3.35 Mya, PP < 0.95; and node 118; HPD: 16.53 – 5.61 Mya, PP = 1). Species in the last two clades are all endemic to West Malesia (Coode 1996b, c).

In the *monocera* clade, two reverse dispersals into Central Malesia were observed in the species *E. grandiflorus* and *E. obtusus* subsp. *obtusus*, estimated at *c*. 12.57 Mya (node 70; HPD: 18.55 – 6.76 Mya, PP = 1) and *c*. 4.07 Mya (node 71; HPD: 8.39 – 0.80 Mya, PP = 1), respectively. Migration of biota across Central Malesia in the late Miocene or early Pliocene between 10 – 5 Mya, may have been facilitated by the formation of Sulawesi (Marshall 1983; Hall 2001b) or exposure of some lowlands within Malesia during episodes of sea-level changes, which acted a dispersal corridors (de Bruyn *et al*. 2014) or provided a substantial landmass for stepping-stone-island dispersal. The zoochory mode of dispersal of some *Elaeocarpus* could have also favoured their dispersal across water barriers (Crayn *et al*. 2015). Transoceanic dispersal of across deep marine areas separating the Sunda and Sahul shelves has

been inferred for various biotic groups (Lohman *et al*. 2011), including zoochorous plant taxa (Crayn *et al*. 2015).

In the middle and late Miocene, lowered sea levels exposed land bridges, which connected the landmasses of Sumatra and Java to mainland Southeast Asia (Hall 1998). In the Pleistocene, sea levels decreased to a maximum of 120 m below the present level, exposing even more land bridges connecting Peninsular Malaysia, Sumatra, Java and Borneo, and leading to the expansion of rainforests (Inger and Voris 2001; Bird *et al*. 2005). As a result, northwards dispersal of *E. nitidus*, *E. petiolatus*, *E. robustus*, *E. stipularis* and possibly *E. angustifolius* complex into Continental Asia may have been facilitated.

Early divergence of *Elaeocarpus* in West Malesia could have occurred between *c*. 12.28 Mya (Fig. 3.3 & Table 3.3: node 37; HPD: 17.17 – 7.95 Mya, PP = 1) and *c*. 7.02 Mya (node 100; HPD: 11.03 – 3.35 Mya, PP = 1) during the Miocene. Radiation of the genus in the region was likely influenced by climatic changes. During the mid Miocene, a warming phase (between 17 – 15 Mya) (Zachos *et al*. 2001) promoted the expansion of megathermal vegetation throughout most of Sundaland (Morley 2007), including West Malesia. This was followed by a global cooling phase (between 14.2 – 13.8 Mya) that reduced sea surface temperatures and led to the expansion of the Antarctic ice-sheets, lasting until the early Pliocene (around 6 Mya) (Zachos *et al*. 2001; Shevenell *et al*. 2004). The level of atmospheric carbon dioxide declined at the same time reducing productivity of terrestrial vegetation (Kürschner *et al*. 2008) and extent of megathermal vegetation in the tropics (Morley 2007). A subsequent gradual transition between global cooling and warming phases took place and lasted until the late Pliocene (*c*. 3.2 Mya) (Zachos *et al*. 2001). The change to a warmer and wetter climate provided a platform for stable increase of the rainforest extent on Sundaland until the mid Pliocene (Heaney 1991; Morley 2007). The relatively recent radiation (between 18.55 Mya – 0 Mya) of *Elaeocarpus* in West Malesia might be linked to these series of global climatic oscillations. Apart from the climatic oscillation, mountain building in Borneo also began to take place at the beginning of the Miocene due to the collisions between Sunda and Sahul shelves (Hall 2002), and between North Borneo and the

continental margin of South China (Hall 2009). Thus the uplift of mountains likely opened up new niches for later colonisation and physical barriers that promoted speciation of *Elaeocarpus*, and made Borneo the second largest centre of diversity after New Guinea.

3.4.2.6 New Caledonia

New Caledonia rifted from Australia during the Late Cretaceous (65 – 80 Mya). The island was submerged for long periods in the Palaeocene and Eocene (Grandcolas *et al*. 2008), then re-emerged and finally reached its present position at *c*. 50 Mya (McLoughlin 2001; Veevers 2001; Neall and Trewick 2008). Recent studies on the terrestrial biota suggest New Caledonia is an old Darwinian island, where colonisation on the island began around 37 Mya, rather than a Gondwanan refugium of relict lineages (Pillon 2012; Swenson *et al*. 2013). The present study indicates that *Elaeocarpus* colonised the island at *c*. 16.92 Mya (Fig. 3.3 & Table 3.3: node 72; HPD: 22.37 – 11.69 Mya, PP = 0.98), apparently once, and subsequently radiated. A similar scenario of single dispersal to post-immersed islands is seen in other groups, such as *Araucaria* Juss. (Araucariaceae), Arecoideae (Arecaceae), *Dacrydium* Lamb. (Podocarpaceae), *Dracophyllum* Labill. (Ericaceae), *Kermadecia* Brongn. & Griseb. and *Sleumerodendron* Virot (Proteaceae), *Metrosideros* Banks ex Gaertn. (Myrtaceae), *Nothofagus* (Nothofagaceae), *Pycnandra* Benth. (Sapotaceae) (summarised in Pillon 2012), and *Styphelia* Sm. (Puente-Lelievre *et al*. in review). Biota on the island possibly derived from lineages in Australia, New Guinea and Malesia as a result of long distance dispersal (Grandcolas *et al*. 2008; Murienne 2009; Bartish *et al*. 2011; Espeland and Murienne 2011; Pillon 2012; Morat *et al*. 2012), possibly via stepping-stone islands between Australia and New Caledonia as suggested by Ladiges and Cantrill (2007).

All New Caledonian species formed a clade, except *E. dognyensis* (Fig. 3.3: node 110). Tirel (1983) in her revision of the New Caledonian *Elaeocarpus* did not highlight any unusual morphological features of *E. dognyensis* in comparison to the rest of the New Caledonian species. This unexpected placement may be explained in two ways: (1), that *Elaeocarpus* dispersed to New Caledonia more than once, as has been

inferred for *Diospyros* L. (Ebenaceae) (Turner *et al*. 2013) and Sapotaceae (Swenson 2014); or (2), experimental error (e.g. sample mix-up or identification error). The sample of *E. dognyensis* used in this study was received from another institution and its identity could not be confirmed during the present study. Further phylogenetic investigations with new samples of *E. dognyensis* collected from the field and addition of the remaining New Caledonian *Elaeocarpus* that were not included in this study will improve insights into the evolution of New Caledonian *Elaeocarpus*.

3.4.2.7 Hawai'i

In this study, *E. bifidus* Hook. & Arn. (Fig. 3.3: node 85), the only *Elaeocarpus* known from Hawai'i so far (Rock 1913; Coode 2004), is estimated to have diverged from the New Guinean group VI clade at *c*. 19.23 Mya (Table 3.3: HPD: 24.56 – 13.88 Mya, PP = 0.98). The Fitch parsimony analysis suggested New Guinea and Pacific islands as the ancestral areas for the *E. bifidus* lineage, and the DEC method suggested the same two areas plus West Malesia. Hawai'i is part of the Hawaiian archipelago situated in the northern Pacific Ocean. America, the closest continent, is more than 2600 km away and the archipelago is considered as one of the most isolated in the world (Harbaugh *et al*. 2009). Numerous molecular phylogenetic analyses have revealed complex dispersal patterns of the Hawaiian flora, with different groups having originated in the western Pacific (*Psychotria* L., Rubiaceae, Nepokroeff *et al*. 2003; *Pittosporum* Banks ex Gaertn., Pittosporaceae, Gemmill *et al*. 2002), Australia (*Scaevola* L., Goodeniaceae, Howarth *et al*. 2003), Southeast Asia (*Lysimachia* L., Myrsinaceae, Hao *et al*. 2004), Africa (*Hesperomannia* A.Gray, Asteraceae, Kim *et al*. 1998; Gossypieae, Malvaceae, Seelanan *et al*. 1997), the Americas (*Gunnera* L., Gunneraceae, Wanntorp and Wanntorp 2003; *Stachys* L., Lamiaceae, Lindqvist and Albert 2002) and the subartic (*Schiedea* Cham. & Schltdl., Caryophyllaceae, Wagner *et al*. 2005; *Viola* L., Violaceae, Ballard and Sytsma 2000). The origin of Pacific *Elaeocarpus* is postulated using molecular evidence for the first time in this study. Thus, to better understand the origin of the Hawaiian *Elaeocarpus* – whether dispersed from the Pacific Islands, New Guinea or even further from the West Malesia as postulated by Fitch parsimony and DEC here, and its dispersal patterns – whether

via zoochory or hydrochory based on the recorded or possible dispersal mechanisms, a more extensive and well represented taxonomic sampling from the Pacific Islands are needed before a conclusion can be made.

3.4.3 Macrofossils of Elaeocarpus

Macrofossil evidence clearly demonstrates that *Elaeocarpus* was present in Australia in the Oligocene. Among the four macrofossils of *Elaeocarpus* used in this study, *E. mackayi*, in particular, merits further investigation. *Elaeocarpus mackayi* exhibits some morphological features found in *E. hylobroma* and members of the *ganitrus* and sect. *Elaeocarpus* clades. This fossil was discovered in Victoria, southern Australia and is estimated to date to between the early Oligocene to Miocene. The sculptured, near-spherical fruit stone is similar to the stone of *E. spackmaniorum* and the *E. angustifolius* complex (Dettman and Clifford 2000); but the stone is usually 3-locular and the sterile locule(s) are compressed; together these are some of the diagnostic characters of the sect. *Elaeocarpus*, although the compressed locule(s) may not resemble a D-shape in transverse section as in the sect. *Elaeocarpus* clade. The sect. *Elaeocarpus* clade has its centre of diversity in Malesia but no extant representative in Australia. Therefore, *E. mackayi* could belong to the ancestor lineage of the clades *ganitrus* and sect. *Elaeocarpus*. The phylogenetic relationships between *E. hylobroma* and the clades *ganitrus*, sect. *Elaeocarpus* and New Zealand are not fully resolved (discussed in section 3.4.2.1) and this species could either be sister to the *ganitrus* clade (based on analysis in MrBayes) or to the clades *ganitrus*, sect. *Elaeocarpus* and New Zealand (based on analysis in BEAST). Phylogenetic analysis of fruit characters may provide better evidence, however, such an analysis has not yet been attempted and was beyond the scope of the present study.

3.5 Conclusions

The divergence times of the Elaeocarpaceae estimated in this study deviate slightly from those previously determined using different calibration points and dating methods. But because the present study included much more intensive and broader

taxon sampling, was based on a larger multi-locus DNA sequence dataset, and utilises an uncorrelated lognormal relaxed clock model and five macrofossils as internal calibration points, the results presented here are likely to be a more accurate estimate of divergence times within Elaeocarpaceae.

The evolutionary history of *Elaeocarpus* is inferred comprehensively for the first time, in time and space. The genus is shown to have originated in Australia in the Eocene and migrated out of Australia to the surrounding regions mostly in the Oligocene and the Miocene. The suggested migration events are broadly consistent with the available geological and climatic data, where migration and reverse migration of the genus via long-distance dispersal were possible due to the effects of (1) Sahul-Sunda collision which reconfigured and created much land within the Malesian region (Hall 1995); (2) climatic changes from cool and dry climates in the Oligocene and early Miocene to warm and wet climates in the mid Miocene leading to the expansion of rainforests on Sundaland (Morley 1998); (3) orogeny in West Malesia, Wallacea and New Guinea (Hall and Nichols 2002; Ufford and Cloos 2005; Hall 2009) which likely provided new niches and promoted speciation; and (4) zoochorous dispersal in the genus (Crayn *et al*. 2015).

All macrofossils used here are assigned to the crown nodes of the clade of extant taxa based on combinations of diagnostic morphological similarities that are unique to each clade. However, assignment of the macrofossils of *Elaeocarpus* were unable to be placed at shallow terminal branches because detailed morphological characters between the fossil and each of the extant taxa were not studied and that is beyond the scope of the present study. Future studies, such as anatomy of the fruit stone and mapping traits of the morphological characters onto phylogenies, may provide better evidence on the placement of these fossils on the phylogenetic tree.

Elaeocarpus is widely distributed, particularly in tropical rainforests and species from some biogeographical regions, such as Madagascar, continental Asia, Wallacea (Sulawesi, Philippines, Lesser Sunda islands, Moluccas), New Guinea and the Pacific islands, are poorly represented in this study. Therefore, understanding of the origins of

and evolution of these floras will become even clearer when these regions are sampled in more depth.

This chapter received valuable contributions from other researchers and they were: Mark Coode (K) who shared his knowledge and opinions generously and the following research institutes or herbaria for field assistance and lending specimens under their care: Herbarium of the Brunei Forestry Center (BRUN), Herbarium Bogoriense (BO), Herbarium of the Forest Research Institute Malaysia (KEP), Forest Research Center Sandakan (SAN), Department of Forestry Sarawak (SAR), Singapore Botanic Gardens (SING) and Smithsonian Tropical Research Institute: Center for Tropical Forest Science in Bukit Timah.

Chapter 4: Morphometric analysis of the West Malesian *polystachyus* group: morphological variation and ecological influences.

ABSTRACT

The *polystachyus* group of *Elaeocarpus* (Elaeocarpaceae) is an informal infrageneric group comprising nine taxa (six species and four varieties) endemic to West Malesia. Whereas the group is distinct within *Elaeocarpus* and clearly diagnosed by a unique combination of morphological character states, taxon boundaries within it are unclear. Morphometric analysis was undertaken to test taxon boundaries within the *polystachyus* group and to explore correlations between morphological similarities within the group and ecology (using phytogeographical affinities at floristic province and subprovince levels as proxies). A total of 39 quantitative and one qualitative variables were scored from 213 herbarium specimens and analysed using classification and ordination techniques. The results of the morphometric analysis are broadly congruent with the existing taxonomy erected using an intuitive approach in that the recognition of six species is supported, but the infraspecific taxa are not supported. The morphological analysis indicates that with the exception of *E. cupreus*, taxa within the group are generally segregated at the floristic provinces (a proxy for general geographical isolation), but not at the floristic subprovinces (a detailed ecological phytogeography based on the combination of floristic composition, floristic influences, elevation and soil types). Each taxon, at least at the species level, appears to be maintaining its own morphological characteristics despite high degree of sympatry, which suggests reproductive isolation at the species level. Additionally, morphological variation is not correlated with ecological phytogeography. Therefore, taxa within the *polystachyus* group, at the species level, are meaningful.

4.1 Introduction

Elaeocarpus L. is the largest genus in the family Elaeocarpaceae. The genus comprises 350 – 400 species widely distributed in the palaeo-tropics, -subtropics and warm temperate regions, from Madagascar eastwards to Asia, southwards to Australia and the Pacific islands (Coode 2004; Coode pers. comm. 2012; Crayn *et al.* 2006). The genus is particularly diverse in West Malesia, a floristic region described by Van Steenis (1950), with *c*. 100 species of *Elaeocarpus* recorded there: *c*. 70 species on Borneo, *c*. 35 on the Malay Peninsula, *c*. 31 on Sumatra, *c*. 14 on Java and *c*. 12 on Palawan (Coode 1984, 1996a – c, 1998, 2001b, d, 2010; Coode pers. comm. 2012).

Six informal infrageneric groups are recognised for the West Malesian species. These are the *acronodia* (Coode 1996b), *coilopetalum* (Coode 1998, 2001b, c), section *Elaeocarpus* (Coode and Weibel 1994; Coode 1996a, 2001a), *ganitrus* (Coode 1998, 2010), *monocera* (Coode 1998) and *polystachyus* (Coode 1996c) groups. Of these, the *polystachyus* group is only group that is endemic to West Malesia (Coode 1996c). This group comprises six species and four varieties: *E. clementis* var. *borneensis* (Ridl.) Coode, *E. clementis* var. *clemensiae* (R. Knuth) Coode, *E. clementis* Merr. var. *clementis*, *E. clementis* var. *kostermansii* Coode, *E*. *cupreus* Merr., *E*. *integripetalus* Miq., *E*. *multinervosus* R. Knuth, *E*. *polyanthus* Ridl. and *E*. *polystachyus* Wall. ex Müll. Berol. (Coode 1996c). It is well-defined and is readily distinguished from other groups within *Elaeocarpus* (Coode 1996c) by a set of distinct morphologies: petals fleshy, disk annular-shaped and weakly developed, stamens numerous (35 – 80) and densely arranged in multiple tiers, anthers without awns (extensions of the anthers' outer lips), ovary of 2- or 3-locules with 4 – 12 ovules in each loculus, embryos curved and seeds with ruminate endosperm (Coode 1996c). In contrast, taxa within the *polystachyus* group are not well-defined; the diagnostic characters used to define taxon boundaries show apparently continuous variation (listed in Table 4.1).

Table 4.1 Comparison of diagnostic morphological characters within the *polystachyus* group complex showing continuous and overlapping character states and unclear taxon boundaries. *Elaeocarpus cupreus* (EC), *E. clementis* var. *borneensis* (ECb), *E. clementis* var. *clemensiae* (ECs), *E. clementis* var. *clementis* (ECc), *E. clementis* var. *kostermansii* (ECk), *E. integripetalus* (EI), *E. multinervosus* (EM), *E. polyanthus* (EP) and *E. polystachyus* (ES) (Regular text = information from Coode 1996c and unpublished data; italic text = observations made in this study).

Characters	EC	ECb	ECs	ECl	ECk	EI	EM	EP	ES
Young twig indument	Glabrous	Densely hairy	Hairy to glabrous	Hairy to glabrous	Densely hairy	Glabrescent	*Hairy*	*Hairy*	*Glabrous or hairy at tip*
Petiole thickness	≤ 1.8 mm thick	*0.8–2.5 mm thick*	*0.6–1.9 mm thick*	*0.8–2.2 mm thick*	1.3–2.3 mm thick	*c. 1.7 mm thick*	≥ 2 mm thick	*1.1–2.5 mm thick*	*1–1.6 mm thick*
Pair of secondary veins	12 pairs	(5–)10–14(–16) pairs	(5–)10–14(–16) pairs	(5–)10–14(–16) pairs	(5–)10–14(–16) pairs	8–11 pairs	12–17 pairs	*10–14 pairs*	*6–13 pairs*
Number of flowers per inflorescences	10–25-flowered	(15–)20–50-flowered	(15–)20–50-flowered	(15–)20–50-flowered	(15–)20–50-flowered	15–20-flowered	30–60-flowered	12–30-flowered	25–38-flowered
Inflorescence indument	Hairy	Densely hairy	Hairy to glabrous	Hairy to glabrous	Densely hairy	Sparsely hairy	*Hairy*	*Hairy*	*Hairy*
Flower sexuality	Unisexual	Bisexual	Bisexual	Bisexual	Bisexual	Bisexual	Unisexual	Bisexual	*Bisexual*
Flower-mery	5	5	5	5	5	4	5	5	4

Flower bud shape	Ovoid-globose	*Globose to ovoid*	Round to ovoid-globose, rounded at apex	Conical or ovoid to broad-ovoid, obtuse to acute at apex	Rounded to obtuse or acute at apex	Broad ovoid, subacute at apex	*Globose*	*Globose to ovoid*	*Globose to ovoid*
Flower bract persistence	Caducous	*Caducous*	Persistent	Persistent	Caducous	Unknown	*Caducous*	*Caducous*	*Caducous*
Petal length	4–5(–5.5) mm long	≥ 5 mm long	4.5–6(–6.5) mm long	5–6(–8) mm long	≥ 5 mm long	5–6 mm long	*4.3–7.5 mm long*	*5.5–6.8 mm long*	*4–7 mm long*
Petal apex shape	Pointed, acute or subacute	Pointed, subacute or 2 or 3-denticulate	Pointed, subacute or 2 or 3-denticulate	Pointed, subacute or 2 or 3-denticulate	Pointed, subacute or 2 or 3-denticulate	Rounded or rarely 1–5-denticulate	Pointed, subacute, subobtuse or occasionally acuminate	Pointed or 1 or 2-denticulate	Rounded, subentire or 1–6-denticulate, rarely attenuate and subacute
Pedicel thickness	< 1 mm thick	< 1 mm thick	< 1 mm thick	< 1 mm thick	< 1 mm thick	> 1.6 mm thick	≤ 1 mm thick	> 1 mm thick	1 mm thick
Number of ovules per locule	5–6	(4–)6(–10)	6(–8)	6(–8)	8	8	8	8–9	10–12
Fruit stalk length	*< 10 mm long*	< 10 mm long	< 10 mm long	< 10 mm long	< 10 mm long	unknown	*< 10 mm long*	> 12 mm long	*3–13 mm long*

Fruit indument	Glabrescent	Glabrous or virtually so	Glabrous or virtually so	Glabrous or virtually so	Glabrous or virtually so	unknown	*Glabrous*	Velvety	*Glabrescent, velvety when young*
Inner mesocarp thickness	≤ 2 mm thick	≤ 2 mm thick	≤ 2 mm thick	≤ 2 mm thick	≤ 2 mm thick	unknown	≤ 2 mm thick	≥ 2 mm thick	2–4 mm thick
Inner mesocarp length	< 2 cm long	< 2 cm long	< 2 cm long	< 2 cm long	< 2 cm long	unknown	< 2 cm long	> 2 cm long	1.1–1.8 cm long
Inner mesocarp thickness	< 2 mm thick	< 2 mm thick	< 2 mm thick	< 2 mm thick	< 2 mm thick	unknown	< 2 mm thick	≥ 2 mm thick	< 2 mm thick

Taxa within the *polystachyus* group are mostly restricted to single landmasses within West Malesia, except *E*. *cupreus* which occurs in Sumatra, Malay Peninsula and Borneo. *Elaeocarpus integripetalus* is endemic to Sumatra, *E*. *polystachyus* to Malay Peninsula, and *Elaeocarpus clementis* (including all of the varieties), *E*. *multinervosus* and *E*. *polyanthus* to Borneo (Fig. 4.1). Some species are poorly understood due to a paucity of scientific records, such as *E. integripetalus* is known only from the type specimen, which was collected from Sumatra (Payakumbuh) in the 1860s, *E*. *clementis* var. *kostermansii* is known only from two localities in Kalimantan, and nearly all known specimens of *E*. *polyanthus* were collected from a single population in the Semengoh Arboretum in Sarawak, Borneo.

Most taxa within the *polystachyus* group are found in lowland dipterocarp forest and in various habitat types, for instance *kerangas* (tropical heath), swamps and seasonally swampy vegetation. *Elaeocarpus clementis* var. *borneensis* and *E. clementis* var. *clementis* are found in montane forest and *E*. *polystachyus* also occurs in coastal vegetation. Phenotypic plasticity (polymorphism), such as variation in morphology, physiology, behaviour or phenology, can result from the influence of the environment (Price *et al*. 2003; Reed *et al*. 2010). Whether morphological polymorphism within the *polystachyus* group is correlated with environmental factors has not yet been investigated.

4.1.1 Aims

Delimitation of species complexes of various plant groups, including *Elaeocarpus sedentarius* D. J. Maynard & Crayn (Maynard *et al*. 2008) and *E. obovatus* G. Don (Baba 2013), has been successfully achieved using morphometric analysis in cases where research failed to discover unique, constant morphological character states (see section 1.1.3). Therefore this study aimed to: 1. test taxon boundaries within the *polystachyus* group using morphometrics based on the phylogenetic framework established in Chapter 2 to provide statistically reliable and repeatable data, and to compare the results of the morphometric analysis with the existing

taxonomy; and 2. assess the relationship between morphology and environment and hence the taxonomic value of morphological variation.

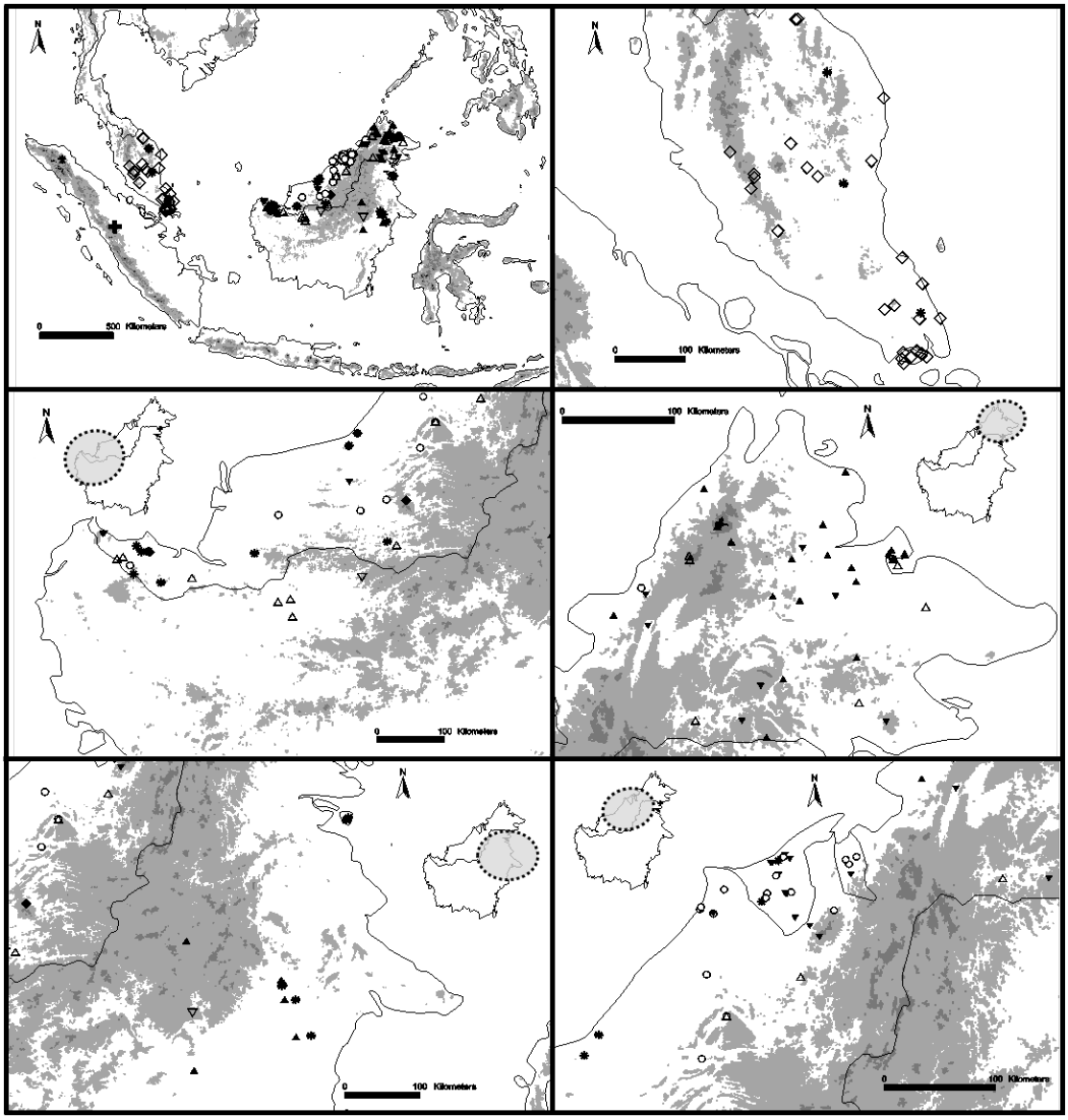


Fig. 4.1 Distribution of the *polystachyus* group: *E. clementis* (var. *borneensis* ▲, var. *clemensiae* ▼, var. *clementis* △ and var. *kostermansii* ▽), *E. cupreus* ✱, *E. integripetalus* ✚, *E. multinervosus* ○, *E. polyanthus* ◆ and *E. polystachyus* ◇. A, Distribution of the nine taxa in West Malesia; B, Distribution in Malay Peninsula; C, Distribution of in West Borneo; D, Distribution of in East Borneo; E, Distribution in Southeast Borneo; F, Distribution in North-Central Borneo.

4.2 Materials and methods

4.2.1 Plant materials

Two hundred and thirteen herbarium specimens (Operational Taxonomic Units, OTUs) were examined in this study. These include 39 *E. clementis* var. *borneensis*, 43 *E. clementis* var. *clemensiae*, 16 *E. clementis* var. *clementis*, 2 *E. clementis* var. *kostermansii*, 29 *E. cupreus*, 1 *E. integripetalus*, 27 *E. multinervosus*, 9 *E. polyanthus* and 47 *E. polystachyus*. These specimens are held at the following herbaria: Herbarium Bogoriense (BO), National Herbarium of Brunei Forestry Centre (BRUN), Herbarium of the Forest Research Institute Malaysia (KEP), Nationaal Herbarium Nederland, Leiden University branch (L), Herbarium of the Forest Research Centre, Sandakan (SAN) and Herbarium of the Singapore Botanic Gardens (SING). Type specimens held at Kew (K) were examined from the high resolution digital images. Collection data and localities of specimens are given in Appendix 4.1 and their geographical location indicated in Figure 4.1. The included specimens were chosen to represent all known collection sites for the species, thus minimising geographical bias.

4.2.2 Pilot study

A pilot study was done to assess the variability and potential utility of a range of morphological characters for defining taxon boundaries within the *polystachyus* group. A subset of 68 specimens (about 32 % of the total 213 specimens studied) representing the geographical and ecological range of the *polystachyus* group was compiled. *Elaeocarpus integripetalus* was excluded because it is known only from the type specimen, which was unavailable during the pilot study. A total of 42 quantitative variables and 25 qualitative variables were scored. Analysis of this pilot dataset followed the methods used for the main analysis.

4.2.3 Character selection and measurement

Based on the results of the pilot investigation, a total of 39 quantitative (15 vegetative and 24 reproductive) and one qualitative (flower sexuality) variables (Table 4.2) that are potentially informative were selected to provide a rigorous test of the existing taxon boundaries (Coode 1996c). The remaining variables examined in the pilot study were excluded from further analysis because either 1) they showed insufficient variation in the morphometric analysis or 2) the character states could not be interpreted unambiguously due to e.g. the influence of the age of the plants and the preservation techniques (Table 4.2).

Measurements were made on mature organs to minimise variation due to age. Leaves were assumed to be mature if they were located behind inflorescences or infructescences, or in the absence of reproductive structures, at the fifth node from the twig tip. For most variables, measurements were made with digital callipers (0 – 150 mm) or a hand-held protractor. For the floral variables, selected flowers were boiled, dissected and measured against graph paper (1 × 1 mm) viewed through a dissecting microscope (WILD Heerbrugg M7 S, Switzerland, magnification 2.5 – 40 ×). Each variable, where possible, was measured five times on each specimen and then averaged. Data scored from different sheets but with the same collector and collection number were pooled and then averaged.

Elaeocarpus integripetalus is known only from the type specimen, therefore all floral variables (variables 16 – 34) were scored from the literature (Coode 1996c, and unpublished data) for this species. Fruit of this species is unknown.

All members of the *polystachyus* group have bisexual flowers (Fig. 4.2A), except *E. cupreus* and *E. multinervosus*, which have unisexual flowers (Fig. 4.2B & C). Floral variables of unisexual and bisexual flowers were incorporated to maximise comparison across specimens, minimise missing data, and minimise potential bias resulting from small sample sizes. Therefore measurements of the sterile stamens on the female flowers were pooled with those of the fertile stamens (variables 24 – 26); whereas in

male flowers (which lack the ovary), StyleLength (variable 30), OvaryLength (31), OvaryWidth (32), LoculiNo (33) and OvuleNo (34) were scored as missing data.

Table 4.2 Definition of variables scored for morphometric analysis and values of the evaluation analyses (i.e. Kruskal-Wallis (p-value), Principal Component Correlation (r^2) and Monte-Carlo Attributes in Ordination (MCAO)) for taxon boundaries, floristic provinces and subprovinces are given. Values for variable 1 – 35 were extracted from the combined vegetative and floral dataset and variable 36 – 40 from the combined vegetative and fruit dataset. * Most informative variables.

No.	Variable	Definition of variable	Taxon boundaries/ Floristic provinces/ Floristic subprovinces (Malay Peninsula)/ Floristic subprovinces (Borneo)		
			p-value	r^2	MCAO
1	TwigWidth	Width of growing twigs (mm)	45.96/ 45.96/ 6.27/ 36.46	0.729/ 0.729/ 0.087/ 0.753	0/ 0/ 85/ 0
2	*PetioleLength	Length of petioles (mm)	45.48/ 45.48/ 13.97/ 35.13	0.688/ 0.688/ 0.576/ 0.754	0/ 0/ 1/ 0
3	*PetioleWidth	Width of petioles measured at widest point (mm)	58.71/ 58.71/ 14.66/ 44.42	0.900/ 0.900/ 0.772/ 0.890	0/ 0/ 0/ 0
4	*DistalPulvinusWidth	Width of the distal pulvinus measured at widest point (mm)	48.39/ 48.39/ 15.94/ 40.77	0.849/ 0.849/ 0.652/ 0.837	0/ 0/ 0/ 0
5	*LeafLength	Length of lamina (mm)	48.39/ 56.11/ 14.27/ 48.80	0.814/ 0.814/ 0.682/ 0.835	0/ 0/ 1/ 0
6	*LeafWidth	Width of lamina measured at widest point (mm)	50.87/ 50.87/ 14.88/ 47.76	0.849/ 0.849/ 0.758/ 0.858	0/ 0/ 0/ 0
7	LeafRatio	LeafLength divided by LeafWidth	18.53/ 18.53/	0.208/ 0.208/	0/ 0/

			5.80/ 25.13	0.112/ 0.475	50/ 0
8	*LeafbaseWidestpoint	Distance along the midrib between leaf base and widest point of the leaf (mm)	50.85/ 50.85/ 13.99/ 48.45	0.820/ 0.820/ 0.636/ 0.838	0/ 0/ 0/ 0
9	LeafbaseAngle	Angle between leaf base and midrib measured at 5 mm distance from the leaf base and 90^{o} from the midrib	25.55/ 25.55/ 10.68/ 28.66	0.184/ 0.184/ 0.471/ 0.007	1/ 1/ 0/ 95
10	MarginteethLength	Distance between tips of teeth on leaf margin (mm)	29.99/ 29.99/ 9.74/ 34.60	0.350/ 0.350/ 0.234/ 0.364	0/ 0/ 23/ 0
11	AcumenLength	Distance between acumen tip and the second last pair of prominent secondary veins (mm)	43.10/ 43.10/ 9.16/ 27.38	0.277/ 0.277/ 0.105/ 0.367	0/ 0/ 86/ 0
12	*AbaxialMidribWidth	Width of midrib on abaxial surface measured at widest point (mm)	47.72/ 47.72/ 12.31/ 41.95	0.818/ 0.818/ 0.583/ 0.829	0/ 0/ 0/ 0
13	SecondaryVeins	Number of pairs of secondary veins	55.31/ 55.31/ 11.19/ 40.11	0.768/ 0.768/ 0.62/ 0.768	0/ 0/ 0/ 0
14	SecondaryVeinAngle	Angle between the second last pair of secondary veins and midrib measured at 90^{o} from the midrib	23.50/ 23.50/ 8.37/ 44.21	0.151/ 0.151/ 0.708/ 0.207	0/ 0/ 0/ 1
15	AbaxialSecondaryVeins Width	Width of secondary veins on abaxial surface measured at widest point (mm)	32.75/ 32.75/ 10.15/ 32.28	0.405/ 0.405/ 0.265/ 0.470	0/ 0/ 14/ 0
16	InflorescenceLength	Length of inflorescence, including terminal flowers (mm)	37.79/ 37.79/ 7.13/ 33.22	0.379/ 0.379/ 0.412/ 0.544	0/ 0/ 3/ 0
17	InflorescenceWidth	Width of inflorescence measured at widest point (mm)	51.43/ 51.43/ 6.74/ 39.19	0.722/ 0.722/ 0.275/ 0.782	0/ 0/ 18/ 0
18	FlowerNo	Number of flowers per inflorescence	31.86/ 31.86/ 11.57/	0.099/ 0.099/ 0.539/	6/ 6/ 1/

			35.40	0.111	9
19	PedicelLength	Length of pedicel (mm)	33.22/ 33.22/ 8.01/ 33.18	0.320/ 0.320/ 0.630/ 0.383	0/ 0/ 0/ 0
20	SepalLength	Length of sepal (mm)	35.01/ 35.01/ 9.49/ 33.29	0.509/ 0.509/ 0.582/ 0.410	0/ 0/ 0/ 0
21	SepalWidth	Width of sepal measured at widest point (mm)	50.64/ 50.64/ 12.22/ 35.71	0.702/ 0.702/ 0.759/ 0.347	0/ 0/ 0/ 0
22	PetalLength	Length of petal (mm)	35.33/ 35.33/ 10.18/ 39.90	0.293/ 0.293/ 0.418/ 0.357	0/ 0/ 1/ 0
23	PetalWidth	Width of petal measured at widest point (mm)	42.23/ 42.23/ 11.67/ 30.29	0.682/ 0.682/ 0.603/ 0.350	0/ 0/ 0/ 0
24	StamenNo	Number of fertile/sterile stamens per flower	30.30/ 30.30/ 7.83/ 39.85	0.253/ 0.253/ 0.315/ 0.288	0/ 0/ 3/ 0
25	FilamentLength	Filament length (mm)	15.91/ 15.91/ 6.88/ 22.25	0.227/ 0.227/ 0.207/ 0.072	0/ 0/ 31/ 18
26	AntherLength	Length of fertile/sterile anthers (mm)	20.86/ 20.86/ 10.64/ 25.41	0.228/ 0.228/ 0.253/ 0.434	0/ 0/ 41/ 0
27	SepalNo	Number of sepals per flower	36.72/ 36.72/ 7.19/ 5.75	0.626/ 0.626/ 0.770/ 0.104	0/ 0/ 0/ 7
28	PetalNo	Number of petals per flower	36.47/ 36.47/ 6.62/ 6.88	0.626/ 0.626/ 0.704/ 0.249	0/ 0/ 0/ 0
29	FlowerMery	Average of SepalNo and PetalNo	41.71/ 41.71/ 7.41/ 9.69	0.686/ 0.686/ 0.791/ 0.248	0/ 0/ 0/ 0

30	StyleLength	Length of style (mm)	13.62/ 13.62/ 10.34/ 10.36	0.112/ 0.112/ 0.540/ 0.219	12/ 12/ 0/ 0
31	OvaryLength	Length of ovary (mm)	8.00/ 8.00/ 3.14/ 21.03	0.149/ 0.149/ 0.375/ 0.364	5/ 5/ 9/ 0
32	OvaryWidth	Width of ovaries measured at widest point (mm)	13.98/ 13.98/ 5.65/ 24.15	0.099/ 0.099/ 0.620/ 0.330	5/ 5/ 0/ 0
33	LoculiNo	Number of loculi per ovary	28.23/ 28.23/ 2.68/ 24.47	0.345/ 0.345/ 0.201/ 0.379	0/ 0/ 35/ 0
34	OvuleNo	Number of ovules per loculus	37.22/ 37.22/ 6.67/ 15.54	0.533/ 0.533/ 0.588/ 0.178	0/ 0/ 0/ 6
35	FlowerSexuality	The presence of bisexual or unisexual flowers	45.60/ 45.60/ 5.10/ 39.34	0.544/ 0.544/ 0.775/ 0.538	0/ 0/ 0/ 0
36	InfructescenceWidth	Width of infructescence measured at widest point (mm)	42.01/ 39.71/ 6.42/ 38.98	0.647/ 0.631/ 0.257/ 0.794	0/ 0/ 18/ 0
37	FruitStalkLength	Length of fruit stalk (mm)	19.38/ 19.33/ 2.44/ 23.56	0.149/ 0.159/ 0.483/ 0.506	1/ 0/ 2/ 0
38	FruitStalkWidth	Width of fruit stalk measured at widest point (mm)	18.53/ 18.19/ 11.47/ 32.98	0.256/ 0.251/ 0.448/ 0.608	0/ 0/ 1/ 0
39	FruitLength	Length of fruit (mm)	21.10/ 12.27/ 8.95/ 29.94	0.124/ 0.127/ 0.341/ 0.571	1/ 2/ 0/ 0
40	FruitWidth	Width of fruit measured at widest point (mm)	23.22/ 18.43/ 8.69/ 26.30	0.041/ 0.218/ 0.153/ 0.583	55/ 0/ 32/ 0

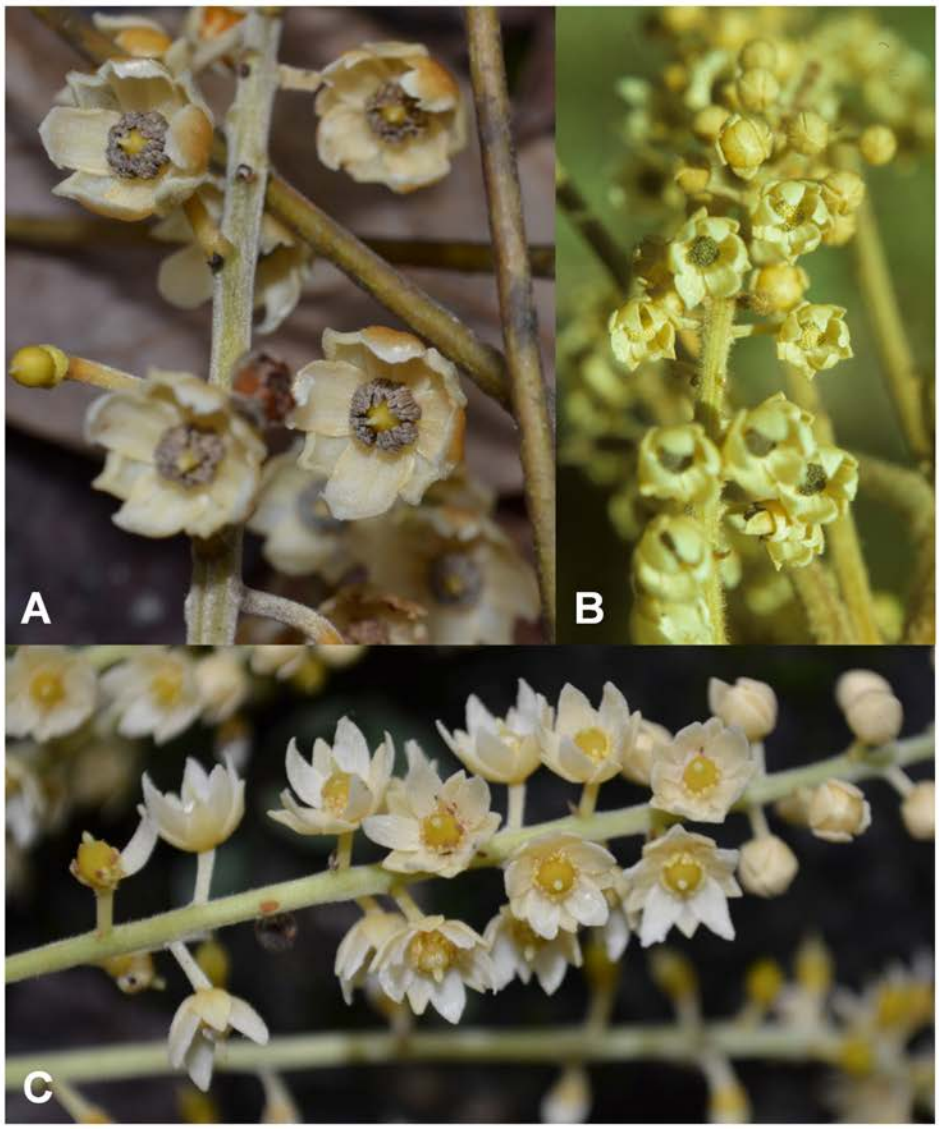


Fig. 4.2 Flower sexuality in the *polystachyus* group. A, Typical bisexual flowers of *Elaeocarpus polystachyus*; B, Male flowers of *E. multinervosus* (photo: Coode); C, Female flowers of *E. multinervosus*.

4.2.4 Missing data

Of the 213 specimens, two (0.9 %) were incomplete for the vegetative dataset, 131 (61.5 %) were incomplete for the floral and the fruit datasets, and eight (3.8 %) could not be assigned unambiguously to one of the phytogeographical subprovinces due to poor locality data. Specimens with incomplete data were excluded from analysis.

4.2.5 Phytogeography

Correlations between morphological variation and ecology were investigated using phytogeographical affinities as broad ecological proxies (Fig. 4.3). Data were analysed at floristic province and subprovince levels. The definition of floristic provinces within West Malesia is based on general geographical isolation and flora composition (dominant plant species) (Ashton 1992; Wong 1998). Three floristic provinces within the distribution range of the *polystachyus* group were identified: Borneo (B), Malay Peninsula (MP) and Sumatra (S). The Malay Peninsula is a floristic area running from Singapore northward to the Alor Star-Songkhla line, a boundary between Alor Star (Northern Peninsular Malaysia) and Songkhla ("Singgora" in Van Steenis, Southern Thailand), which marks a significant change in floristic composition between West Malesia and the mainland Asia (Fig. 4.3) (Van Steenis 1950). However, species of the *polystachyus* group are absent from Southern Thailand, therefore the term "Malay Peninsula" in this study is referring to Peninsular Malaysia and Singapore only.

The floristic subprovince classification of Wong (1998) is a detailed ecological phytogeography based on the combination of flora composition, floristic influences (element of the dominant plant species), elevation of habitat and edaphic types. Eleven floristic subprovinces were identified in the *polystachyus* group: Lowland Flora in Borneo (B), East Coast Sabah Floristic Subprovince (ECSS), *Kerangas* Forest (K), Mountain Flora in Borneo (MFB), Mountain Flora in Malay Peninsula (MFMP), Lowland Flora in Malay Peninsula (MP), Seasonally Flooded Riverside (R), Riau Pocket (RP),

Sumatra (S), Seasonal Asiatic-Australasian Intrusion in North Borneo (SAAI), Seasonal Asiatic Intrusion (SAI). Only two Sumatran OTUs were available, therefore they were analysed together with the Malay Peninsula dataset.

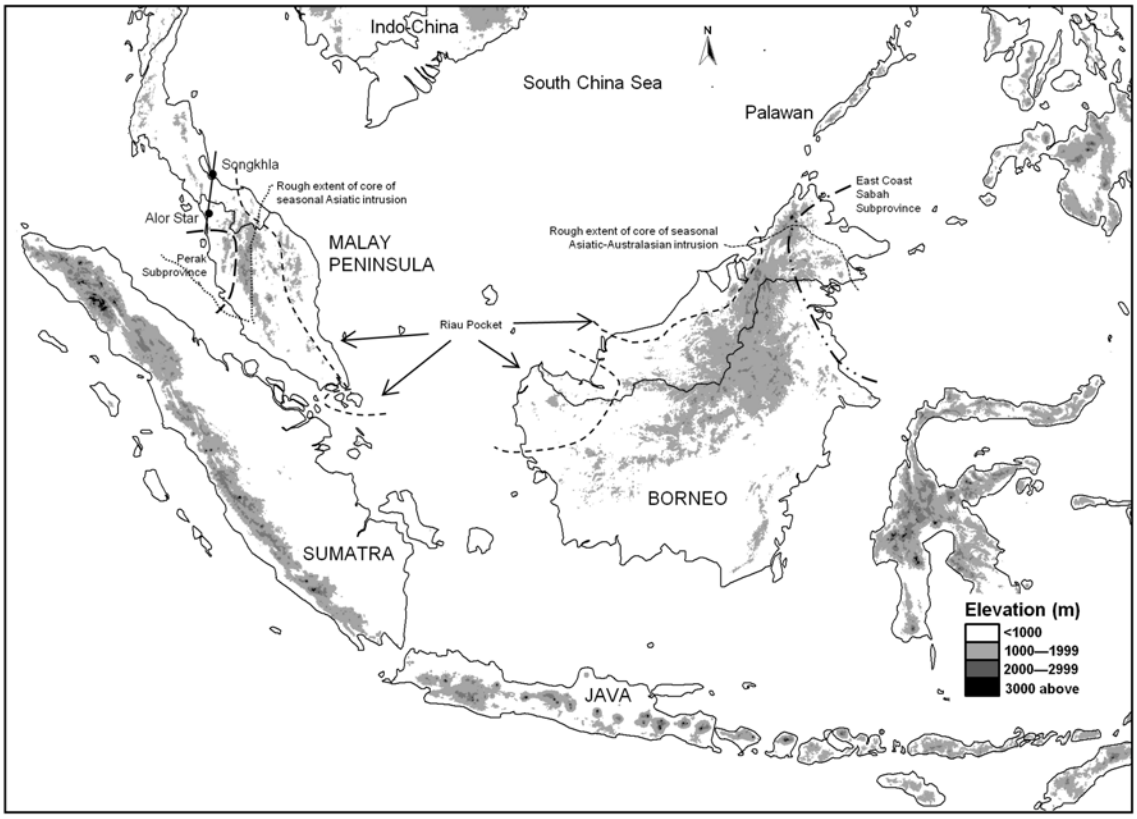


Fig. 4.3 Phytogeographical areas at the floristic province and subprovince levels in the distribution range of the *polystachyus* group (modified from Wong 1998, with permission from Institute of Botany, Academia Sinica Monograph Series).

4.2.6 Data analysis

The data were compiled using Microsoft® Excel® for Mac 2011 version 14.1.4 (Microsoft Corporation 2010), and ordination and classification analyses performed on the untransformed data using PATN version 3.12 (Blatant Fabrications Pty. Ltd.). Five datasets were constructed and analysed: vegetative, floral, fruit (each analysed

independently), vegetative plus floral (hereafter named veg+fl) and vegetative plus fruit (hereafter named veg+fr) variables.

The ordination analysis used Semi-Strong Hybrid Multidimensional Scaling (SSH MDS). SSH MDS uses a combination of metric and non-metric regressions, first by randomly distributing the OTUs in the selected three-dimensional space followed by a series of measuring and comparing the distances between each pair of OTUs and redistributing until all of them are "best-fit" in the space (Belbin 1993). The following default parameters were applied for the analyses: cut-off value 0.9, 10 random starts, 1235 random seed and a maximum of 50 iterations. These parameters can accommodate the asymptotic tails in a linear relationship, common in biological data (Belbin 1993). Minimum Spanning Trees (MST), which link any two OTUs at their shortest distance, were mapped on the ordination plots to provide a better view of the ordination structure for those OTUs that do not fit well in the space (Gower and Ross 1969).

An alternative morphometric assessment was done using the classification analysis. A distance matrix was estimated using the Gower metric association measure, which allows different measurement scales and zero values (Gower 1971). The Agglomerative Hierarchical Clustering with Flexible UPGMA algorithm was used, where OTUs were paired based on the estimated nearest distance (or highest degree of similarity based on selected variables) (Sneath and Sokal 1973). The analysis applied a beta value of -0.1 (default in PATN) to neutralise the underestimation of large association values and to control the perceived association values as the agglomeration process runs (Belbin 1984).

The most informative variables for segregating the OTUs in the ordination analysis were determined using Principal Component Correlation (PCC) and Monte-Carlo Attributes in Ordination (MCAO). PCC employs a multiple linear regression approach to find the best-fit of selected variables in three-dimensional ordination space and estimates the degree of co-linearity between two variables. The coefficient of determination is an estimation of the proportion of variance between two variables,

where r^2 = 0.6 is used as the cut-off point to identify the most informative variables in the analyses (Belbin 1993). MCAO evaluates the PCC results by randomly redistributing the variables in the ordination space and calculates the frequency with which the r^2 value exceeds its true value over 100 iterations (default in PATN). A low frequency indicates that the two variables are highly correlated (Belbin 1993). For the classification analysis, variables were evaluated using Kruskal-Wallis One-Way Analysis of Variance by Ranks (KW). KW is a non-parametric method suitable for non-normally distributed data (Kruskal and Wallis 1952).

4.3 Results

Results of the morphometric analysis based on the vegetative, floral, fruit, veg+fl and veg+fr datasets were highly similar in their clustering patterns for both ordination and classification analyses. The results of vegetative, floral and fruit datasets are not shown here.

4.3.1 Analysis of combined vegetative and floral data

The three-dimensional ordination plot shows spatial clustering patterns among the specimens studied and retrieved six discrete clusters corresponding to six taxa at the species level (Fig. 4.4, stress value = 0.1392; Supplementary Figure 4.1). All of the OTUs assigned to *E. cupreus* and *E. multinervosus* formed discrete clusters, respectively. Most of the OTUs assigned to *E. clementis, E. polyanthus* and *E. polystachyus* formed clusters with their conspecifics. *Elaeocarpus integripetalus* was represented by a single specimen, which came out independently.

The fifteen most informative variables (r^2 > 0.6) contributing to the clustering patterns were determined by the evaluation analyses and are summarised in Table 4.2. These are (arranged from the highest to the lowest r^2 value): PetioleWidth (variable 3), LeafWidth (6), DistalPulvinusWidth (4), LeafbaseWidestpoint (8), AbaxialMidribWidth (12), LeafLength (5), SecondaryVeins (13), TwigWidth (1), InflorescenceWidth (17), SepalWidth (21), PetioleLength (2), FlowerMery (29),

PetalWidth (23), PetalNo (28) and SepalNo (27). All of the variables are correlated, except variables 21 and 23 and the direction of the vectors is shown next to the ordination plots (Fig. 4.4).

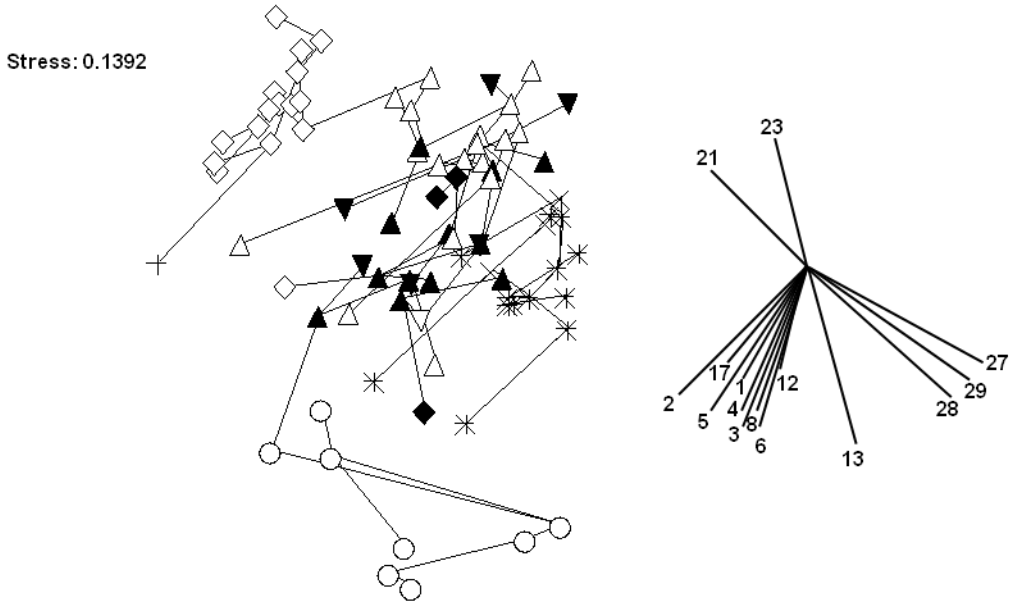


Fig. 4.4 Minimum spanning tree (MST) and Semi-strong Hybrid Multidimensional Scaling on the two-dimensional plane (axis 1 and 3) viewed through three-dimensional ordination space showing relationship of the nine taxa within the *polystachyus* group based on the combined vegetative and floral variables: *E. clementis* (var. *borneensis* ▲, var. *clemensiae* ▼, var. *clementis* △ and var. *kostermansii* ▽), *E. cupreus* ❋, *E. integripetalus* ┼, *E. multinervosus* ○, *E. polyanthus* ◆ and *E. polystachyus* ◇. A, Combined vegetative and floral dataset. Vectors of the most informative variables (r^2 > 0.6) as determined by PCC are shown in the direction of correlation. Variables are: TwigWidth (variable 1), PetioleLength (2), PetioleWidth (3), DistalPulvinusWidth (4), LeafLength (5), LeafWidth (6), LeafbaseWidestpoint (8), AbaxialMidribWidth (12), SecondaryVeins (13), InflorescenceWidth (17), SepalWidth (21), PetalWidth (23), SepalNo (27), PetalNo (28), FlowerMery (29).

The dendrogram is generally less resolved compared to the ordination plot. Nine major clusters were retrieved (Fig. 4.5). Cluster 1A mostly comprised OTUs assigned to *E. clementis*, with two OTUs of *E. polyanthus* (S 36493 and Omar 352) and one of *E. polystachyus* (FRI 717); whereas cluster 2A comprised only OTUs assigned to *E. clementis*. Cluster 3A mostly comprised OTUs of *E. cupreus*, with an OTU of *E. clementis* (WKM 298), while the rest of the OTUs assigned to *E. cupreus* were placed in Cluster 4A. Cluster 5A consisted of the single OTU of *E. integripetalus*. Other than FRI 717, the OTUs assigned to *E. polystachyus* were all in the Cluster 6A. Cluster 7A, 8A and 9A comprised only OTUs of *E. multinervosus*, except in the last cluster, where an OTU of *E. polyanthus* (S 27991) was nested here.

The set of variables considered most informative for separating clusters as indicated by the KW value is slightly different from the ordination analysis. PetalNo (28, p-value = 36.47) and SepalNo (27, p-value = 36.72) are considered less informative than FlowerSexuality (35, p-value = 45.60) and AcumenLength (11, p-value = 43.10) (Table 4.2).

4.3.2 Analysis of combined vegetative and fruit data

The three-dimensional ordination plot retrieved five major clusters corresponding to five taxa at the species level (*E. integripetalus* omitted) (Fig. 4.6, stress value = 0.1393; Supplementary Figure 4.2). Two clusters were distinctive, comprising only the OTUs of their conspecifics (i.e. *E. cupreus* and *E. polyanthus*). The remaining three clusters comprised a majority of the OTUs of *E. clementis*, *E. multinervosus* and *E. polystachyus*, respectively, but there were some OTUs assigned to a different taxon nested within them.

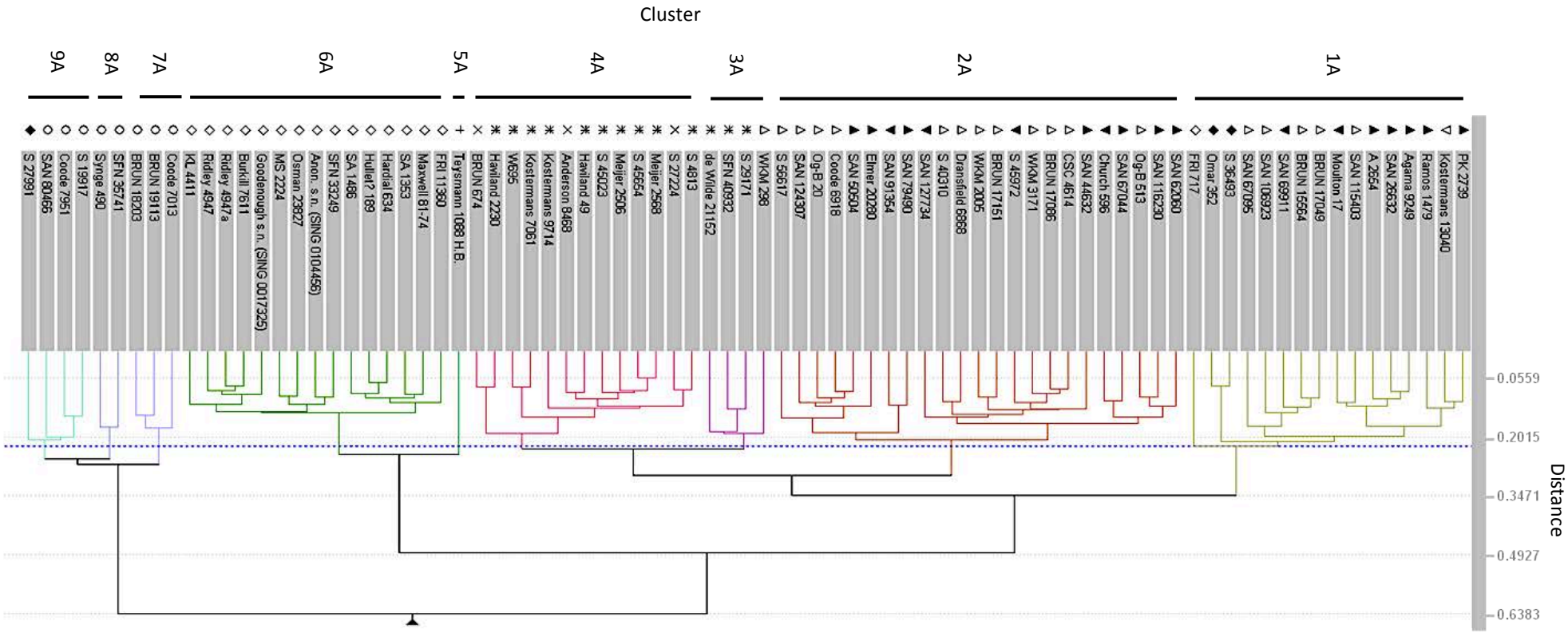


Fig. 4.5 Dendrograms of the nine taxa within the *polystachyus* group produced from the Agglomerative Hierarchical Clustering with Flexible UPGMA technique and distance estimated with Gower metric association measure based on the combined vegetative and floral dataset. OTUs represent the following taxa: *E. clementis* (var. *borneensis* ▲, var. *clemensiae* ▼, var. *clementis* △ and var. *kostermansii* ▽), *E. cupreus* ✱, *E. integripetalus* ＋, *E. multinervosus* ○, *E. polyanthus* ◆ and *E. polystachyus* ◇. Blue dotted line indicates grouping cut off point.

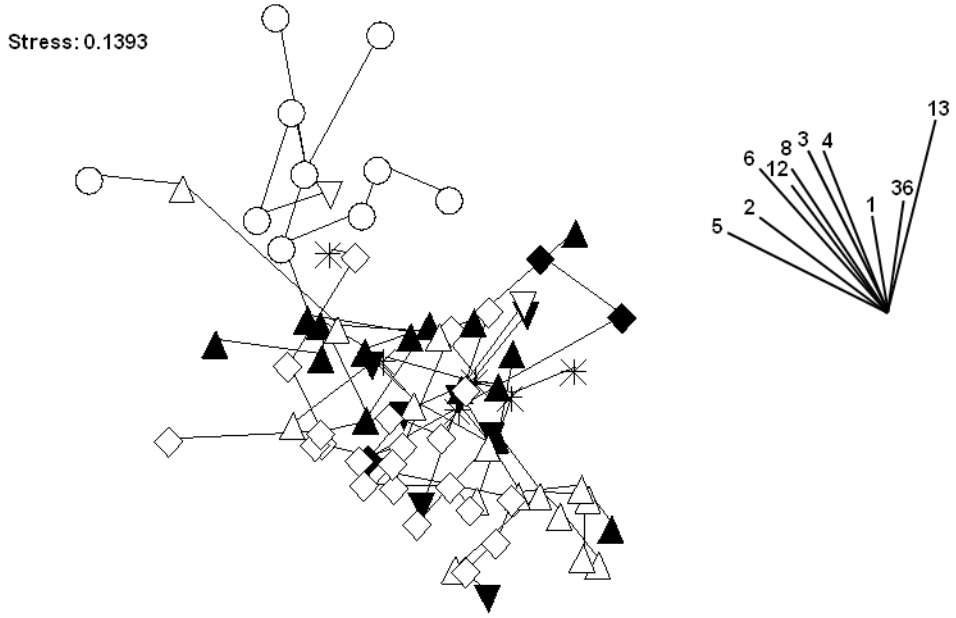


Fig. 4.6 Minimum spanning tree (MST) and Semi-strong Hybrid Multidimensional Scaling on the two-dimensional plane (axis 1 and 3) viewed through three-dimensional ordination space showing relationship of the nine taxa within the *polystachyus* group based on the combined vegetative and fruit variables: *E. clementis* (var. *borneensis* ▲, var. *clemensiae* ▼, var. *clementis* △ and var. *kostermansii* ▽), *E. cupreus* ✱, *E. integripetalus* ┼, *E. multinervosus* ○, *E. polyanthus* ◆ and *E. polystachyus* ◇. A, Vegetative + floral dataset. Vectors of the most informative variables (r^2 > 0.6) as determined by PCC are shown in the direction of correlation. Variables are: TwigWidth (variable 1), PetioleLength (2), PetioleWidth (3), DistalPulvinusWidth (4), LeafLength (5), LeafWidth (6), LeafbaseWidestpoint (8), AbaxialMidribWidth (12), SecondaryVeins (13) and InfructescencesWidth (36).

The ten most informative variables (r^2 > 0.6) contributing to the clustering patterns were determined in the evaluation analyses (Table 4.2). These are (arranged from the highest to the lowest r^2 value): LeafWidth (6), DistalPulvinusWidth (4), LeafLength (5), PetioleWidth (3), AbaxialMidribWidth (12), LeafbaseWidestpoint (8), TwigWidth (variable 1), PetioleLength (2), SecondaryVeins (13) and FruitStalkLength (36), and all of them are correlated. The direction of the vectors is shown next to the ordination plots in Fig. 4.6.

Nine major clusters were retrieved in the classification analysis corresponding to four taxa at the species level; OTUs assigned to *E. cupreus* did not form a discrete cluster and *E. integripetalus* was not included because fruits of this species are unknown (Fig. 4.7). The results of the classification analysis showed less resolution groups compared to the ordination analysis. Cluster 1B comprised mostly the OTUs assigned to *E. clementis*, with four OTUs of *E. cupreus* (KEP 108999, Wiriadinata 651, Amdjah 1070 and Meijer 2506) and one of *E. polyanthus* (Ghazalli 13408) nested within it. Cluster 2B comprised mostly the OTUs of *E. polystachyus*, with four OTUs of *E. clementis* (SFN 26475, CWL 413, Church 596 and Mahyar 798) and two of *E. cupreus* (FRI 15900 and SFN 32414) nested within it. Cluster 3B comprised OTUs of *E. clementis* and only of their conspecifics. Cluster 4B, 5B and 6B comprised the OTUs of *E. multinervosus*, but in the first cluster, four OTUs assigned to *E. clementis* (A 1612, S 38206, S 38905 and Kostermans 13101) were nested here. Cluster 7B and 8B each comprised of a single OTU of *E. polystachyus* and *E. clementis*, respectively. Cluster 9B comprised only the OTUs of *E. polyanthus*.

The set of most informative characters as indicated by the KW values (Table 4.2) is similar to that from the ordination analysis, except that MarginteethLength (variable 10, p-valued = 43.37) is evaluated as more informative than FruitStalkLength (36, p-valued = 42.01).

4.3.3 Informative variables

On the evaluation of the utility of the individual variables in separating the *polystachyus* specimens, nine were found to have strong effects in both the veg+fl and veg+fr analyses. These are: TwigWidth (variable 1), PetioleLength (2), PetioleWidth (3), DistalPulvinusWidth (4), LeafLength (5), LeafWidth (6), LeafbaseWidestpoint (8), AbaxialMidribWidth (12) and SecondaryVeins (13). Two variables (10 and 11) show high KW values in the veg+fr and veg+fl analyses, respectively, but the r^2 values are very low (Table 4.2). The floral (variable 17, 21, 23, 27, 28, 29 and 35) and fruit (36) variables are less informative in all analyses in comparison to the vegetative ones. The

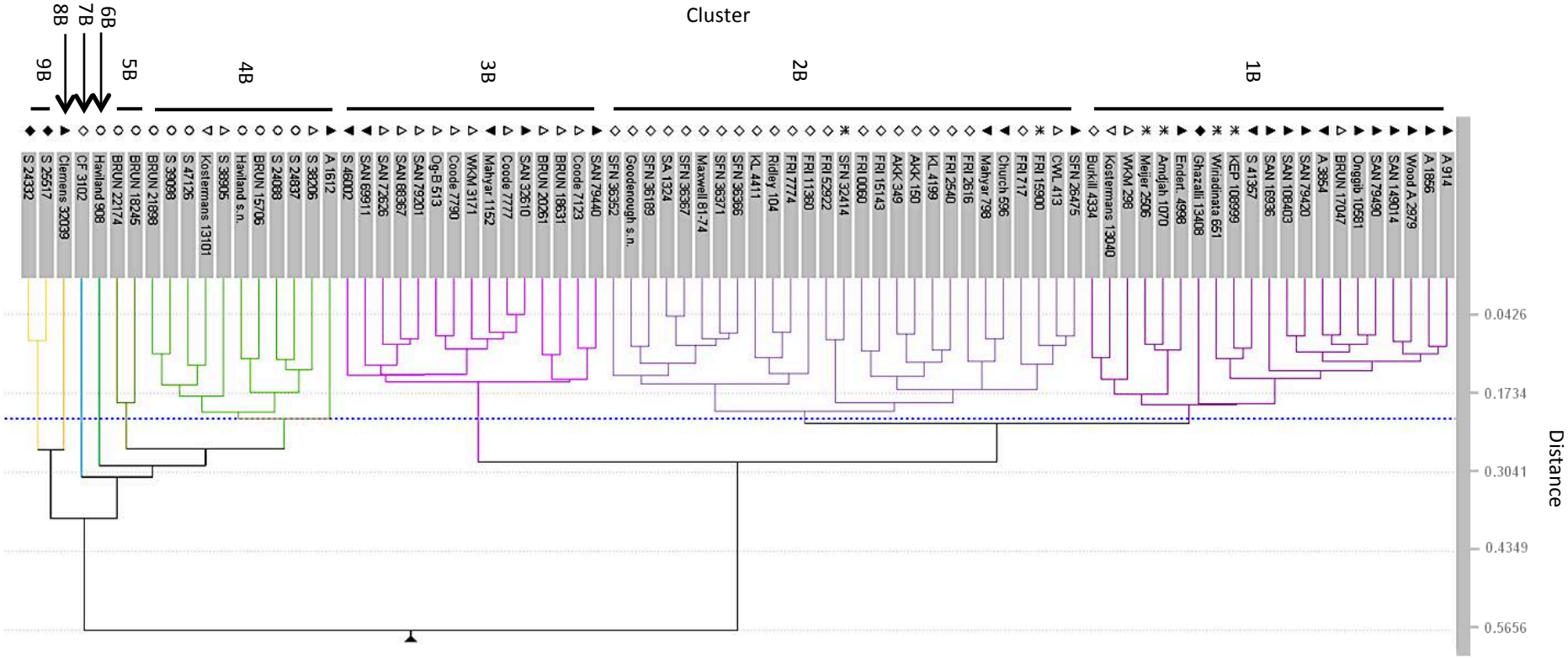


Fig. 4.7 Dendrograms of the nine taxa within the *polystachyus* group produced from the Agglomerative Hierarchical Clustering with Flexible UPGMA technique and distance estimated with Gower metric association measure based on the combined vegetative and fruit dataset. OTUs represent the following taxa: *E. clementis* (var. *borneensis* ▲, var. *clemensiae* ▼, var. *clementis* △ and var. *kostermansii* ▽), *E. cupreus* ✱, *E. integripetalus* ┼, *E. multinervosus* ○, *E. polyanthus* ◆ and *E. polystachyus* ◇. Blue dotted line indicates grouping cut off point.

six taxa or groups are best differentiated by a combination of vegetative and reproductive characters.

Three variables in the veg+fl analysis (variables 30, 31 and 32) and six (7, 14, 37, 39 and 40) in the veg+fr analysis showed very low KW values and thus were the least informative in resolving clusters. However, removing these variables did not obviously affect the results of the ordination and classification analyses (results not shown).

4.3.4 Morphological polymorphism

Two floral variables, flower-mery and flower sexuality (variables 29 and 35), exhibited polymorphism in nine of the 213 specimens examined (4.2 %). In the 4-merous flower species, two specimens of *E. polystachyus* (FRI 717 and SA 1353) from the Malay Peninsula have 4 or 5 sepals and 5 petals. On the other hand, in the 5-merous species, two specimens of *E. clementis* var. *clemensiae* (Dransfield 6868 and SAN 67095) from Borneo have 4 sepals and 5 petals; one specimen of *E. multinervosus* (SFN 35741) from Borneo has 4 sepals and 5 petals.

In the bisexual-flowered species, one specimen of *E. clementis* var. *clemensiae* (S 21889) from Borneo has both male and bisexual flowers. The male flowers have numerous stamens, densely arranged in multiple tiers towards the flower centre, with slightly shorter stamens rather than an ovary in the centre. The possibility of the ovary having fallen or been removed from the flowers was discounted because no scar on the receptacle was observed. Similarly, in the unisexual flowered species, three specimens of *E. cupreus* (BRUN 130, BRUN 553 and BRUN 674) from Borneo show inflorescences with up to 5 bisexual flowers amongst the much more numerous unisexual flowers.

Ordination and classification analyses were done using the vegetative dataset to test the effects of polymorphism on the results. Five of the specimens with polymorphic flower-mery (SA 1353, Dransfield 6868, SAN 67095, SFN 35741) or flower

sexuality (S 21889) grouped with their conspecific OTUs, while another four (polymorphic flower-mery: FRI 717; flower sexuality: BRUN 130, BRUN 553 and BRUN 674) are nested in mixed clusters (data not shown). SFN 35741 was unplaced in the analyses (BRUN 130 and BRUN 553 were missing the floral variables, thus omitted). Nonetheless, the results of these analyses were not significantly different from the veg+fl dataset.

4.3.5 Correlation between morphology and phytogeography

The morphological analysis indicates that with the exception of Cluster 3A, all of the clusters within the *polystachyus* group are generally endemic to a single floristic province. *Elaeocarpus cupreus* is the most widely distributed taxon within the *polystachyus* group occurring throughout the distribution range of the group, while the remaining taxa are geographically more restricted: *E. integripetalus* is endemic to Sumatra, *E. polystachyus* is endemic to the Malay Peninsula and *E. clementis*, *E. multinervosus* and *E. polyanthus* are endemic to Borneo. Correlation between morphological variation and floristic provinces using the veg+fl and veg+fr datasets is shown in Fig. 4.8A (stress value = 0.1392, Supplementary Figure 4.3) and 4.8B (stress value = 0.1445, Supplementary Figure 4.4), respectively. *Elaeocarpus integripetalus*, the single specimen of which lacks fruit, was omitted in all analyses of the veg+fr dataset. The direction of the vectors is shown next to each ordination plot.

No apparent clustering patterns are discernible at the floristic subprovince level and all taxa appear to be generalists instead of subprovince-specific. In the Sumatra and Malay Peninsula provinces, the three taxa (*E. cupreus*, *E. integripetalus* and *E. polystachyus*) are distributed in five floristic subprovinces: MFMP, MP, RP, S and SAI; while in the Borneo province, the seven taxa (*E. clementis* var. *borneensis*, *E. clementis* var. *clemensiae*, *E. clementis* var. *clementis* and *E. clementis* var. *kostermansii*, *E. cupreus*, *E. multinervosus* and *E. polyanthus*) are recorded in seven floristic subprovinces: B, ECSS, K, MFB, R, RP and SAAI. The correlations between morphological variation and floristic subprovinces in the Malay Peninsula and Sumatra are shown in Fig. 4.9A (veg+fl dataset, stress value = 0.1263, Supplementary Figure

4.5) and 4.9B (veg+fr dataset, stress value = 0.1164, Supplementary Figure 4.6), whereas for Borneo, the ordination plots are in Fig. 4.10A (veg+fl dataset, stress value = 0.1478, Supplementary Figure 4.7) and 4.10B (veg+fr dataset, stress value = 0.1164, Supplementary Figure 4.8).

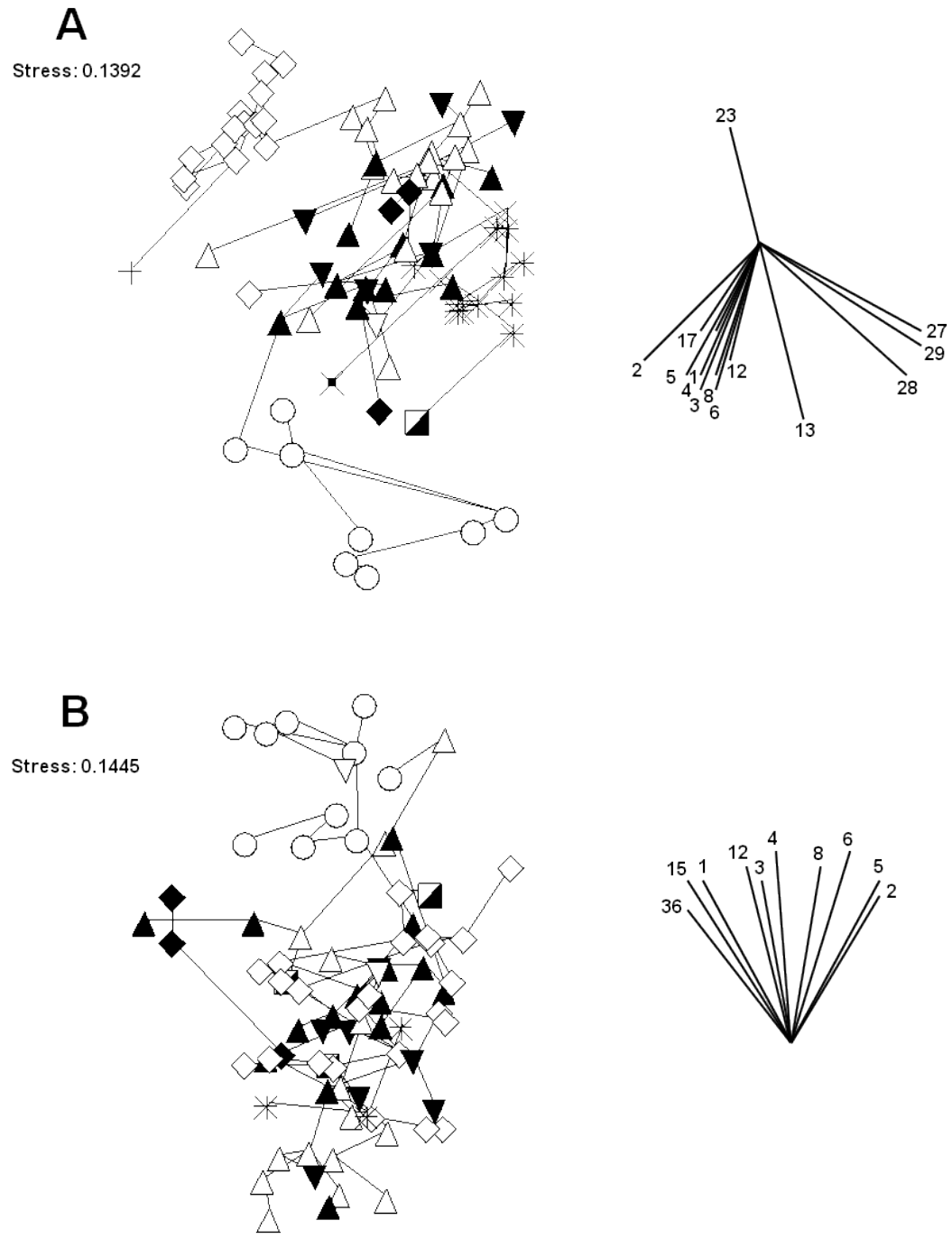


Fig. 4.8. Minimum spanning tree (MST) and Semi-strong Hybrid Multidimensional Scaling on the two-dimensional plane viewed through three-dimensional ordination space showing correlation between morphological variation and floristic provinces: *E. clementis* in Borneo (var. *borneensis* △, var. *clemensiae* ▽, var. *clementis* ▲ and var. *kostermansii* ▼), *E. cupreus* in Sumatra ✹, *E. cupreus* in Malay Peninsula ◪, *E. cupreus* in Borneo ✳, *E. integripetalus* in Sumatra ┼, *E. multinervosus* in Borneo ○, *E. polyanthus* in Borneo ◆ and *E. polystachyus* in Malay Peninsula ◇. A, Combined

vegetative and floral dataset (axis 1 and 2). Vectors of the most informative variables (r^2 > 0.6) as determined by PCC are shown in the direction of correlation. Variables are: TwigWidth (variable 1), PetioleLength (2), PetioleWidth (3), DistalPulvinusWidth (4), LeafLength (5), LeafWidth (6), LeafbaseWidestpoint (8), AbaxialMidribWidth (12), SecondaryVeins (13), InflorescenceWidth (17), PetalWidth (23), SepalNo (27), PetalNo (28) and FlowerMery (29); B, Combined vegetative and fruit dataset (axis 2 and 3). Vectors of the most informative variables (r^2 > 0.6) as determined by PCC are shown in the direction of correlation. Variables are: TwigWidth (variable 1), PetioleLength (2), PetioleWidth (3), DistalPulvinusWidth (4), LeafLength (5), LeafWidth (6), LeafbaseWidestpoint (8), AbaxialMidribWidth (12), AbaxialSecondaryVeinsWidth (15) and InfructescenceWidth (36).

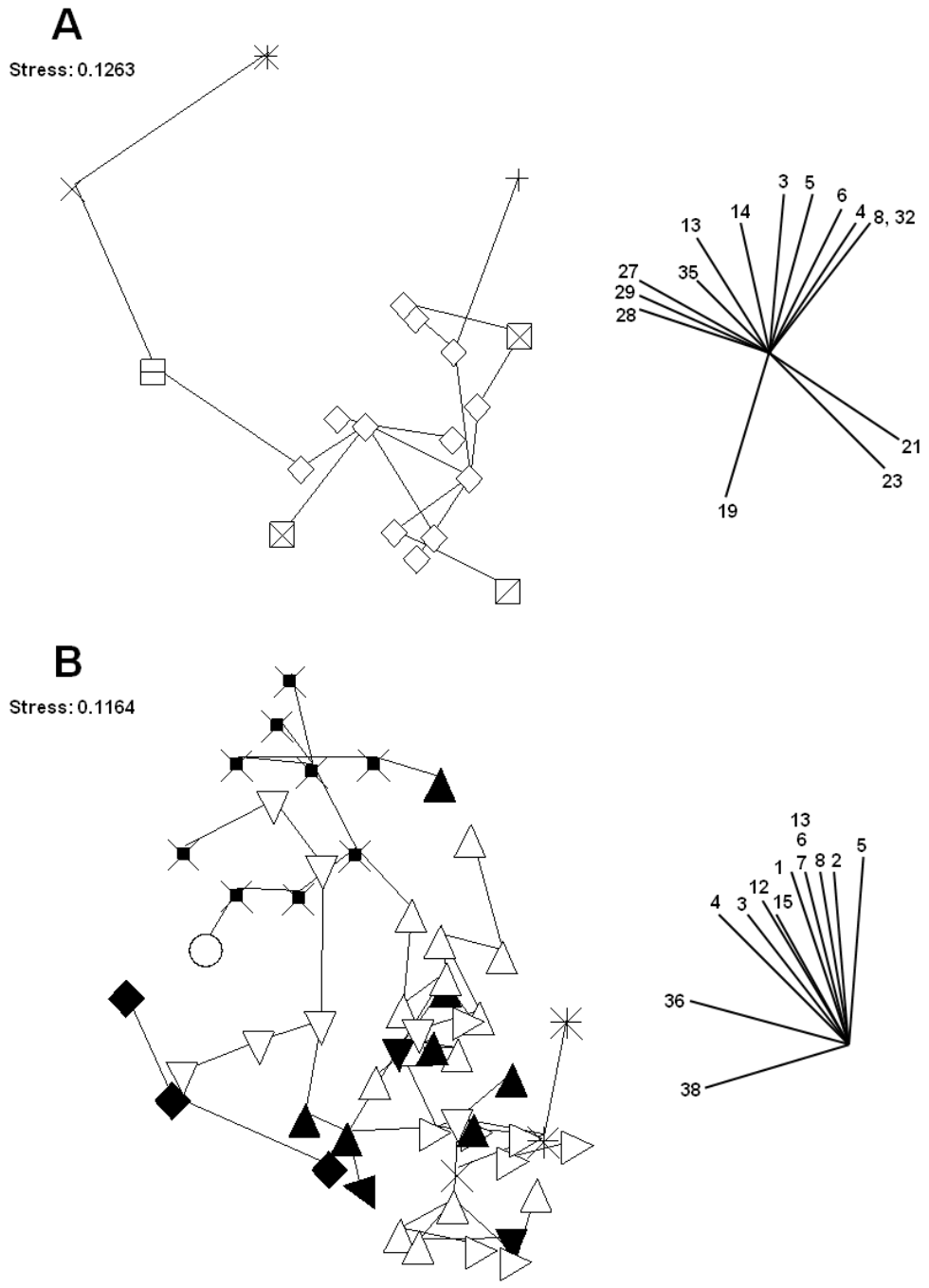


Fig. 4.9 Minimum spanning tree (MST) and Semi-strong Hybrid Multidimensional Scaling on the two-dimensional plane viewed through three-dimensional ordination space showing correlation between morphological variation and floristic subprovinces in Sumatra and Malay Peninsula: *E. cupreus* in Sumatra (S) ×, *E. cupreus* in Riau Pocket (RP) in vegetative + floral dataset ⋇ or Lowland Flora in Malay Peninsula in vegetative + fruit dataset ✳, *E. integripetalus* in S ┼, *E. polystachyus* in Mountain Flora in Malay Peninsula ☒, *E. polystachyus* in Lowland Flora in Malay Peninsula ◇, *E. polystachyus* in

RP ⧄ and *E. polystachyus* in Seasonal Asiatic Intrusion ⧅ . A, Combined vegetative and floral dataset (axis 1 and 2). Vectors of the most informative variables (r^2 > 0.6) as determined by PCC are shown in the direction of correlation. Variables are: PetioleWidth (variable 3), DistalPulvinusWidth (4), LeafLength (5), LeafWidth (6), LeafbaseWidestpoint (8), SecondaryVeins (13), SecondaryVeinAngle (14), PedicelLength (19), SepalWidth (21), PetalWidth (23), SepalNo (27), PetalNo (28), FlowerMery (29), OvaryWidth (32) and FlowerSexuality (35); B, Combined vegetative and fruit dataset (axis 1 and 2). Vectors of the most informative variables (r^2 > 0.6) as determined by PCC are shown in the direction of correlation. Variables are: (variable 1), PetioleLength (2), PetioleWidth (3), DistalPulvinusWidth (4), LeafLength (variable 5), LeafWidth (6), LeafRatio (7), LeafbaseWidestpoint (8), AbaxialMidribWidth (12), SecondaryVeins (13), AbaxialSecondaryVeinsWidth (15), InfructescencesWidth (36) and FruitStalkWidth (38).

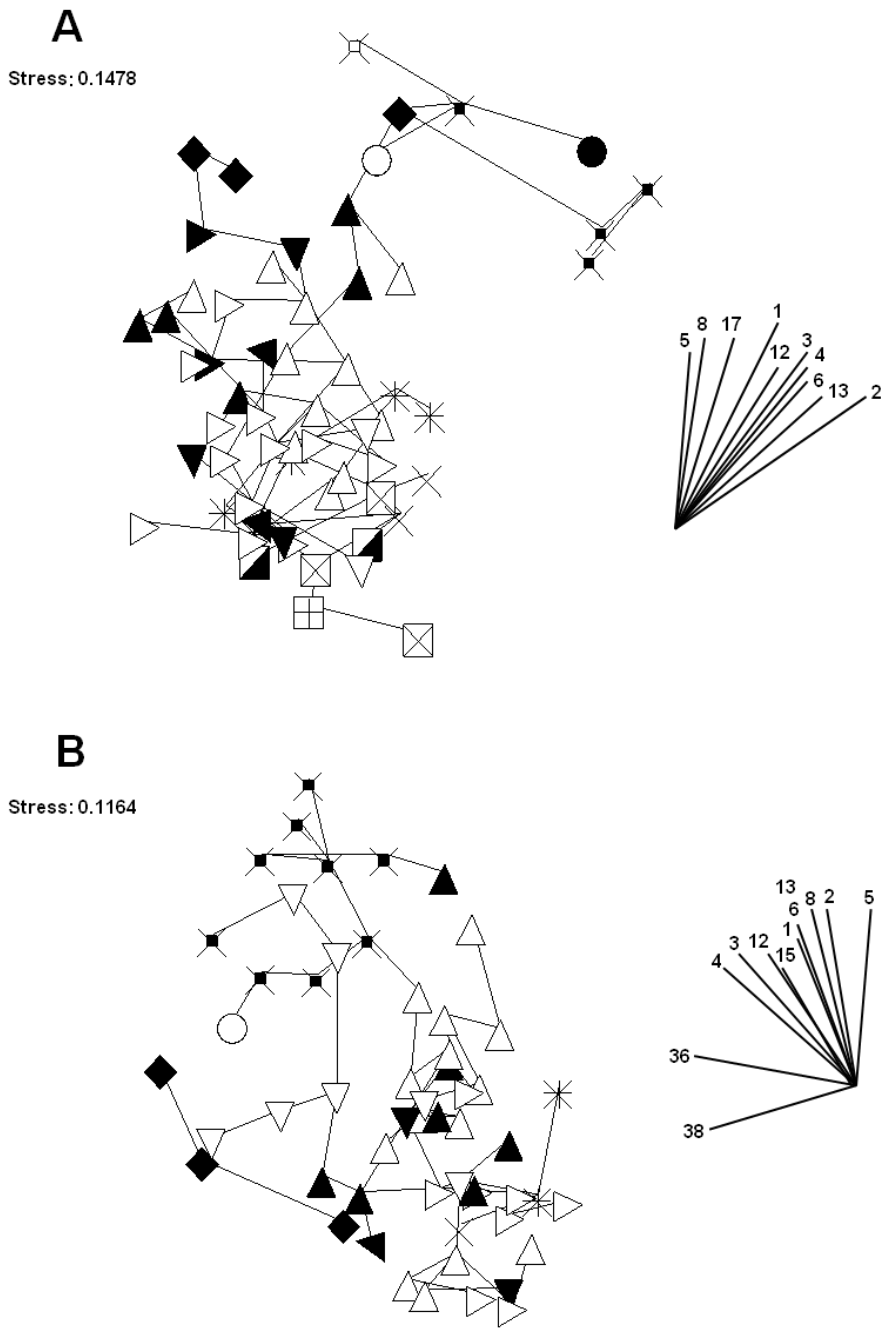


Fig. 4.10 Minimum spanning tree (MST) and Semi-strong Hybrid Multidimensional Scaling in the two-dimensional plane viewed through three-dimensional ordination space showing correlation between morphological variation and floristic subprovinces in Borneo: *E. clementis* in Lowland Flora in Borneo (B) ▲, *E. clementis* in East Coast Sabah Floristic Subprovince (ECSS) △, *E. clementis* in *Kerangas* Forest (K) ▼, *E. clementis* in Mountain Flora in Borneo (MFB) ▽, *E. clementis* in Seasonally Flooded Riverside (R) ▶, *E. clementis* in Riau Pocket (RP) ▷, *E. clementis* in Seasonal Asiatic-

Australasian Intrusion in North Borneo (SAAI) ◥, *E. cupreus* in B ✳, *E. cupreus* in ECSS ×, *E. cupreus* in RP ⊠, *E. cupreus* typical form in B ⊞, *E. cupreus* typical form in RP ◪, *E. multinervosus* in RP ✶, *E. multinervosus* in B ○, *E. multinervosus* in K ●, *E. multinervosus* in SAAI ⍟ and *E. polyanthus* in RP ◆. A, Combined vegetative and floral dataset (axis 1 and 2). Vectors of the most informative variables (r^2 > 0.6) as determined by PCC are shown in the direction of correlation. Variables are: TwigWidth (variable 1), PetioleLength (2), PetioleWidth (3), DistalPulvinusWidth (4), LeafLength (5), LeafWidth (6), LeafbaseWidestpoint (8), AbaxialMidribWidth (12), SecondaryVeins (13) and InflorescenceWidth (17); B, Combined vegetative and fruit dataset (axis 1 and 3). Vectors of the most informative variables (r^2 > 0.6) as determined by PCC are shown in the direction of correlation. Variables are: TwigWidth (variable 1), PetioleLength (2), PetioleWidth (3), DistalPulvinusWidth (4), LeafLength (5), LeafWidth (6), LeafbaseWidestpoint (8), AbaxialMidribWidth (12), SecondaryVeins (13), AbaxialSecondaryVeinsWidth (15), InfructescenceWidth (36) and FruitStalkWidth (38).

4.4 Discussion

4.4.1 Taxon boundaries within the polystachyus *group*

Taxon boundaries within the *polystachyus* group are confirmed by the morphometric analysis, in that six clusters are retrieved which broadly correspond to the existing (Coode 1996c) species concepts of *E. clementis*, *E. cupreus*, *E. integripetalus*, *E. multinervosus*, *E. polyanthus* and *E. polystachyus*. The methods employed here have the advantage of being statistically reliable and repeatable compared with the traditional intuitive taxonomic method. Moreover, the morphometric analysis constitutes a test of the existing view of the relationships between taxa, allows for the identification of outlying specimens, and more thoroughly and systematically evaluates the taxonomic value of each of the morphological characters scored.

Elaeocarpus clementis is considered to be morphologically highly heterogeneous, with four varieties (var. *borneensis*, var. *clemensiae*, var. *clementis* and var. *kostermansii*) currently recognised. However, the ordination and classification analyses did not separate the varieties into discrete clusters. Geographical discontinuities are not particularly obvious and some populations of the varieties are sympatric, except in Brunei where only var. *clemensiae* is present. Coode (1996c) differentiated the four varieties using the following characters: indumentum on young twigs and inflorescences, patterns of leaf primary venation, shape of bracts, shape of flower buds, length of petals, and number of ovules per loculus. The last two characters are included in the present study (as variables 22 and 34, respectively), but they do not appear to be informative in differentiating the specimens (Table 4.2). The other four characters were excluded due to limitation in interpreting the character states unambiguously (e.g. degree of indumentum) and age bias (e.g. changes due to different growing stages). Nonetheless, this study of mainly quantitative characters does not support the recognition of varieties within *E. clementis* as there is no evidence of morphological discontinuities between varieties.

Elaeocarpus clementis is morphologically highly similar to *E. cupreus*. The former is a Bornean endemic whereas the latter is more widespread throughout Sumatra, Malay Peninsula and Borneo. The two can only be distinguished on flower sexuality - *E. cupreus* has unisexual flowers, whereas *E. clementis* has bisexual flowers (Coode 1996c). Ordination analysis of the veg+fl dataset allows a better delimitation between them, where *E. cupreus* is segregated as a discrete clusters (Fig. 4.4), while in the classification analysis all of the specimens formed a cluster, except one *E. clementis* specimen (WKM 298) fell on the edge of this cluster (Fig. 4.5). Even though FlowerSexuality (variable 35) did not score a high r^2 and p-value in the evaluation analysis (Table 4.2), exclusion of this variable resulted in *E. cupreus* and *E. clementis* being less clearly separated (results not shown).

Two informal forms of *E. cupreus* have been proposed: a typical form and a pubescent form (Coode 1996c). The typical form is endemic to Borneo and is characterised by the midrib of the leaves that slightly raised and sepals that have silky, flattened hairs (adpressed-sericeous) on the outer surface (Coode 1996c). The pubescent form is more widespread, distributed in Sumatra, Malay Peninsula and Borneo and is characterised by the leaf midrib that is strongly raised and sepals with silky, upright hairs (felty or velvety) on the outer surface (Coode 1996c). All of the *E. cupreus* specimens included in this study are of the pubescent form, except BRUN 674, S 27224 and Anderson 8468, which are of the typical form, although only flowering materials are available. The veg+fl (Fig. 4.4 & 4.5) and veg+fr (Fig. 4.6 & 4.7) analyses did not segregate the two forms. Furthermore, results of this study show no evident morphological discontinuity or phytogeographical separation to support taxonomic recognition of the two forms (Fig. 4.8, 4.9 and 4.10). Hence it seems best not to recognise the two forms taxonomically.

Elaeocarpus integripetalus is known only from a single specimen (Teysmann 1088 H.B.) collected in the 19th century in Payakumbuh, Sumatra. Fruit is unknown for this species. Coode (1996c) noted that this taxon is morphologically similar to *E. polystachyus* in having 4-merous flowers and rounded petal apices (the remaining taxa have 5-merous flowers and acute petal apices), but it differs from the latter in having

fewer flowers per inflorescence (15 – 20 flowers) and fewer ovules per loculus (8 ovules). In the ordination (Fig. 4.4) and classification (Fig. 4.5) analyses, *E. integripetalus* clusters next to the *polystachyus* cluster indicating it is similar to but distinct from *E. polystachyus*. Thus the morphometric results support the current taxonomy of this species.

The species status of *E. multinervosus* appears to be well supported by the morphometric analysis, particularly that of the veg+fl dataset. Typically, members of this taxon are vegetatively larger or more robust than those of the rest of the taxa in the *polystachyus* group, showing thicker twigs (variable 1), longer and thicker petioles (2 & 3), larger distal pulvini (4), larger leaf size (5 & 6), larger distance between leaf base and the widest point of the leaf (8), thicker midribs on the abaxial surface (12) and more pairs of secondary veins (13). In the classification analyses of the veg+fl and veg+fr datasets, this taxon shows minor overlap with *E. polyanthus* (Fig. 4.5) and *E. clementis* (Fig. 4.7), respectively. Additionally, the fruit of *E. multinervosus* is not particularly different from the rest of the *polystachyus* group taxa. While this does not invalidate the species status of *E. multinervosus*, it means that several vegetative characters in combination are required to distinguish *E. multinervosus* from the rest of the taxa within the group. Since only *E. cupreus* and *E. multinervosus* have unisexual flowers, the inclusion of this character in the combination could help delimit the taxa, although flower sexuality does not appear to be one of the most informative characters in both of the ordination and classification analyses.

Elaeocarpus polyanthus is a poorly known taxon that has been collected from only two localities in Sarawak (Semengoh Arboretum, Kuching) and the Kelabit Highland. Ordination and classification analyses of the veg+fl data do not resolve the three specimens in a cluster (Fig. 4.4 & 4.5, Supplementary Figure 4.1). In the former, all of the three specimens are nested in the *clementis* cluster, but in the latter, two specimens are nested in the *clementis* cluster and one in the *multinervosus* cluster. Ordination and classification analyses treat the data differently, which may explain different results. For instance, in the evaluation of the most informative variables,

variables 27 (SepalNo, r^2 = 0.626) and 28 (PetalNo, r^2 = 0.626) are among the most informative variables in the PCC analysis, but not in the KW analysis (KW = 36.72 and 36.47, respectively), whereas variables 11 (AcumenLength, KW = 43.10) and 35 (FlowerSexuality, KW = 45.60) are among the most informative variables in the KW analysis, but not in the PCC analysis (r^2 = 0.277 and 0.544, respectively) (Table 4.2). Despite these differences, the results indicate that *E. polyanthus* and *E. clementis* resemble each other in their vegetative and floral characters. In the veg+fr analysis, two of the three *E. polyanthus* specimens formed a discrete cluster (Fig. 4.6 & 4.7, Supplementary Figure 4.2), although all of them are closely related to *E. clementis*, thus conforming to Coode's (1996c) taxonomic concepts - "*Considered by Weibel to be a variety of* E. clementis*, I believe it is as distinct a "species" as any in the* polystachyus *group; robust, persistently hairy with fruits much larger than most for the group and themselves persistently gingery-hairy.*"

Elaeocarpus polystachyus is the most clearly delimited of the examined taxa, showing clear separation in both of the ordination and classification analyses. *Elaeocarpus polystachyus* is endemic to the Malay Peninsula and can be delimited from the rest of the taxa on its wider sepals (variable 21, to 3.5 mm) and petals (23, to 4 mm). Coode used the number of ovules per loculus to delimit *E. polystachyus* from the rest of the *polystachyus* taxa (*E. polystachyus* has 10 – 12 ovules per loculus, the rest typically have up to 9 ovules per loculus) (Table 4.1). This character is included here (variable 34), but the KW analysis suggests it is not among the most informative characters when the whole dataset is analysed (p-value = 37.22, r^2 = 0.533).

4.4.2 Useful characters for delimiting taxa and morphological polymorphism

When used in combination, the 40 morphological variables scored in this study segregated most of the OTUs into clusters that correspond to existing taxa.

The analyses suggest that the vegetative variables are generally more useful for differentiating clusters compared to the reproductive ones – eight out of 12 most informative variables are vegetative (PetioleLength, 2; PetioleWidth, 3;

DistalPulvinusWidth, 4; LeafLength, 5; LeafWidth, 6; LeafbaseWidestpoint, 8; AbaxialMidribWidth, 12) (Table 4.2). The set of most informative variables is highly similar in the PCC and KW results, except for a few minor differences that are mostly in the magnitude of the values. Fruit variables are less useful in defining taxon boundaries, except in the case of *E. clementis* and *E. polyanthus*.

The low percentage (4.2 %) of morphological polymorphism (flower-mery, variable 29 and flower sexuality, 35) observed in this study does not seem to have had a strong influence on the results of morphometric analysis of the *polystachyus* group. Removing the two floral variables did not obviously alter the clustering patterns in the ordination plots and dendrograms.

4.4.3 Correlation between morphological variation and phytogeography

The morphometric assessment in this study reveals that all of the clusters retrieved here are broadly congruent with the floristic provinces, except Cluster 3A (comprising some of the OTUs of *E. cupreus*). For example, Cluster 5A (*E. integripetalus*) is endemic to Sumatra, Clusters 6A, 2B and 8B (comprising OTUs of *E. polystachyus*) is endemic to the Malay Peninsula and the remaining clusters (comprising OTUs of *E. clementis*, *E. cupreus*, *E. multinervosus* and *E. polyanthus*) are endemic to Borneo (Fig. 4.5 & 4.7).

Within the *polystachyus* group, some taxa are sympatric. For instance, within the Malay Peninsula province, *E*. *cupreus* and *E*. *polystachyus* co-exist in the RP subprovince; whereas within the Borneo province, *E*. *clementis* var. *clementis*, *E*. *clementis* var. *clemensiae*, *E*. *cupreus* and *E*. *multinervosus* and *E*. *polyanthus* co-exist in the B and K subprovinces; *E*. *clementis* var. *borneensis* and *E*. *clementis* var. *clementis* co-exist in the SAAI subprovince; and lastly, *E*. *cupreus*, *E*. *clementis* var. *borneensis* and *E. multinervosus* co-exist in the ECSS subprovince. Clustering patterns of the morphometric analysis show that co-occurrence does not seem to result in morphological intermediates; each taxon, at least at the species level, appears to be

maintaining its own morphological characteristics. This suggests that the species taxa studied here are reproductively isolated.

This study did not reveal any obvious correlation between the morphological variation and the detailed ecological phytogeography (based on the combination of floristic composition, floristic influences, elevation and soil types). The clusters retrieved are broadly congruent both with the current taxonomic concepts at the species level (Coode 1996c) as well as congruent with the general geographical isolation at the floristic province level, even for the Bornean taxa, which show higher morphological variation. This suggests that the morphological differences observed are genetically controlled and heritable, rather than ecologically influenced.

The phytogeographical data used in this study were mainly determined from herbarium records, where the broad phytogeographical categories (at both floristic province and subprovince levels) may have limitations in the analysis for determining correlations between morphology and ecology. Detailed ecological factors, abiotic or biotic, that influence morphological variation within the *polystachyus* group may not be accurately represented here. Abiotic (e.g. niche temperature, amount of sunlight and rainfall, and pH of the soil) and biotic (e.g. the presence of pollinators and seed dispersers) factors collected from careful field observations will greatly improve the precision of the analysis in determining correlations between morphology and ecology.

4.4.4 Variation within the polystachyus *group: speculative discussion*

The results of this study do not support the a priori hypothesis of strong ecological influence on the origins and maintenance of morphological variation in the *polystachyus* group. Two alternative hypotheses can be considered: hybridisation, and incomplete recent speciation.

Elaeocarpus is thought to be insect pollinated (Matthews and Endress 2002; Weber 1994) and the seeds are vertebrate dispersed (Crome 1975, 1976; Aggarwal

2002; Maynard 2004; Khan *et al*. 2005). The former mechanism favours short distance genetic exchange, whereas the latter can mediate either short or long distance genetic exchange. However, the possibility of hybridisation either between or within the zones of sympatry of the *polystachyus* group does not seem to be supported by the results of this morphometric analysis. Specimens with morphological polymorphism or intermediate character states are not restricted to the zones of sympatry, for instance SAN 67095 (*E*. *clementis* var. *clemensiae* in Sungai Tongod, Labuk Sugud, Sabah), SFN 35741 (*E. multinervosus* in Sungai Sebunut, Sarawak), SA 1353 (*E. polystachyus* in Bukit Timah, Singapore) and FRI 717 (*E. polystachyus* in Slim Hills, Perak) were collected outside of the zones of sympatry. Finally, high morphological variation within species is mainly found within Borneo, yet the results of the morphological analysis show that each Bornean taxon seems to have broadly maintained its characteristics at the species level despite a high degree of sympatry.

An alternative explanation for the pattern of morphological variation in the *polystachyus* group is recent incomplete speciation. The branch lengths of the phylogeny of *Elaeocarpus* are short and the phylogenetic relationships of many species are unresolved (Chapter 2, Fig. 2.4). Short internal branches may indicate rapid diversification and result in poorly resolved phylogenetic relationships (Richardson *et al*. 2001). Although *E. clementis*, *E. cupreus* and *E. multinervosus* formed a highly supported clade, within the clade relationships are not resolved. Additionally, *E. polystachyus* is not part of this Bornean clade. Nonetheless, the low resolution among some species in the current phylogeny of *Elaeocarpus* (Chapter 2) and the absence of *E. clementis* var. *kostermansii*, *E. integripetalus* and *E. polyanthus* in this phylogeny, means relationships within the *polystachyus* group merit further investigation using more variable and informative DNA markers.

4.5 Conclusions

The morphometric analysis delimited clusters within the *polystachyus* group which are generally congruent with the existing species taxonomy. However, recognition of the existing infraspecific taxa is not supported by this study.

The findings of the morphometric analysis provided preliminary evidence that morphological variation is not correlated with ecology. Each taxon generally maintained its own morphological characteristics at the species level. Future studies, particularly using molecular evidence, are needed to confirm the results of the morphometric analysis.

Chapter 5: General Conclusions

5.1 Overview

The general aims of this thesis were to explore phylogenetic relationships and infer evolutionary history within the *Elaeocarpus* alliance − *Elaeocarpus* L., *Aceratium* DC. and *Sericolea* Schltr. − and within the genus *Elaeocarpus*. The molecular phylogenetic framework provided a strong basis to test current infrageneric systems of *Elaeocarpus* based on morphology, and to test taxon boundaries within the West Malesian *E. polystachyus* Wall. ex Müll.Berol. complex using a morphometric approach. In achieving these aims, this study was the first to:

1. produce a solid foundational multi-locus molecular phylogeny of the *Elaeocarpus* alliance based on three plastid (*trnL-trnF*, *psbA-trnH* and *trnV-ndhC*) and one nuclear (*Xdh*) DNA markers incorporating extensive sampling of representatives from various biogeographic regions;
2. use Next-Generation Sequencing (NGS) technology in building molecular datasets for phylogeny reconstruction in Elaeocarpaceae;
3. trace transformation of seed morphology on the phylogenetic tree of *Elaeocarpus* to identify potential morphological synapomorphies to define molecular clades;
4. estimate divergence times within *Elaeocarpus* based on genetic information, fossil evidence and large scale sampling at the family level;
5. undertake historical biogeography reconstructions to investigate the origins and to explain the present day geographic distribution patterns of *Elaeocarpus*;
6. test taxon boundaries within the West Malesian *Elaeocarpus polystachyus* complex using morphometric analysis and explore correlations between morphological variation and ecology (using phytogeographical affinities at floristic province and subprovince levels as proxies).

5.2 Phylogenetics

Since understanding relationships between and within *Elaeocarpus* is a vital step in inferring evolutionary history and explaining present day distribution patterns, this study used DNA sequence data to produce a comprehensive phylogeny of the *Elaeocarpus* alliance (Chapter 2). Reconstruction of the phylogeny then allowed investigation of morphological character state transformation (Chapter 2), estimation of divergence times, and reconstruction of the historical biogeography (Chapter 3) of *Elaeocarpus*.

Phylogenetic relationships within the *Elaeocarpus* alliance were well resolved. *Elaeocarpus* was shown to be monophyletic and sister to *Aceratium*. Within *Elaeocarpus*, the first solid foundational phylogenetic tree was reconstructed using a multi-locus molecular dataset and an extensive sample of species from various biogeographic regions (Chapter 2). Current infrageneric concepts (based on morphology) within the genus were tested using this molecular evidence. The results show that *E. holopetalus* F.Muell. is sister to the remainder of *Elaeocarpus*. *Elaeocarpus holopetalus* has a combination of atypical morphological characters and is placed in the monotypic group X in the current infrageneric systems (Coode 1984), which is consistent with the molecular results generated in this study. Within the remainder of *Elaeocarpus*, 13 lineages were identified: *E. sedentarius* D.J.Maynard & Crayn, the group VI, group VII, group XI subgroup B, *acronodia*, c*oilopetalum*, sect. *Elaeocarpus*, g*anitrus*, *monocera*, o*bovatus*, *polystachyus*, New Zealand, and New Caledonian groups. All clades were broadly congruent with the current infrageneric systems and supported by combinations of morphological characters. However, group V subgroup D in the current infrageneric systems was shown to be paraphyletic. The representatives of this subgroup included in this study, viz. *E. arnhemicus* F.Muell., *E. obovatus* G.Don, *Elaeocarpus* sp. "Mt Bellenden Ker", *E. dentatus* (J.R. & G.Forst.) Vahl, *E. hookerianus* Raoul and *E. polydactylus* Schltr., segregated into three clades: the *obovatus* (consisting of the first three taxa), New Zealand (the next two taxa) and *ganitrus* (the last taxon) clades.

Reconstruction of seed morphological transformation using parsimony indicated that both the curved embryo and ruminate endosperm are homoplasious within *Elaeocarpus* (Chapter 2). Further investigations are needed to identify other morphological synapomorphies that can be used to define molecular clades for the purpose of constructing a practical, phylogenetic classification.

5.3 Divergence times and historical biogeography

The large-scale phylogeny of the family Elaeocarpaceae allowed more precise estimation of the spatio-temporal evolution of the genus *Elaeocarpus*. Analyses indicated that *Elaeocarpus* diverged from its sister *Aceratium* in the Eocene in Australia. Combinations of geological events and zoochorous dispersal were postulated to have played major roles in subsequent migrations of some groups of *Elaeocarpus* out of Australia (Chapter 3). The evolutionary history of only five of the 12 genera of Elaeocarpaceae has been explicitly examined in studies to date – *Aristotelia* L'Hér*., Tetratheca* Sm. and *Tremandra* R.Br. (Crayn *et al*. 2006), *Sloanea* L. (Niissalo 2011) and *Elaeocarpus* (this study). For the remaining genera, the phylogenetic framework resolved in the present study provides the necessary foundation for such studies.

Understanding the evolutionary history of *Elaeocarpus* will help to explain its current day distribution patterns, particularly for those rare and narrowly distributed taxa. This knowledge combined with information on dispersal and pollination mechanisms can be a very useful tool for conservation planning and management. For example, a comprehensive study on a rare and narrowly endemic species in Australia, *E. sedentarius*, showed that rarity is strongly correlated with dispersal limitation rather than habitat preference (Rossetto *et al*. 2008) and studies on some Australian Wet Tropics *Elaeocarpus* species showed that fruit size influences the potential dispersal distance (Rossetto *et al*. 2009).

The evolution of *Elaeocarpus* advances our knowledge of the biogeography of Asian-Australasian lineages of flowering plants. Vicariance and the mode of dispersal

are two major factors in the events of flora exchange between the Sahul-Sunda shelves. Collision of the Sahul-Sunda shelves in the late Oligocene (Hall 2009) changed the configurations of land and sea, where dispersal corridors (e.g. islands of Wallacea and exposure of some lowlands within Malesia) were formed and overwater distance between the two shelves was greatly reduced (Hall 2002); subsequently, encouraged exchange of biota between the two shelves (Marshall 1983; Hall 2001b, 2012; de Bruyn *et al*. 2014; Crayn *et al*. 2015). The zoochory mode of dispersal has favoured the dispersal of viable propagules across water barriers after the Sahul-Sunda convergence (Crayn *et al*. 2015).

5.4 Morphological variation in species complexes

Morphological variation can be associated with genetic inheritance or ecological factors. Traditional intuitive taxonomic approaches may be limited in their ability to differentiate between these two causes of morphological variation. Population genetics is a powerful approach to determine association between morphological variation and genetic inheritance in species complexes (e.g. Kazan *et al*. 1993 (Fabaceae); Rex *et al*. 2007 (Bromeliaceae); Baba 2013 (Elaeocarpaceae); Puente-Lelievre 2013 (Ericaceae); Griffin and Hoffman 2014 (Poaceae)). Based on field observations, many populations of the West Malesian *polystachyus* group investigated here are likely extinct in Peninsular Malaysia and Borneo due to conversion of natural forests to other land uses. The population genetics approach is unfeasible for this study because extraction of DNA of suitable quality and quantity from herbarium materials of *Elaeocarpus* using current technology has not been successful. The morphometric analysis, which is statistically reliable and repeatable, has been widely applied to resolve species complexes in flowering plants (e.g. Möller *et al*. 2007 (Taxaceae); Kenfack 2011 (Meliaceae); Gibson *et al*. 2012 (Droceraceae); Hurry *et al*. 2012 (Poaceae)), including two species complexes of *Elaeocarpus* in Australasia (Maynard *et al*. 2008; Baba 2013). The results of this study demonstrate that morphometric analysis is a useful approach to resolve species complexes when obtaining molecular data is unfeasible.

5.5 Future directions

The phylogenetic framework established in this study is based on a broad geographical sample of species, but is depauperate in species from Madagascar, continental Asia, Wallacea (Sulawesi, Philippines, Lesser Sunda islands, Moluccas), New Guinea and the Pacific islands. A more thorough understanding of the evolutionary and biogeographical history of the genus will only be achieved when this sampling bias is rectified.

A number of clades within *Elaeocarpus* were robustly resolved, but relationships among the majority of *Elaeocarpus* species are not yet clear due to low sequence divergence. Therefore, new and informative DNA markers are needed to provide better resolution, particularly among the shallow branches of the phylogeny.

The results of this study show that NGS can help overcome the problem of paralogy in phylogenetic studies and suggest that NGS methods deserve to be more extensively utilised in phylogenetic studies at higher taxonomic levels. In the case of paralogy, NGS is more powerful than the traditional Sanger Sequencing technology, because it has the ability to interpret a mixture of templates. However for the methods to be more widely adopted, more efficient and less expensive lab and data cleaning protocols need to be developed.

The assessment of evolutionary transformations of seed morphology within *Elaeocarpus* provides a good evaluation of the utility of this approach for testing the current infrageneric systems. Further investigations should be done for other morphological characters used in the systems to identify synapomorphies and their taxonomic value. This assessment is not restricted to the identification of synapomorphies within the extant taxa, it also offers an opportunity for morphological comparison between the extant and fossil taxa, which could lead to a more accurate fossil calibration points, thus giving better estimation of the molecular divergence times.

Investigation of the correlations between morphological variation, ecology and geography as observed in the *polystachyus* group provides a tangible example of the analytical capability of morphometric analysis. Thus it is applicable as an alternative approach to resolve species complexes. Notwithstanding, the morphometrics approach will be simpler for a geographically restricted complex than a widespread complex. Further insights into the genetic basis of morphological variation in the *polystachyus* group and other species complexes in *Elaeocarpus* may be achieved through the comparative analysis of population genetic data when advances in methods of extraction and analysis of DNA allow for the use of herbarium materials.

References

Aggarwal S (2002) *Elaeocarpus*. In: Van Valkenburg JLCH and Bunyapraphatsara N (eds), *Plant Resources of South-East Asia 12(2)*, *Medicinal and poisonous plants*. pp. 241 – 246. Prosea Foundation, Bogor.

Albert JS and Reis RE (2011) *Historical biogeography of Neotropical freshwater fishes*. University of California Press, Berkeley.

Alfaro ME, Zoller S and Lutzoni F (2003) Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Molecular Biology and Evolution* **20**, 255 – 266.

Ali JR and Aitchison JC (2008) Gondwana to Asia: Plate tectonics, paleogeography and the biological connectivity of the Indian subcontinent from the Middle Jurassic through latest Eocene (166-35 Ma). *Earth-Science Reviews* **88**, 145 – 166.

Altekar G, Dwarkadas S, Huelsenbeck JP and Ronquist F (2004) Parallel metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* **20**, 407 – 415.

Archibald JM, Logsdon JM and Doolittle WF (1999) Recurrent paralogy in the evolution of archaeal chaperonins. *Current Biology* **9**, 1053 – 1056.

Arensburger P, Buckley TR, Simon C, Moulds M and Holsinger KE (2004) Biogeography and phylogeny of the New Zealand cicada genera (Hemiptera: Cicadidae) based on nuclear and mitochondrial DNA data. *Journal of Biogeography* **31**, 557 – 569.

Ashton PS (1992) Plant conservation in the Malesian region. In: Yap SK and Lee SW (eds), *Proceedings of the International Conference on Tropical Biodiversity "In Harmony with Nature"*. pp. 86 – 93. Malayan Nature Society, Kuala Lumpur.

Australian plant census (2015) http://www.anbg.gov.au/chah/apc/. Accessed 15 January 2015.

Baba Y (2013) Evolution, systematics and taxonomy of *Elaeocarpus* (Elaeocarpaceae) in Australasia. *PhD thesis*. James Cook University, Queensland.

Baba Y and Crayn D (2012) *Elaeocarpus hylobroma* (Elaeocarpaceae): a new species endemic to mountain tops in north-east Queensland, Australia. *Kew Bulletin* 67, 743 – 750.

Baldwin BG and Sanderson MJ (1998) Age and rate of diversification of the Hawaiian silverword alliance (Compositae). *Proceedings of the National Academy of Sciences of the United States of America* **95**, 9402 – 9406.

Ballard HE and Sytsma KJ (2000) Evolution and biogeography of the woody Hawaiian violets (*Viola*, Violaceae): Arctic origins, herbaceous ancestry and bird dispersal. *Evolution* **54**, 1521 – 1532.

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

Bayer C and Appel O (1996) Occurrence and taxonomic significance of ruminate endosperm. *The botanical review* **62**, 301 – 310.

Beebe NW and Cooper RD (2002) Distribution and evolution of the *Anopheles punctulatus* group (Diptera: Culicidae) in Australia and Papua New Guinea. *International Journal for Parasitology* **32**, 563 – 574.

Belbin L (1984) FUSE, a FORTRAN 5 program for agglomerative fusion on micro-computers. *Computers & Geosciences* **10**, 361 – 384. doi: 10.1016/0098-3004(84)90039-6

Belbin L (1993) *PATN analysis package. Users' and technical reference manuals*. CSIRO Division of Wildlife and Ecology, Canberra.

Bell CD and Donoghue MJ (2005) Dating the diversification of Dipsacales: comparing models, genes, and evolutionary implications. *American Journal of Botany* **92**, 284 – 296.

Bell CD, Soltis DE and Soltis PS (2005) The age of the angiosperms: a molecular timescale without a clock. *Evolution* **59**, 1245 – 1258.

Bell CD, Soltis DE and Soltis PS (2010) The age and diversification of the angiosperms re-revisited. *American Journal of Botany* **97**, 1296 – 1303. doi:10.3732/ajb.0900346

Bentham G and Hooker JD (1862) *Genera Plantarum ad exemplaria imprimis in herbariis kewensibus servata definita*. Volume 1, part 1, pp. 239 – 240. A. Black, London.

Berbee ML, Carmean DA and Winka K (2000) Ribosomal DNA and resolution of branching order among the ascomycota: how many nucleotides are enough? *Molecular Phylogenetics and Evolution* **17**, 337 – 344.

Bergh NG and Linder HP (2009) Cape diversification and repeated out-ofsouthern Africa dispersal in paper daises (Asteraceae-Gnaphalieae). *Molecular Phylogenetics and Evolution* **51**, 5 – 18.

Biffin E, Hill RS and Lowe AJ (2010) Did kauri (Agathis: Araucariaceae) really survive the Oligocene drowning of New Zealand? *Systematic Biology* **59**, 594 – 602.

BirdLife International (2015) Species factsheet: *Heteralocha acutirostris*. http://www.birdlife.org. Accessed on 21 January 2015.

Bird MI, Taylor D and Hunt C (2005) Paleoenvironments of insular Southeast Asia during the Last Glacial Period: a savanna corridor in Sundaland. *Quaternary Science Reviews* **24**, 2228 – 2242.

Blackith RE and Reyment RA (1971) *Multivariate morphometrics*. Academic Press, London, New York.

Blattner FR (1999) Direct amplification of the entire ITS region from poorly preserved plant material using recombinant PCR. *Biotechnology* **27**, 1180 – 1185.

Bromham L (2008) *Reading the story in DNA, a beginner's guide to molecular evolution*. Oxford University Press Inc., New York.

Brongniart AT and Gris (1861) Description de quelques Eléocarpées de la Nouvelle Calédonie. *Bulletin de la Société Botanique de France* **8**, 198 – 203.

Brown RP and Yang Z (2011) Rate variation and estimation of divergence times using strict and relaxed clocks. *BMC Evolutionary Biology* **11**, 271. doi:10.1186/1471-2148-11-271

Burrows CJ (1995) A macrofossil flora from sediments in a lagoon marginal to Lake Coleridge, Canterbury, New Zealand. *New Zealand Journal of Botany* **33**, 519 – 522.

Burrows CJ (1997) An interglacial macrofossil flora from Schulz Creek, north Westland, New Zealand. *New Zealand Journal of Botany* **35**, 187 – 194.

Byrne M, Steane DA, Joseph L, Yeates DK, Jordan GJ, Crayn D, Aplin K, Cantrill DJ, Cook LG, Crisp MD, Keogh JS, Melville J, Moritz C, Porch N, Sniderman JMK, Sunnucks P and Weston PH (2011) Decline of a biome: evolution, contraction, fragmentation, extinction and invasion of the Australian mesic zone biota. *Journal of Biogeography* **38**, 1635 – 1656.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD and Costello EK (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**, 335 – 336. doi: 10.1038/nmeth0510-335

Chapple DG, Ritchie PA and Daugherty CH (2009) Origin, diversification, and systematics of the New Zealand skink fauna. *Molecular Phylogenetics and Evolution* **52**, 470 – 487.

Christenhusz MJM, Fay MF, Clarkson JJ, Gasson P, Can JM, Jiménez Barrios JB and Chase MW (2010) Petenaeaceae, a new angiosperm family in Huerteales with a

distant relationship to *Gerrardina* (Gerrardinaceae). *Botanical Journal of the Linnean Society* **164**, 16 – 25. doi: 10.1111/j.1095-8339.2010.01074.x

Christophel DC, Harris WK and Syber AK (1987) The Eocene flora of the Anglesea locality, Victoria. *Alcheringa* **11**, 303 – 323.

Church GM (2006) Genomes for all. *Scientific American* **294**, 46 – 54. doi: 10.1038/scientificamerican0106-46

Collinson ME, Manchester SR, Wilde V and Hayes P (2010) Fruit and seed floras from exceptionally preserved biotas in the European Paleogene. *Bulletin of Geosciences* **85**, 155 – 162.

Coode MJE (1978) A conspectus of Elaeocarpaceae in Papuasia. *Brunonia* 1, 131 – 298.

Coode MJE (1980) An obscure new species of *Elaeocarpus* from Central New Guinea. *Kew Bulletin* **35**, 247 – 251.

Coode MJE (1984) *Elaeocarpus* in Australia and New Zealand. *Kew Bulletin* **39**, 509 – 586.

Coode MJE (1985) *Aristotelia* and *Vallea*, closely related in Elaeocarpaceae. *Kew Bulletin* **40**, 479 – 507.

Coode MJE (1987) *Crinodendron*, *Dubouzetia* and *Peripentadenia*, closely related in Elaeocarpaceae. Kew Bulletin **42**, 777 – 814.

Coode MJE (1995) *Elaeocarpus* in the Flora Malesiana area – *E*. *kraengensis* and ten new species from Sulawesi. *Kew Bulletin* **50**, 267 – 294.

Coode MJE (1996a) *Elaeocarpus* for Flora Malesiana – notes, new taxa and combinations in sect. *Elaeocarpus*: 2. *Kew Bulletin* **51**, 83 – 101.

Coode MJE (1996b) *Elaeocarpus* for Flora Malesiana – notes, new taxa and combinations in the *Acronodia* group. *Kew Bulletin* **51**, 267 – 300.

Coode MJE (1996c) *Elaeocarpus* for Flora Malesiana: the '*Polystachyus* group'. *Kew Bulletin* **51**, 649 – 666.

Coode MJE (1998) *Elaeocarpus* for Flora Malesiana – notes, new taxa, names and combinations for Borneo. *Kew Bulletin* **53**, 83 – 128.

Coode MJE (2001a) *Elaeocarpus* in New Guinea – new taxa in the *Debruynii* subgroup of the *Monocera* group. Contribution to the Flora of Mt Jaya, V. *Kew Bulletin* **56**, 449 – 460.

Coode MJE (2001b) *Elaeocarpus* for Flora Malesiana – the *Fissipetalum* group in central Malesia. *Kew Bulletin* **56**, 461 – 463.

Coode MJE (2001c) *Elaeocarpus* for Flora Malesiana – the *E*. *stipularis* complex, *E*. *nitidus* group and *E*. *barbulatus*. *Kew Bulletin* **56**, 513 – 565.

Coode MJE (2001d) *Elaeocarpus* for Flora Malesiana: the *Coilopetalum* group in Sulawesi & Maluku. *Kew Bulletin* **56**, 837 – 874.

Coode MJE (2001e) *Elaeocarpus* for Flora Malesiana – the *Coilopetalum* group in the Lesser Sunda Islands. *Kew Bulletin* **56**, 875 – 883.

Coode MJE (2001f) *Elaeocarpus* for Flora Malesiana – the *Verticellatae* subgroup of the *Monocera* group and a new Philippine species. *Kew Bulletin* **56**, 885 – 901.

Coode MJE (2002) *Elaeocarpus* in New Guinea: *E. sericoloides* (*Fissipetalum* Group), *E. royenii* & *E. multisectus* (Sect. *Elaeocarpus*). Contributions to the Flora of Mt Jaya, IX. *Kew Bulletin* **57**, 925 – 935.

Coode MJE (2003) *Elaeocarpus* for Flora Malesiana: two new taxa in the *Coloides* subgroup of the *Monocera* group from New Guinea. *Kew Bulletin* **58**, 453 – 458.

Coode MJE (2004) Elaeocarpaceae. In: Kubitzki K, KU Kramer, PS Green, JG Rohwer, V Bittrich (eds), *The families and genera of vascular plants, flowering plants dicotyledons: Celastrales, Oxalidales, Rosales, Cornales, Ericales*. Volume 6, pp. 135 – 144. Springer-Verlag.

Coode MJE (2005) *Elaeocarpus* for Flora Malesiana: *E. crenulatus*, *E. myrtoides* and *E. amabilis* from New Guinea, reconsidered. *Kew Bulletin* **60**, 305 – 311.

Coode MJE (2007a) *Elaeocarpus* for Flora Malesiana: *E. forbesii*, a Philippine species with unusual features, and some possible 'associates'. *Kew Bulletin* **62**, 321 – 328.

Coode MJE (2007b) *Elaeocarpus* for Flora Malesiana: new information on *Elaeocarpus* from Borneo and Sulawesi. *Kew Bulletin* **62**, 329 – 332.

Coode MJE (2010) *Elaeocarpus* for Flora Malesiana: new taxa and understanding in the *Ganitrus* group. *Kew Bulletin* **65**, 1 – 45.

Coode MJE and Weibel R (1994) *Elaeocarpus* for Flora Malesiana – notes, new taxa and combinations in sect. *Elaeocarpus*: 1. *Kew Bulletin* **49**, 235–259.

Cook LG and Crisp MD (2005) Not so ancient: the extant crown group of Nothofagus represents a post-Gondwanan radiation. *Proceedings of the Royal Society of London B* **272**, 2535 – 2544.

Cooper A and Cooper RA (1995) The Oligocene bottleneck and New Zealand biota: genetic record of a past environmental crisis. *Proceedings of the Royal Society of London B* **261**, 293 – 302.

Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J and Ward R (2001) Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* **409**, 704 – 707.

Costion C, Edwards W, Ford A, Metcalfe DJ, Cross H, Harrington M, Richardson JE, Hilbert DW, Lowe A and Crayn D (2014) Using phylogenetic diversity to identify ancient rain forest refugia and diversification zones in a biodiversity hotspot. *Diversity and Distributions* Early Online. doi: 10.1111/ddi.12266

Couvreur TLP, Chatrou LW, Sosef MSM and Richardson JE (2008) Molecular phylogenetics reveal multiple tertiary vicariance origins of the African rain forest trees. *BMC Biology* **6**, 54.

Cox CB and Moore PD (1993) *Biogeography: an ecological and evolutionary approach*. Fifth edition. Blackwell Scientific Publications, Oxford.

Crayn DM, Rossetto M and Maynard DJ (2006) Molecular phylogeny and dating reveals on Oligo-Miocene radiation of dry-adapted shrubs (former Tremandraceae) from rainforest tree progenitors (Elaeocarpaceae) in Australia. *American Journal of Botany* **93**, 1328 – 1342.

Crayn DM, Costion C and Harrington MG (2015) The Sahul – Sunda floristic exchange: dated molecular phylogenies document Cenozoic intercontinental dispersal dynamics. *Journal of Biogeography* **42**, 11 – 24.

Crisci JV, Katinas L and Posadas P (2003) *Historical biogeography: an introduction*. Harvard University Press, Cambridge.

Crome FHJ (1975) The ecology of fruit pigeons in tropical northern Queensland. *Australian Wildlife Research* **2**, 155 – 85.

Crome FHJ (1976) Some observations on the biology of the cassowary in northern Queensland. *Emu* **76**, 8 – 14.

de Bruyn M, Stelbrink B, Morley RJ, Hall R, Carvalho GR, Cannon CH, van den Bergh G, Meijaard E, Metcalfe I, Boitani L, Maiorano L, Shoup R and von Rintelen T (2014) Borneo and Indochina are Major Evolutionary Hotspots for Southeast Asian Biodiversity. *Systematic Biology* **63**, 879 – 901.

Dettmann ME and Clifford HT (2000) The fossil record of *Elaeocarpus* L. fruits. *Memoirs of the Queensland Museum* **46**, 461 – 497.

Drummond AJ, Ho SYW, Phillips MJ and Rambaut A (2006) Relaxed phylogenetics and

dating with confidence. *PLoS Biology* **4**, e88. doi: 10.1371/journal.pbio.0040088

Drummond AJ and Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**, 214. doi: 10.1186/1471-2148-7-214

Drummond AJ and Rambaut A (2008) *BEAST version 1.4.8*. http://beast.bio.ed.ac.uk/

Emshwiller E and Doyle JJ (1999) Chloroplast-expressed glutamine synthetase (ncpGS): potential utility for phylogenetic studies with an example from Oxalis (Oxalidaceae). *Molecular Phylogenetics and Evolution* **12**, 310 – 319.

Erdei B and Rakosi L (2009) The Middle Eocene flora of Csordakút (N Hungary). *Geologica Carpathica* **60**, 43 – 57.

Ericson PGP, Christidis L, Cooper A, Irestedt M, Jackson J, Johansson US and Norman JA (2002) A Gondwanan origin of passerine birds supported by DNA sequences of the endemic New Zealand wrens. *Proceedings of the Royal Society of London B* **269**, 235 – 241.

Espeland M and Murienne J (2011) Diversity dynamics in New Caledonia: towards the end of the museum model? *BMC Evolutionary Biology* **11**, 254.

Farris JS, Albert VA, Källersjö M, Lipscomb DL and Kluge AG (1996) Parsimony jackknifing outperforms neighbor-joining. *Cladistics* **12**, 99 – 124.

Feliner GN and Rossello JA (2007) Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Molecular Phylogenetics and Evolution* **44**, 911 – 919.

Felsenstein J (1985) Confidence intervals on phylogenies: An approach using the bootstrap. *Evolution* **39**, 783 – 791.

Felsenstein J (2004) *Inferring phylogenies*. Sinauer Associates, Sunderland, Massachusetts.

Fitch WM (1971) Toward defining the course of evolution: minimum change for a specific tree topology. *Systematic Zoology* **20**, 406 – 416.

Fu L, Niu B, Zhu Z, Wu S and Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150 – 3152. doi: 10.1093/bioinformatics/bts565

Gagnepain F, Humbert H and Lecomte H (1907 – 1912) *Flore générale de I'Indo-Chine*. Volume 1. Masson, Paris.

Gamerman D and Lopes HF (2006) *Markov Chain Monte Carlo: stochastic simulation for Bayesian inference*. Second edition. Chapman and Hall/CRC, Boca Raton, Florida.

García-Sandoval R (2014) Why some clades have low bootstrap frequencies and high Bayesian posterior probabilities. *Israel Journal of Ecology and Evolution* **60**, 41 – 44. doi: 10.1080/15659801.2014.937900

Gascuel O and Steel M (2006) Neighbor-joining revealed. *Molecular Biology and Evolution* **23**, 1997 – 2000.

Gemmill CEC, Allan GJ, Wagner WL and Zimmer EA (2002) Evolution of insular Pacific Pittosporum (Pittosporaceae): origin of the Hawaiian radiation. *Molecular Phylogenetics and Evolution* **22**, 31 – 42.

Gibson R, Conn BJ and Bruhl JJ (2012) Morphological evaluation of the *Drosera peltata* complex (Droseraceae). *Australian Systematic Botany* **25**, 49 – 80.

Gielly L, Yuan YM, Kupfer P and Taberlet P (1996) Phylogenetic use of noncoding regions in the genus *Gentiana* L. Chloroplast trnL (UAA) intron versus nuclear ribosomal internal transcribed spacer sequences. *Molecular Phylogenetics and Evolution* **5**, 460 – 466.

Givnish TJ, Barfuss MHJ, Van Ee B, Riina R, Schulte K, Horres R, Gonsiska PA, Jabaily RS, Crayn DM, Smith JAC, Winter K, Brown GK, Evans TM, Holstm BK, Luther H, Till W, Zizka G, Berry PE and Sytsma KJ (2014) Adaptive radiation, correlated and contingent evolution, and net species diversification in Bromeliaceae. *Molecular Phylogenetics and Evolution* **71**, 55 – 78.

Glassburn TA (1995) A new palaeontological technique describing temporal shape variation in Miocene bivalves. *Palaeontology* **38**, 133 – 151.

Global Biodiversity Information Facility (2014) http://www.gbif.org. Accessed 23 December 2014.

Goebel K (1993) *Organographic der Pflanzen*. Third edition, volume 3. Fischer, Jena.

Goloboff PA, Farris JS, Kallersjo M, Oxelman B, Ramirez MJ and Szumik CA (2003) Improvements to resampling measures of group support. *Cladistics* **19**, 324 – 332.

Gower JC (1971) A general coefficient of similarity and some of its properties. *Biometrics* **27**, 857 – 871. doi: 10.2307/2528823

Gower JC and Ross RJS (1969) Minimum spanning trees and single linkage cluster analysis. *Applied Statistics* **18,** 54 – 64.

Gradstein FM and Ogg JG (2004) Geologic time scale 2004 – why, how, and where next! *Lethaia* **37**, 175 – 181. doi: 10.1080/00241160410006483

Grandcolas P, Murienne J, Robillard T, Desutter-Grandcolas L, Jourdan H, Guilbert E

and Deharveng L (2008) New Caledonia: a very old Darwinian island? *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**, 3309 – 3317.

Graur D (2003) Single-base mutation. *Nature encyclopedia of the human genome*. pp. 287 – 290. Macmillan Publishers Ltd, Nature Publishing Group.

Graur D and Li WH (2000) *Fundamentals of molecular evolution*. Second edition. Sinauer Associates Inc., Sunderland, Massachusetts.

Griffin PC, Robin C and Hoffman AA (2011) A next-generation sequencing method for overcoming the multiple gene copy problem in polyploid phylogenetics, applied to *Poa* grasses. *BMC Biology* **9**, 1 – 19.

Griffin PC and Hoffman AA (2014) Limited genetic divergence among Australian alpine *Poa* tussock grasses coupled with regional structuring points to ongoing gene flow and taxonomic challenges. *Annals of Botany* **113**, 953 – 965. doi:10.1093/aob/mcu017

Grudinski M, Wanntorp L, Pannell CM and Muellner-Riehl AN (2014) West to east dispersal in a widespread animal-dispersed woody angiosperm genus (*Aglaia*, Meliaceae) across the Indo-Australian Archipelago. *Journal of Biogeography* **41**, 1149 – 1159.

Hably L. (2007) New European records and first evidence of epidermal structures of *Sloanea* L. from the Italian Oligocene – revised type and original material of principi 1916. *Rivista Italiana di Paleontologia e Stratigrafia* **113**, 449 – 457.

Hall N (2007) Advanced sequencing technologies and their wider impact in microbiology. The *Journal of Experimental Biology* **209**, 1518 – 1525. doi: 10.1242/jeb.001370

Hall R (1998) The plate tectonics of Cenozoic SE Asia and the distribution of land and sea. In: Hall R and Holloway JD (eds), *Biogeography and geological evolution of SE Asia*. pp. 99 – 131. Backhuys Publishers, Leiden.

Hall R (2001a) Cenozoic reconstructions of SE Asia and the SW Pacific: changing patterns of land and sea. In: Metcalfe I, Smith JMB, Morwood M and Davidson ID (eds), *Faunal and floral migrations and evolution in SE Asia-Asutralasia*. pp. 35 – 56. Swets and Zeitlinger Publisher, Lisse.

Hall R (2001b) Extension during late Neogene collision in east Indonesia and New Guinea. *Journal of the Virtual Explorer* **4**, 1 – 13.

Hall R (2002) Cenozoic geological and plate tectonic evolution of Southeast Asia and the Southwest Pacific: computer-based reconstructions, model and animations. *Journal of Asian Earth Sciences* **20**, 353 – 434.

Hall R (2009) Southeast Asia's changing palaeogeography. *Blumea* **54**, 148 – 161. doi:10.3767/000651909X475941

Hall R (2012) Sundaland and Wallacea: geology, plate tectonics and palaeogeography. In: Gower D, Johnson K, Richardson JE, Rosen B, Rüber L and Williams SE (eds), *Biotic Evolution and Environmental Change in Southeast Asia*. Cambridge University Press, Cambridge.

Hall R and Nichols G (2002) Cenozoic sedimentation and tectonics in Borneo: climatic influences on orogenesis. In: Jones SJ and Frostick L (eds), *Sediment flux to basins: causes, controls and consequences*. Special publications 191, pp. 5 – 22. Geological Society, London.

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.

Hao G, Yuan Y-M, Hu C-M, Ge X-J and Zhao N-X (2004) Molecular phylogeny of *Lysimachia* (Myrsinaceae) based on chloroplast trnL-F and nuclear ribosomal ITS sequences. *Molecular Phylogenetics and Evolution* **31**, 323 – 339.

Harbaugh DT, Wagner WL, Allan GJ and Zimmer EA (2009) The Hawaiian Archipelago is a stepping stone for dispersal in the Pacific: an example from the plant genus Melicope (Rutaceae). *Journal of Biogeography* **36**, 230 – 241. doi:10.1111/j.1365-2699.2008.02008.x

Harpke D and Peterson A (2008) 5.8S motifs for the identification of pseudogenic ITS regions. *Botany* **86**, 300 – 305.

Heaney LR (1991) A synopsis of climatic and vegetational change in Southeast Asia. *The International Journal of Climate Change* **19**, 53 – 61.

Heibl C and Renner SS (2012) Distribution Models and a Dated Phylogeny for Chilean Oxalis Species Reveal Occupation of New Habitats by Different Lineages, not Rapid Adaptive Radiation. *Systematic Biology* **61**, 823 – 834.

Herbert G and Easteal S (1996) Relative rates of nuclear DNA evolution in human and Old World monkey lineages. *Molecular Biology and Evolution* **13**, 1054 – 1057.

Hesberg C, Hänsch R, Mendel RR and Bittner F (2004) Tandem orientation of duplicated xanthine dehydrogenase genes from Arabidopsis thaliana: differential gene expression and enzyme activities. *The Journal of Biological Chemistry* **279**, 13547 – 13554. doi: 10.1074/jbc.M312929200.

Hill KC and Hall R (2003) Mesozoic-Cenozoic evolution of Australia's New Guinea margin in a west Pacific context. In: Hillis RR and Müller RD (eds), *Evolution and dynamics of the Australian Plate*. pp. 259 – 283. Geological Society of Australia

Special Publication 22 and Geological Society of America Special Paper 372, Sydney.

Hill RS (2004) Origins of the southeastern Australian vegetation. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences* **359**, 1537 – 1549.

Hillis DM (1987) Molecular versus morphological approaches to systematics. *Annual Review of Ecology and Systematics* **18**, 23 – 42.

Hillis DM, Heath TA and St John K (2005) Analysis and visualisation of tree space. *Systematic Biology* **54**, 471 – 482.

Ho SYM (2007) Calibrating molecular estimates of substition rates and divergence times in birds. *Journal of Avian Biology* **38**, 409 – 414. doi: 10.1111/j.0908-8857.2007.04168.x

Ho SY, Saarma U, Barnett R, Haile J and Shapiro B (2008) The effect of inappropriate calibration: three case studies in molecular ecology. *PLoS One* **3**, e1615. doi: 10.1371/journal.pone.0001615

Hodges SA and Arnold ML (1994) Columbines: a geographically widespread species flock. *Proceedings of the National Academy of Sciences* **91**, 5129 – 5132.

Howarth DG, Gustafsson MHG, Baum DA and Motley TJ (2003) Phylogenetics of the genus *Scaevola* (Goodeniaceae): Implications for the dispersal patterns across the Pacific Basin and colonization of the Hawaiian Islands. *American Journal of Botany* **90**, 915 – 923.

Huelsenbeck JP, Larget B, Miller RE and Ronquist F (2002) Potential applications and pitfalls of bayesian inference of phylogeny. *Systematic Biology* **51**, 673 – 688.

Huelsenbeck JP and Rannala B (2004) Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Systematic Biology* **53**, 904 – 913.

Hug LA and Roger AJ (2007) The impact of fossils and taxon sampling on ancient molecular dating analyses. *Molecular Biology and Evolution* **24**, 1889 – 1897.

Hurry CR, Walsh NG and Murphy DJ (2012) A taxonomic review of *Triodia bunicola* and *T. scariosa* (Poaceae: Chloridoideae), based on morphological and molecular data. *Australian Systematic Botany* **25**, 304 – 312.

Illumina Inc. (2013) *16S metagenomic sequencing library preparation guide*. (support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf)

Inger RF and Voris HK (2001) The biogeographical relations of the frogs and snakes of Sundaland. *Journal of Biogeography* **28**, 863 – 891.

Janssens SB, Knox EB, Huysmans S, Smets EF and Merckx VSFT (2009) Rapid radiation of *Impatiens* (Balsaminaceae) during Pliocene and Pleistocene: result of a global climate change. *Molecular Phylogenetics and Evolution* **52**, 806 – 824.

Johnson DJ (2004) *The geology of Australia*. Cambridge University Press, Cambridge.

Katoh K and Frith MC (2012) Adding unaligned sequences into an existing alignment using MAFFT and LAST. *Bioinformatics* **28**, 3144 – 3146. doi: 10.1093/bioinformatics/bts578

Katoh K and Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772 – 780. doi: 10.1093/molbev/mst010

Kazan K, Manners JM and Cameron DF (1993) Genetic relationships and variation in the *Stylosanthes guianensis* species complex assessed by random amplified polymorphic DNA. *Genome* **36**, 43 – 49.

Kenfack D (2011) Resurrection in *Carapa* (Meliaceae): a reassessment of morphological variation and species boundaries using multivariate methods in a phylogenetic context. *Botanical Journal of the Linnean Society* **165**, 186 – 221.

Khan ML, Putul B and Tripathi RS (2005) Effects of forest disturbance on fruit set, seed dispersal and predation of Rudraksh (*Elaeocarpus ganitrus* Roxb.) in northeast India. *Current Science* **88**, 133 – 142.

Kim H-G, Keeley SC, Vroom PS and Jansen RK (1998) Molecular evidence for an African origin of the Hawaiian endemic *Hesperomannia (*Asteraceae). *Proceedings of the National Academy of Sciences* **95**, 15440 – 15445.

Kim K-J and Jansen RK (1995) *ndhF* sequence evolution and the major clades in the sunflower family. *Proceedings of the National Academy of Sciences* **92**, 10379 – 10383.

Kim S, Park C-W, Kim Y-D and Suh Y (2001) Phylogenetic relationships in family Magnoliaceae inferred from *ndhF* sequences. *American Journal of Botany* **88**, 717 – 728.

Kishino H, Thorne JL and Bruno WJ (2001) Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Molecular Biology and Evolution* **18**, 352 – 361.

Kruskal WH and Wallis WA (1952) Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association* **47**, 583 – 621.

Kumar S and Subramanian S (2002) Mutation rates in mammalian genomes.

Proceedings of the National Academy of Sciences **99**, 803 – 808.

Kürschner WM, Kvaèek Z and Dilcher DL (2008) The impact of Miocene atmospheric carbon dioxide fluctuations on climate and the evolution of terrestrial ecosystems. *Proceedings of the National Academy of Sciences USA* **105**, 449 – 453.

Kvaček Z (2002) Late Eocene landscape, ecosystems and climate in northern Bohemia with particular reference to the locality of Kučlín near Bílina. *Bulletin of the Czech Geological Survey* **77**, 217 – 236.

Kvaček Z, Hably L and Manchester SR (2001) *Sloanea* (Elaeocarpaceae) fruits and foliage from the Early Oligocene of Hungary and Slovenia. *Palaeontographica Abteilung B Palaeophytologie* **259**, 113 – 124.

Ladiges PY and Cantrill D (2007) New Caledonia – Australian connections: biogeographic patterns and geology. *Australian Systematic Botany* **20**, 383 – 389.

Larget B and Simon DL (1999) Markov Chain Monte Carlo Algorithms for the Bayesian Analysis of Phylogenetic Trees. *Molecular Phylogenetics and Evolution* **16**, 750 – 759.

Lee DW (1991) Ultrastructureal basis and function of iridescent blue colour of fruits in *Elaeocarpus*. *Nature* **349**, 260 – 263.

Leiberman BS (2005) Geobiology and paleobiogeography: tracking the conevolution of the earth and its biota. *Palaeogeography, Palaeoclimatology, Palaeoecology* **219**, 23 – 33.

Life Technologies Corporation (2014) http://www.lifetechnologies.com/. Accessed 11 June 2014.

Lihová J, Shimizu K and Marhold K (2006) Allopolyploid origin of Cardamine asarifolia (Brassicaceae): incongruence between plastid and nuclear ribosomal DNA sequences solved by a single-copy nuclear gene. *Molecular Phylogenetics and Evolution* **39**, 759 – 786.

Lindqvist C and Albert VA (2002) Origin of the Hawaiian endemic mints within North American *Stachys* (Lamiaceae). *American Journal of Botany* **89**, 1709 – 1724.

Lohman DJ, de Bruyn M, Page T, von Rintelen K, Hall R, Ng PKL, Shih H-T, Carvalho GR and von Rintelen T (2011) Biogeography of the Indo-Australian Archipelago. *Annual Review of Ecology and Systematics* **42**, 205 – 226.

Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M and Usadel B (2012) RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Research* **40**, W622 – W627. doi: 10.1093/nar/gks540

Maddison WP and Maddison DR (2011) Mesquite: a modular system for evolutionary analysis. version 2.75. http://mesquiteproject.org. (725 modules).

Magallón S, Gómez-Acevedo S, Sánchez-Reyes LL and Hernández-Hernández T (2015) A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist* (Early View 23 Jan 2015). doi: 10.1111/nph.13264

Manchester SR and Kvaček Z (2009) Fruits of *Sloanea* (Elaeocarpaceae) in the Paleogene of North America and Greenland. *International Journal of Plant Sciences* **170**, 941 – 950.

Marcus LF (1990) Traditional morphometrics. In: Rohlf FJ and Bookstein FL (eds), *Proceedings of the Michigan morphometrics workshop*. Special publication number 2, pp. 77 – 122. The University of Michigan Museum of Zoology, Ann Arbor.

Marshall AG (1983) Bats, flowers and fruit: evolutionary relationships in the Old World. *Biological Journal of the Linnean Society* **20**, 115 – 135.

Marshall CR (1990) The Fossil Record and Estimating Divergence Times between Lineages: Maximum Divergence Times and the Importance of Reliable Phylogenies. *Journal of Molecular Evolution* **30**, 400 – 408.

Masters MT (1874) Tiliaceae. In: Hooker JD (ed), *Flora of British India*. Volume 1, pp. 400 – 409. L. Reeve, London.

Maynard D (2004) A molecular phylogeny for Australian *Elaeocarpus* (Elaeocarpaceae) and the affinities of a putative new taxon. *Honours thesis*. University of New South Wales, New South Wales.

Maynard D, Crayn D, Rossetto M, Kooyman R and Coode M (2008) *Elaeocarpus sedentarius* sp. nov. (Elaeocarpaceae) – morphometric analysis of a new, rare species from eastern Australia. *Australian Systematic Botany* **21**, 192 – 200.

McDowall RM (2008) Process and pattern in the biogeography of New Zealand – a global microcosm? *Journal of Biogeography* **35**, 197 – 212.

McGlone MS (2005) Goodbye Gondwana. *Journal of Biogeography* **32**, 739 – 740.

McLoughlin S (2001) The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. *Australian Journal of Botany* **49**, 271 – 300.

Mesquite Project Team (1999 – 2014) *Mesquite: A modular system for evolutionary analysis*. http://mesquiteproject.wikispaces.com/. Accessed 20 June 2014.

Möller M, Gao L-M, Mill RR, Li D-Z, Hollingsworth ML and Gibby M (2007) Morphometric analysis of the *Taxus wallichiana* complex (Taxaceae) based on

herbarium material. *Botanical Journal of the Linnean Society* **155**, 307 – 335.

Morat P, Jaffré T, Tronchet F, Munzinger J, Pillon Y, Veillon J-M and Chalopin M (2012) The taxonomic reference base 'FLORICAL' and characteristics of the native vascular flora of New Caledonia. *Adansonia* **34**, 179 – 221.

Morley RJ (2002) Tertiary vegetational history of Southeast Asia, with emphasis on the biogeographical relationships with Australia. In: Kershaw P, David B, Tapper N, Penny D, Brown J (eds), *Bridging Wallace's line: the environmental and cultural history and dynamics of the SE-Asian-Australian region*. pp. 49 – 60. Reiskirchen, Catena Verlag.

Morley RJ (2003) Interplate dispersal paths for megathermal angiosperms. *Perspectives in Plant Ecology, Evolution and Systematics* **6**, 5 – 20.

Morley RJ (2007) Cretaceous and Tertiary climate change and the past distribution of megathermal rainforests. In: Flenley JR, Bush MB (eds), *Tropical rainforest responses to climatic changes*. pp. 1 – 31. Springer, Berlin.

Morrone JJ and Crisci JV (1995) Historical biogeography: introduction of methods. *Annual Review of Ecology and Systematics* **26**, 373 – 401.

Morton CM (2011) Newly sequenced nuclear gene (Xdh) for inferring angiosperm phylogeny. *Annals of the Missouri Botanical Gardens* **98**, 63 – 89. doi: 10.3417/2008107

Müller KF (2003) PRAP - computation of Bremer support for large data sets. *Molecular Phylogenetics and Evolution* **31**, 780 – 782. doi: 10.1016/j.ympev.2003.12.006

Murienne J (2009) Testing biodiversity hypotheses in New Caledonia using phylogenetics. *Journal of Biogeography* **36**, 1433 – 1434.

Neall VE and Trewick SA (2008) The age and origin of the Pacific islands: a geological overview. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**, 3293 – 3308.

Nelson G (1975) Reviews: biogeography, the vicariance paradigm, and continental drift. *Systematic Zoology* **24**, 489 – 504.

Nelson GJ and Ladiges PY (1996) Paralogy in cladistic biogeography and analysis of paralogy-free subtrees. *American Museum Novitates* **3167**, 1 – 58.

Nepokroeff M, Sytsma KJ, Wagner WL and Zimmer EA (2003) Reconstructing ancestral patterns of colonization and dispersal in the Hawaiian understory tree genus *Psychotria* (Rubiaceae): a comparison of parsimony and likelihood approaches. *Systematic Biology* **52**, 820 – 838.

Niissalo MA (2011) Biogeography and phylogeny of *Sloanea*. *MSc thesis*. The University of Edinburgh, Edinburgh.

Nixon KC (1999) The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics* **15**, 407 – 414.

Nylander JAA (2008) MrModeltest 2.3. https://www.abc.se/~nylander/mrmodeltest2/mrmodeltest2.html. Accessed 20 June 2012.

Nylander JAA, Olsson U, Alström P and Sanmartín I (2008) Accounting for phylogenetic uncertainty in biogeography: a Bayesian approach to Dispersal-Vicariance Analysis of the thrushes (Aves: Turdus). *Systematic Biology* **57**, 257 – 268.

Pacific Biosciences of California Inc (2003) http://www.pacificbiosciences.com/. Accessed 11 June 2014.

Pereira SL, Johnson KP, Clayton DH and Baker AJ (2007) Mitochondrial and nuclear DNA sequences support a Cretaceous origin of Columbiformes and a dispersal-driven radiation in the Paleogene. *Systematic Biology* 56, 656 – 672.

Periasamy K (1962) The ruminate endosperm. Development and types of rumination. In: *Plant embryology, a symposium*. pp. 62 – 74. C.S.I.R., New Delhi.

Petriella B (1972) Estudio de maderas petrificadas del Terciario Inferior del area Central de Chubut (Cerro Bororó). *Revista del Museo de La Plata, Sección Paleontología* **6**, 159 – 254.

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

Pillon Y, Hopkins HCF, Munzinger J and Chase MW (2009) A molecular and morphological survey of generic limits of *Acsmithia* and *Spiraeanthemum* (Cunoniaceae). *Systematic Botany* **34**, 141 – 148.

Pole M (1994) The New Zealand flora - entirely long-distance dispersal? *Journal of Biogeography* **21**, 625 – 635.

Posada D (2006) ModelTest Server: a web-based tool for the statistical selection of models of nucleotide substitution online. *Nucleic Acids Research* **34**, W700 – W703 (Web Server Issue). doi:10.1093/nar/gkl042

Posada D and Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology* **53**, 793 – 808.

Posada D and Crandell KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817 – 818.

Poulin E, Palma AT and Féral JP (2002) Evolutionary versus ecological success in Antarctic benthic invertebrates. *Trends in Ecology and Evolution* **17**, 218 – 222.

Prakash U and Dayal R (1964) Fossil woods resembling *Elaeocarpus* and *Leea* from the Deccan Intertrappean beds of Mahurzaei near Nagpur. *Palaeobotanist* **12**, 121 – 127.

Price TD, Qvarnström A and Irwin DE (2003) The role of phenotypic plasticity in driving genetic evolution. *Proceedings of the Royal Society B: Biological Sciences* **270 (1523)**, 1433 – 40. doi:10.1098/rspb.2003.2372

Puente-Lelievre C (2013) Systematics and biogeography of the Styphelieae (Epacridoideae, Ericaceae). *PhD thesis*. James Cook University, Queensland.

Puente-Lelievre C, Harrington MG, Brown EA, Kuzmina M and Crayn DM (2013) Cenozoic extinction and recolonization in the New Zealand flora: the case of the fleshy-fruited epacrids (Styphelieae, Styphelioideae, Ericaceae). *Molecular Phylogenetics and Evolution* **66**, 203 – 214.

Rambaut A (2014) FigTree version 1.4.2: Tree figure drawing tool. http://tree.bio.ed.ac.uk/software/figtree/. Accessed 20 July 2014.

Raven PH and DI Axelrod (1974) Angiosperm biogeography and past continental movements. *Annals of the Missouri Botanical Garden* **61**, 539 – 673.

Ree RH, Moore BR, Webb CO and Donoghue MJ (2005) A Likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution* **59**, 2299 – 2311.

Ree RH and Smith SA (2008) Maximum Likelihood inference of geographic range evolution by dispersal, local extinction, and Cladogenesis. *Journal of Systematic Palaeontology* **57**, 4 – 14.

Reed TE, Waples RS, Schindler DE, Hard JJ and Kinnison MT (2010) Phenotypic plasticity and population viability: the importance of environmental predictability. *Proceedings of the Royal Society B* **277**, 3391 – 3400. doi: 10.1098/rspb.2010.0771

Rex M, Patzolt K, Schulte K, Zizka G, Vásquez R, Ibisch PL and Weising K (2007) AFLP analysis of genetic relationships in the genus *Fosterella* L. B. Smith (Pitcairnioideae, Bromeliaceae). *Genome* **50**, 90 – 105.

Reyment RA (1991) *Multidimensional Paleobiology*. Pergamon Press, New York.

Richardson JE, Pennington RT, Pennington TD and Hollingsworth PM (2001) Rapid diversification of a species-rich genus of neotropical rain forest trees. *Science* **293**, 2242 – 2245.

Ridley HN (Reprint 1990) *The dispersal of plants throughout the world*. Bishen Singh Mahendra Pal Singh, Dehra Dun, India.

Roche Diagnostics Corporation (2014) www.454.com. Accessed 11 June 2014.

Rock JF (1913) *The indigenous trees of the Hawaiian Islands*. Patronage, Honolulu, Hawaii.

Rohlf JF and Marcus LF (1993) A revolution in morphometrics. *Tree* **8**, 129 – 132.

Ronquist F, Huelsenbeck J and Teslenko M (2011) *Draft MrBayes version 3.2 manual: tutorials and model summaries*. November 15. URL: http://mrbayes.sourceforge.net/manual.php

Rossetto M, Grossb CL, Jones R and Hunter J (2004) The impact of clonality on an endangered tree (*Elaeocarpus williamsianus*) in a fragmented rainforest. *Biological Conservation* **117**, 33 – 39.

Rossetto M, Crayn D, Ford A, Ridgeway P and Rymer P (2007) The comparative study of range-wide genetic structure across related, co-distributed rainforest trees reveals contrasting evolutionary histories. *Australian Journal of Botany* **55**, 416 – 424.

Rossetto M, Kooyman R, Sherwin W and Jones R (2008) dispersal limitations, rather than bottlenecks or habitat specificity, can restrict the distribution of rare and endemic rainforest trees. *American Journal of Botany* **95**, 321 – 329.

Rossetto M, Crayn D, Ford A, Mellick R and Sommerville K (2009) The influence of environment and life-history traits on the distribution of genes and individuals: a comparative study of 11 rainforest trees. *Molecular Ecology* **18**, 1422 – 1438. doi: 10.1111/j.1365-294X.2009.04111.x

Rozefelds AC and Christophel DC (1996a) *Elaeocarpus* (Elaeocarpaceae) endocarps from the Early to Middle Miocene Yallourn formation of Eastern Australia. *Muelleria* **9**, 229 – 237.

Rozefelds AC and Christophel DC (1996b) *Elaeocarpus* (Elaeocarpaceae) endocarps from the Oligo-Miocene of Eastern Australia. *Papers and Proceedings of the Royal Society of Tasmania* **130**, 41 – 48.

Rozefelds AC and Christophel DC (2002) Cenozoic *Elaeocarpus* (Elaeocarpaceae) fruits from Australia. *Alcheringa* **26**, 261 – 274. doi: 10.1080/03115510208619256

Sachse M (2005) A remarkable fossiliferous mass flow deposit in the Eocene Eckfeld Maar (Germany) – sedimentological, taphonomical, and palaeoecological considerations. *Facies* **51**, 173 – 184.

Sanderson MJ (1997) A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution* **14**, 1218 – 1231.

Sanderson MJ (2002) Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution* **19**, 101 – 109.

Sanderson MJ (2003) r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301 – 302.

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

Sang T, Crawford DJ and Stuessy TF (1997) Chloroplast DNA phylogeny, reticulate evolution and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany* **84**, 1120 – 1136.

Sauquet H, Ho SYW, Gandolfo MA, Jordan GJ, Wilf P, Cantrill DJ, Bayly MJ, Bromham L, Brown GK, Carpenter RJ, Lee DM, Murphy DJ, Sniderman JMK and Udovicic F (2012) Testing the Impact of Calibration on Molecular Divergence Times Using a Fossil-Rich Group: The Case of *Nothofagus* (Fagales). *Systematic Biology* **61**, 289 – 313.

Savolainen V, Fay MF, Albach DC, Backlund A, van der Bank M, Cameron KM, Johnson SA, Llédo MD, Pintaud JC, Powell M, Sheahan MC, Soltis DE, Soltis PS, Weston P, Whitten WM, Wurdack KJ and Chase MW (2000) Phylogeny of the eudicots: a nearly complete familial analysis based on *rbcL* gene sequences. *Kew Bulletin* **55**, 257 – 309.

Schlechter R (1916) Die Elaeocarpaceen Papuasiens. In: Lauterbach C (ed), Beitrage zur Flora von Papuasien V. *Botanische Jahrbucher* **54**, 92 – 155.

Schwery O, Onstein RE, Bouchenak-Khelladi Y, Xing Y, Carter RJ and Linder HP (2014) As old as the mountains: the radiations of the Ericaceae. *New Phytologist*, online in advance of print. Accessed 15 January 2015.

Seelanan T, Schnabel A and Wendel JF (1997) Congruence and consensus in the cotton tribe (Malvaceae). *Systematic Biology* **22**, 259 – 290.

Shaw J, Lickey EB, Schilling EE and Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in Angiosperms: the tortoise and the hare III. *American Journal of Botany* **94**, 275 – 288.

Shepherd LD and Lambert DM (2007) The relationships and origins of the New Zealand wattlebirds (Passeriformes, Callaeatidae) from DNA sequence analysis. *Molecular*

Phylogenetics and Evolution **43**, 480 – 492.

Shevenell AE, Kennett JP and Lea DW (2004) Middle Miocene southern ocean cooling and Antarctic cryosphere expansion. *Science* **305**, 1766 – 1770.

Simmons MP and Ochoterena H (2000) Gaps as characters in sequence-based phylogenetics analyses. *Systematic Biology* **49**, 369 – 381.

Simpson MG (2006) *Plant systematics*. Elsevier Academic Press, Canada.

Small R, Cronn R and Wendel J (2004) Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* **17**, 145 – 170.

Smith AC (1944) Studies of Papuasian plants, VI. *Journal of the Arnold Arboretum* **25**, 104 – 121, 222 – 298.

Smith AC (1953) Studies of Pacific Island plants, XV – the genus *Elaeocarpus* in the New Hebrides, Fiji, Samoa, and Tonga. *Contributions from the United States National Herbarium* **30**, 523 – 575.

Sneath PHA and Sokal RR (1973) *Numerical taxonomy – the principles and practice of numerical classification*. W.H. Freeman, San Francisco.

Stevens PF (2013) Angiosperm Phylogeny Website, version 13. http://www.mobot.org/MOBOT/research/APweb/. Accessed 20 December 2014.

Strimmer KS (1997) Maximum likelihood methods in molecular phylogenetics. *PhD thesis*. Ludwig Maximilian University, Munich.

Swenson U, Nylinder S and Munzinger J (2013) Towards a natural classification of Sapotaceae subfamily Chrysophylloideae in Oceania and Southeast Asia based on nuclear sequence data. *Taxon* **62**, 746 – 770.

Swenson U, Nylinder S and Munzinger J (2014) Sapotaceae biogeography supports New Caledonia being an old Darwinian island. *Journal of Biogeography* **41**, 797 – 809.

Swofford DL (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Sun Y, Skinner DZ, Liang GH and Hulbert SH (1994) Phylogenetic analysis of sorghum and related taxa using internal transcribed spacers of nuclear ribosomal DNA. *Theoretical and Applied Genetics* **89**, 26 – 32.

Taberlet P, Gielly L, Pautou G and Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* **17**, 1105 – 1109.

Tang Y (1992) Notes on the genus *Elaeocarpus* Linn. from China. *Acta Phytotaxonomica Sinica* **30**, 385 – 404.

Tang Y and Phengklai C (2007) Elaeocarpaceae. In: Wu Z, Raven PH and Hong D (eds), *Flora of China*. Volume 12, pp. 223 – 239. Online volume http://www.efloras.org/. Accessed 20 June 2012.

Tate J and Simpson BB (2003) Paraphyly of Tarasa (Malvaceae) and diverse origins of the polyploid species. *Systematic Botany* **28**, 723 – 737.

Thomas DC, Hughes M, Phutthai T, Ardi WH, Rajbhandary S, Rubite R, Twyford AD and 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.

Tirel C (1983) Une nouvelle espéce d'Elaeocarpus (Elaeocarpaceae) de Nouvelle-Calédonie. *Bulletin du Museum National d' Histoire Naturelle Section B, Adansonia, no 3, botanique, phytochimie* **5**, 283 – 287.

Tirel C (1985) Éléocarpacées. In: Humbert H and Leroy J-F (eds), *Flore de Madagascar et des Comores: plantes vasculaires / pubiliée sous les auspices du gouvernement général de Madagascar et sous la direction de H.Humbert*. pp. 1 – 116. Muséum National D'Historire Naturelle, Paris.

Tripp EA and McDade LA (2014) A Rich Fossil Record Yields Calibrated Phylogeny for Acanthaceae (Lamiales) and Evidence for Marked Biases in Timing and Directionality of Intercontinental Disjunctions. *Systematic Biology* **63**, 660 – 684. doi:10.1093/sysbio/syu029

Turner B, Munzinger J, Duangjai S, Temsch EM, Stockenhuber R, Barfuss MHJ, Chase MW and Samuel R (2013) Molecular phylogenetics of New Caledonian *Diospyros* (Ebenaceae) using plastid and nuclear markers. *Molecular Phylogenetics and Evolution* **69**, 740 – 763.

Tutar Y (2012) Pseudogenes. *Comparative and Funtional Genomics* **2012**, Article ID 424526, 4 pages. doi:10.1155/2012/424526

Ufford AQ van and Cloos M (2005) Cenozoic tectonics of New Guinea. *American Association of Petroleum Geologists Bulletin* **89**, 119 – 140.

Van Steenis CGGJ (1949) General considerations. In: Van Steenis CGGJ (ed), *Flora Malesiana*. Series I, volume 4, pp. XLVII, Noordhoff-Kolff N.V., Batavia.

Van Steenis CGGJ (1950) The delimitation of Malesia and its main plant geographic regions. *Flora Malesiana Series I*, 1, lxxvi – cvi.

Van Welzen PC and Slik JW (2009) Patterns in species richness and composition of plant families in the Malay Archipelago. *Blumea* **54**, 166 – 171. doi: 10.3767/000651909X475969

Van Welzen PC, Strijk JS, Van Konijnenburg-Van Cittert JHA, Nucete M and Merckx VSFT (2014) Dated phylogenies of the sister genera *Macaranga* and *Mallotus* (Euphorbiaceae): congruence in historical biogeographic patterns? *PLoS ONE*, **9**, e85713.

Veevers JJ (2001) *Atlas of billion-year earth history of Australia and neighbours in Gondwanaland*. Gemoc Press, Sydney.

Wagner WL, Weller SG and Sakai AK (2005) Monograph of *Schiedea* (Caryophyllaceae–Alsinoideae). *Systematic Botany Monographs* **72**, 1 – 169.

Wanntorp L and Wanntorp H-E (2003) The biogeography of *Gunnera* L.: vicariance and dispersal. *Journal of Biogeography* **30**, 979 – 987.

Waters JM and McDowall RM (2005) Phylogenetics of the Australasian mudfishes: evolution of an eel-like body plan. Molecular Phylogenetics and Evolution **37**, 417 – 425.

Waters JM and Craw D (2006) Goodbye Gondwana? New Zealand biogeography, geology, and the problem of circularity. *Systematic Biology* **55**, 351 – 356.

Waterston RH *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520 – 562.

Weber E (1994) The reproductive biology of *Elaeocarpus largiflorens* C. White subsp. *largiflorens* (Elaeocarpaceae). *Honours thesis*. James Cook University, Queensland.

Wen J, Zhang J-Q, Nie Z-L, Zhong Y and Sun H (2014) Evolutionary diversifications of plants on the Qinghai-Tibetan Plateau. *Frontiers in Genetics* **5**, article 4. doi: 10.3389/fgene.2014.00004

Weibel RW (1968) Morphologie de l'embryon et de la graine des *Elaeocarpus*. *Candollea* **23**, 101 – 108.

Weibel R (1971) Espèces nouvelles du genre Elaeocarpus provenant de la Nouvelle-Guinée et de ses dépendances, des îles Salomons et de Célèbes. *Candollea* **26**, 359 – 382.

Weir JT and Schluter D (2008) Calibrating the avian molecular clock. *Molecular Ecology* **17**, 2321 – 2328. doi: 10.1111/j.1365-294X.2008.03742.x

White TJ, Bruns T, Lee S and Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ

and White TJ (eds), *PCR protocols a guide to methods and applications*. pp. 315 – 322. Academic Press, San Diego, California.

Whitten WM, Williams NH and Chase MW (2000) Subtribal and generic relationship of Maxillarieae (Orchidaceae) with emphasis on Stanhopeinae: combined molecular evidence. *American Journal of Botany* **87**, 1842 – 1856.

Wikström N, Savolainen V and Chase MW (2001) Evolution of the angiosperms: calibrating the family tree. *Proceedings of the Royal Society of London B: Biological Sciences* **268**, 2211 – 2220. doi: 10.1098/rspb.2001.1782.

Wilgenbusch JC, Warren DL and Swofford DL (2004) AWTY: A system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference. http://ceb.csit.fsu.edu/awty. Accessed 14 June 2012.

Winkworth RC, Wagstaff SJ, Glenny D and Lockhart PJ (2005) Evolution of the New Zealand mountain flora: origins, diversification and dispersal. *Organisms Diversity and Evolution* **5**, 237 – 247.

Wong KM (1998) Patterns of plant endemism and rarity in Borneo and the Malay Peninsula. In: Peng CI and Lowry II PP (eds), Rare, threatened, and endangered floras of Asia and the Pacific rim. *Academia Sinica Monograph* **16**, 139 – 169.

Wu CI and Li WH (1985) Evidence for higher rates of nucleotide substitution in rodents than in man. *Proceedings of the National Academy of Sciences of the United States of America* **82**, 1741 – 1745.

Yang Z (2006) Computational molecular evolution. Oxford University Press, New York.

Yi S, Ellsworth DL and Li WH (2002) Slow molecular clocks in Old World monkeys, apes, and humans. *Molecular Biology and Evolution* **19**, 2191 – 2198.

Zachos J, Pagani M, Sloan L, Thomas E and Billups K (2001) Trends, rhythms, and aberration in global climate 65 Ma to present. *Science* **292**, 686 – 693.

Zuckerkandl E and Pauling L (1962) Evolutionary Divergence and Convergence in Proteins. In: Bryson V and Vogel HJ (eds), *Evolving Genes and Proteins*. Academic Press, New York.

Zwickl DJ (2008) GARLI manual (program version 0.96, manual revision 1). http://www.nescent.org/wg/garli. Accessed 2 May 2013.

Zwickl DJ (2011) GARLI 2.0. https://www.nescent.org/wg_garli/main_page. Accessed 2 May 2013.

Appendix 1.1 Comparison of diagnostic morphological characters within the *Polystachyus* group complex showing continuous and overlapped character states and unclear taxon boundaries.

Elaeocarpus cupreus (EC), *Elaeocarpus clementis* var. *borneensis* (ECb), *Elaeocarpus clementis* var. *clemensiae* (ECs), *Elaeocarpus clementis* var. *clementis* (ECc), *Elaeocarpus clementis* var. *kostermansii* (ECk), *E. integripetalus* (EI), *Elaeocarpus multinervosus* (EM), *Elaeocarpus polyanthus* (EP) and *Elaeocarpus polystachyus* (ES) (Regular text = information extracted from Coode 1996c and unpublished data; italic text = observations in this study)

Characters	EC	ECb	ECs	ECl	ECk	EI	EM	EP	ES
Young twigs	Glabrous	Densely hairy	Hairy to glabrous	Hairy to glabrous	Densely hairy	Glabrescent	*Hairy*	*Hairy*	*Glabrous or hairy at tip*
Petioles	≤ 1.8 mm thick	*0.8–2.5 mm thick*	*0.6–1.9 mm thick*	*0.8–2.2 mm thick*	1.3–2.3 mm thick	*c. 1.7 mm thick*	≥ 2 mm thick	*1.1–2.5 mm thick*	*1–1.6 mm thick*
Secondary veins	12 pairs	(5–)10–14(–16) pairs	(5–)10–14(–16) pairs	(5–)10–14(–16) pairs	(5–)10–14(–16) pairs	8–11 pairs	12–17 pairs	*10–14 pairs*	*6–13 pairs*
Inflorescences	10–25-flowered	(15–)20–50-flowered	(15–)20–50-flowered	(15–)20–50-flowered	(15–)20–50-flowered	15–20-flowered	30–60-flowered	12–30-flowered	25–38-flowered
indumentum	Hairy	Densely hairy	Hairy to glabrous	Hairy to glabrous	Densely hairy	Sparsely hairy	*Hairy*	*Hairy*	*Hairy*
Flower sexuality	Unisexual	Bisexual	Bisexual	Bisexual	Bisexual	Bisexual	Unisexual	Bisexual	*Bisexual*
Flower-mery	5	5	5	5	5	4	5	5	4

Flower buds	Ovoid-globose; bracts caducous	*Globose to ovoid; bracts caducous*	Round to ovoid-globose, rounded at apex; bracts with lateral teeth or lobes	Conical or ovoid to broad-ovoid, obtuse to acute at apex; bracts entire	Rounded to obtuse or acute at apex; bracts caducous	Broad ovoid, subacute at apex; bracts unknown	*Globose; bracts caducous*	*Globose to ovoid; bracts caducous*	*Globose to ovoid; bracts caducous*
Petals	4–5(–5.5) mm long	≥ 5 mm long	4.5–6(–6.5) mm long	5–6(–8) mm long	≥ 5 mm long	5–6 mm long	*4.3–7.5 mm long*	*5.5–6.8 mm long*	*4–7 mm long*
Petals' apexes	Pointed, acute or subacute	Pointed, subacute or 2 or 3-denticulate	Pointed, subacute or 2 or 3-denticulate	Pointed, subacute or 2 or 3-denticulate	Pointed, subacute or 2 or 3-denticulate	Rounded or rarely 1–5-denticulate	Pointed, subacute, subobtuse or occasionally acuminate	Pointed or 1 or 2-denticulate	Rounded, subentire or 1–6-denticulate, rarely attenuate and subacute
Pedicels	< 1 mm thick	< 1 mm thick	< 1 mm thick	< 1 mm thick	< 1 mm thick	> 1.6 mm thick	≤ 1 mm thick	> 1 mm thick	1 mm thick
Ovules	5–6	(4–)6(–10)	6(–8)	6(–8)	8	8	8	8–9	10–12
Fruit stalk	*< 10 mm long*	< 10 mm long	< 10 mm long	< 10 mm long	< 10 mm long	unknown	*< 10 mm long*	> 12 mm long	*3–13 mm long*
Fruit indumentum	Glabrescent	Glabrous or virtually so	Glabrous or virtually so	Glabrous or virtually so	Glabrous or virtually so	unknown	*Glabrous*	Velvety	*Glabrescent, velvety when young*
Mesocarp	≤ 2 mm thick	≤ 2 mm thick	≤ 2 mm thick	≤ 2 mm thick	≤ 2 mm thick	unknown	≤ 2 mm thick	≥ 2 mm thick	2–4 mm thick
Stone	< 2 cm long	< 2 cm long	< 2 cm long	< 2 cm long	< 2 cm long	unknown	< 2 cm long	> 2 cm long	1.1–1.8 cm long
Wall	< 2 mm thick	< 2 mm thick	< 2 mm thick	< 2 mm thick	< 2 mm thick	unknown	< 2 mm thick	≥ 2 mm thick	< 2 mm thick

Appendix 2.1 Details of samples used in the phylogenetic and molecular dating analyses. Provided for each sample is the taxon, authority, infrageneric assignment (according to Baba and Crayn 2012, Coode 1978, 1984, 1996a – d, 1998, 2001c, e, f, 2010, pers. comm. 2014, Coode and Weibel 1994 and Weibel 1968), voucher details and herbarium where deposited, DNA sample number and institution where extracted (G and CNS: extraction done at CNS; NSW: extraction done at NSW and donated for this study). + New sequence generated in this study; FS Unpublished sequences donated by Ferry Slik (pers. comm. 2011); NA Information not available or sequence not generated. GenBank accession numbers are provided for sequences obtained from that source. AK (Auckland War Memorial Museum), BRI (Queensland Herbarium), BRUN (Brunei Forestry Centre), CM (Carnegie Museum of Natural History Herbarium), CNS (Australian Tropical Herbarium), K (Royal Botanic Gardens, Kew), KEP (Forest Research Institute Malaysia, Kepong), KYO (Graduate School of Science, Kyoto University), MEL (Royal Botanic Gardens, Melbourne), MO (Missouri Botanical Garden), NOU (Institut de Recherche pour le Développement), NSW (National Herbarium of New South Wales), PERTH (Western Australian Herbarium), PTBG (National Tropical Botanical Garden, Kalaheo), SAN (Forest Research Centre, Sandakan), SAR (Department of Forestry, Kuching), UCBG (University of California Botanical Garden at Berkeley).

Species	**Group (subgroup)**	**Voucher (herbarium where deposited)**	**DNA extraction no.**	***psbA-trnH* spacer**	***trnL-trnF* region**	***trnV-ndhC* spacer**	***Xdh1***	***Xdh2***
Elaeocarpaceae								
Elaeocarpus								
Elaeocarpus acrantherus Merr.	*coilopetalum*	Borneo, S.N. Phoon *et al.* 422 (CNS, SAN)	G 01719	+	+	+	+	+
Elaeocarpus alaternoides Brongn. & Gris	*dicera*	New Caledonia, D.M. Crayn 749 (NSW)	NSW 60	+	KJ631296	KJ658421	+	+
Elaeocarpus angustifolius Blume	V (A)	Peninsular Malaysia, S.N. Phoon *et*	G 437	+	+	+	+	+

		al. 025 (CNS, KEP)						
Elaeocarpus angustifolius Blume	V (A)	New Guinea, D.M. Crayn 572 (NSW)	NSW 153	+	KJ631297	KJ658422	+	NA
Elaeocarpus angustifolius Blume	V (A)	India, NA (herbarium accession no.: NSW 710750) (NSW)	NSW 429	+	KJ631298	KJ658423	+	+
Elaeocarpus angustifolius Blume	V (A)	New Guinea, R.J. Johns 10685 (K)	NA	NA	DQ444689	NA	NA	NA
Elaeocarpus angustipes R.Knuth	*coilopetalum*	Borneo, J.B. Sugau SAN 153672 (CNS, SAN)	G 480	+	+	+	+	+
Elaeocarpus austro-yunnanensis Hu	Unassigned	China, NA (Sequence no.: China114)	NA	FS	NA	NA	NA	NA
Elaeocarpus arnhemicus F.Muell.	V (D)	Australia, Y. Baba 341 (CNS)	CNS 1852	NA	KJ631300	KJ658425	+	+
Elaeocarpus bakaianus Coode	V (C)	New Guinea, D.M. Crayn 584 (NSW)	NSW 168	+	NA	NA	NA	NA
Elaeocarpus bancroftii F.Muell.	VI (B)	Australia, Y. Baba 351 (CNS)	CNS 1856	+	KJ631301	KJ658426	+	+
Elaeocarpus bancroftii F.Muell.	VI (B)	Australia, D.M. Crayn 502 (NSW)	NA	NA	DQ444685	NA	NA	NA
Elaeocarpus barbulatus R.Knuth	*Elaeocarpus*	Borneo, S.N. Phoon *et al.* 375 (CNS, SAR)	G 01718	+	+	+	NA	NA
Elaeocarpus beccarii A.DC. subsp. *beccarii*	*Elaeocarpus*	Borneo, S.N. Phoon *et al.* 411 (CNS, SAN)	G 01178	+	+	+	+	+
Elaeocarpus bifidus Hook. & Arn.	Unassigned	Hawaii, Trauernicht 649 (PTBG)	CNS QC6	+	KJ631302	KJ658427	+	NA
Elaeocarpus brachypodus	VI	New Caledonia, Y. Pillon 71 (NOU)	NSW 1012	+	KJ631303	+	+	+

Guillaumin								
Elaeocarpus brunnescens R.Knuth	*Elaeocarpus*	Borneo, S.N. Phoon *et al.* 385 (CNS, SAN)	G 01414	+	+	+	NA	NA
Elaeocarpus brunnescens R.Knuth	*Elaeocarpus*	Borneo, S.N. Phoon *et al.* 419 (CNS, SAN)	G 01415	+	+	+	+	+
Elaeocarpus bullatus Tirel	Unassigned	New Caledonia, J. Munzinger 2906 (NOU)	NSW 1026	+	KJ631304	KJ658429	+	NA
Elaeocarpus carolinae B.Hyland & Coode	VII	Australia, A. Ford 4444 (CNS)	NSW 2735	NA	KJ631305	KJ658430	+	+
Elaeocarpus carolinensis Koidz.	Unassigned	Caroline Island, Lorence 10004 (PTBG)	CNS QG6	+	KJ631306	KJ658431	+	+
Elaeocarpus chinensis Hook.*f*. ex Benth.	Unassigned	China, C.C. Pang 51 (CNS)	G 01653	NA	NA	NA	+	+
Elaeocarpus chrysophyllus Merr.	*acronodia*	Borneo, S.N. Phoon *et al.* 269 (CNS, SAR)	G 01185	+	+	+	NA	+
Elaeocarpus clementis var. *borneensis* (Ridl.) Coode	*polystachyus*	Borneo, S.N. Phoon *et al.* 378 (CNS, SAN)	G 01179	+	+	+	+	+
Elaeocarpus clementis var. *clemensiae* (R.Knuth) Coode	*polystachyus*	Borneo, S.N. Phoon *et al.* 202 (CNS, BRUN)	G 01172	+	+	+	+	+
Elaeocarpus clementis Merr. var. *clementis*	*polystachyus*	Borneo, S.N. Phoon *et al.* 295 (CNS, SAR)	G 01187	+	+	+	+	+
Elaeocarpus coumbouiensis Guillaumin	Unassigned	New Caledonia, Y. Pillon *et al.* 388 (NOU)	NSW 2281	+	KJ631307	KJ658432	+	NA

Elaeocarpus coorangooloo J.F. Bailey & C.T. White	VI (E)	Australia, Y. Baba 695 (CNS)	CNS 1939	+	KJ631308	KJ658433	+	+
Elaeocarpus cristatus Coode	*monocera* (*obtusus*)	Brunei, S.N. Phoon *et al.* 262 (CNS, BRUN)	G 01177 G 01721	+	+	+	+	+
Elaeocarpus culminicola Warb.	VII	Australia, Y. Baba 350 (CNS)	CNS 1854	+	KJ631310	KJ658435	NA	NA
Elaeocarpus culminicola Warb.	VII	Australia, D.M. Crayn 499 (NSW)	NSW 121	+	+	+	NA	NA
Elaeocarpus cupreus Merr.	*polystachyus*	Peninsular Malaysia, S.N. Phoon *et al.* 341 (CNS, KEP)	G 01166	+	+	+	+	+
Elaeocarpus dentatus (J.R. & G.Forst.) Vahl	V (D)	New Zealand, M. Renner *s.n.* (perhaps no voucher)	NSW 578	+	KJ675689	KJ658436	NA	NA
Elaeocarpus decipiens Hemsl.	Unassigned	China, living tree tag no.: 0124040	NA	HQ415428	NA	NA	NA	NA
Elaeocarpus decipiens Hemsl.	Unassigned	China, NA	NA	HQ426998	NA	NA	NA	NA
Elaeocarpus dognyensis Guillaumin	Unassigned	New Caledonia, Munzinger & McPherson 751 (MO)	G 508	+	+	+	NA	NA
Elaeocarpus dolichobotrys Merr.	*Elaeocarpus*	Borneo, J.B. Sugau SAN 153654 (CNS, SAN)	G 478	+	+	+	+	+
Elaeocarpus dongnaiensis Pierre	*Elaeocarpus*	Vietnam, Nguyen Van Du HNK 1118 (K)	NSW 1028	+	+	+	+	+
Elaeocarpus dubius A.DC.	Unassigned	China, C.C. Pang 47 (CNS)	G 01651	NA	NA	NA	+	+
Elaeocarpus elliffii B.Hyland & Coode	XI (B)	Australia, D.M. Crayn 884 (NSW)	NSW 892	NA	KJ675691	KJ658438	NA	NA

Elaeocarpus eumundi F.M.Bailey	VII	Australia, A. Ford 4459 (CNS)	NSW 624	+	KJ675692	KJ658439	+	NA
Elaeocarpus eumundi F.M.Bailey	VII	Australia, D.M. Crayn 505 (NSW)	NA	NA	DQ444682	NA	NA	NA
Elaeocarpus euneurus Stapf ex Ridl.	*acronodia*	Borneo, S.N. Phoon *et al.* 201 (CNS, BRUN)	G 01171	+	+	+	+	+
Elaeocarpus fairchildii Merr.	VI (D)	New Guinea, S.N. Phoon *et al.* 139 (CNS)	G 01189	+	+	+	+	+
Elaeocarpus ferrugineus (Jack) Steud.	*acronodia*	Peninsular Malaysia, S.N. Phoon *et al.* 015 (CNS, KEP)	G 427	+	+	+	NA	NA
Elaeocarpus ferrugineus (Jack) Steud.	*acronodia*	Singapore, S.N. Phoon *et al.* 061 (CNS)	G 467	+	+	+	+	NA
Elaeocarpus ferruginiflorus C.T.White	XI (B)	Australia, D.M. Crayn 882 (NSW)	NSW 1754	+	KJ675693	+	+	NA
Elaeocarpus ferruginiflorus C.T.White	XI (B)	Australia, G. Fensom 401 (NSW)	NA	NA	DQ444692	NA	NA	NA
Elaeocarpus floribundus Blume	*Elaeocarpus*	Peninsular Malaysia, S.N. Phoon *et al.* 008 (CNS, KEP)	G 420	+	+	+	+	+
Elaeocarpus floribundus Blume	*Elaeocarpus*	Singapore, S.N. Phoon *et al.* 054 (CNS)	G 462	+	NA	+	+	+
Elaeocarpus floribundus Blume	*Elaeocarpus*	Borneo, J.B. Sugau SAN 153675 (CNS, SAN)	G 482	+	+	+	+	NA
Elaeocarpus floribundus Blume	*Elaeocarpus*	Origin unknown, S.N. Phoon *et al.* 141 (CNS)	G 01413	+	+	+	NA	NA

Elaeocarpus foveolatus F.Muell.	XI (B)	Australia, D.M. Crayn 856 (NSW)	NSW 1988	+	KJ675694	+	+	NA
Elaeocarpus foveolatus F.Muell.	XI (B)	Australia, D.M. Crayn 819 (NSW)	NSW 1835	+	NA	NA	NA	NA
Elaeocarpus foveolatus F.Muell.	XI (B)	Australia, P.D. Hind 6265 (NSW)	NA	NA	DQ444691	NA	NA	NA
Elaeocarpus geminiflorus Brongn. & Gris	Unassigned	New Caledonia, J. Munzinger 2866 (NOU)	NSW 1023	+	KJ675695	+	NA	NA
Elaeocarpus glaber Blume	*Elaeocarpus*	Borneo, Garden collector *s.n.*, living tree tag no.: XXIV.A.149 (CNS)	G 486	+	+	+	NA	NA
Elaeocarpus glaber Blume	*Elaeocarpus*	Borneo, S.N. Phoon *et al.* 239 (CNS, BRUN)	G 01174	+	+	+	+	+
Elaeocarpus gordonii Tirel	Unassigned	New Caledonia, Y. Pillon & C. Grignon 300 (NOU)	NSW 2276	+	KJ675697	KJ658443	NA	NA
Elaeocarpus grandiflorus Sm.	*monocera* (*obtusus*)	Singapore, S.N. Phoon *et al.* 068 (CNS)	G 473	+	+	+	+	+
Elaeocarpus grandis F.Muell.	V (A)	Australia, P.I. Ford 27569 (BRI)	NSW 232	+	KJ675700	KJ658446	+	+
Elaeocarpus grahamii F.Muell.	VII	Australia, A. Ford & G. Sankowsky 4313 (CNS)	NSW 133	+	+	KJ658444	NA	NA
Elaeocarpus griffithii (Wight) A.Gray	*coilopetalum*	Borneo, S.N. Phoon *et al.* 322 (CNS, SAR)	G 01713	+	+	+	+	+
Elaeocarpus gummatus Guillaumin	Unassigned	New Caledonia, Y. Pillon & A. Vergnes 260 (NOU)	NSW 2279	+	KJ675701	KJ658447	+	+
Elaeocarpus holopetalus F.Muell.	X	Australia, J.M. Allen *s.n.* (herbarium accession no.: NSW 605470) (NSW)	NSW 8	+	KJ675702	KJ658448	+	+

Elaeocarpus hookerianus Raoul	V (D)	New Zealand, J.M. Allen *s.n.* (herbarium accession no.: NSW 605721) (NSW)	NA	NA	DQ444686	NA	NA	NA
Elaeocarpus hookerianus Raoul	V (D)	New Zealand, C.D. Kilgour 787 (AK)	CNS 2113	+	KJ675703	KJ658449	+	+
Elaeocarpus hortensis var. *neocaledonica* Tirel	VI	New Caledonia, J. Munzinger 2968 (NOU)	NSW 1019	+	KJ675704	KJ658450	+	NA
Elaeocarpus hylobroma Y.Baba & Crayn	V	Australia, D.M. Crayn 838 (NSW)	NSW 2088	+	KJ675705	KJ658451	+	+
Elaeocarpus hylobroma Y.Baba & Crayn	V	Australia, D.M. Crayn 888 (NSW)	NSW 2106	NA	NA	+	NA	NA
Elaeocarpus cf. *inopinatus* Coode	*Elaeocarpus*	Borneo, A. van der Ent *Elaeocarpus* sp. 2 (CNS)	G 01190	+	+	+	+	+
Elaeocarpus jacobsii Coode	*acronodia*	Borneo, S.N. Phoon *et al.* 435 (CNS, SAN)	G 01184	+	+	+	+	NA
Elaeocarpus japonicus Siebold	Unassigned	Japan, K. Aoki 010974 (KYO)	NSW 488	NA	NA	KJ658452	NA	NA
Elaeocarpus japonicus Siebold	Unassigned	China, living tree tag no.: 0319031	NA	HQ415431	NA	NA	NA	NA
Elaeocarpus johnsonii F.Muell. ex C.T. White	IV	Australia, W.W. Cooper 2122 (CNS)	CNS 1937	+	KJ675706	+	+	NA
Elaeocarpus jugahanus Coode	*coilopetalum*	Borneo, S.N. Phoon *et al.* 410 (CNS, SAN)	G 01180	+	+	+	+	+
Elaeocarpus jugahanus Coode	*coilopetalum*	Borneo, S.N. Phoon *et al.* 424 (CNS, SAN)	G 01717	+	NA	NA	NA	NA

Elaeocarpus kerstingianus Schltr.	*dicera*	Caroline Island, Perlman 21433 (PTBG)	CNS QE6	+	KJ675706	KJ658454	+	+
Elaeocarpus kirtonii F.Muell. ex F.M.Bailey	VII	Australia, D.M. Crayn 501 (NSW)	NSW 106	+	DQ444687	KJ658455	NA	NA
Elaeocarpus knuthii Merr. subsp. *knuthii*	*acronodia*	Borneo, C.D. Kilgour 859 (CNS)	G 400	+	+	+	+	+
Elaeocarpus kusanoi Koidz.	*coilopetalum*	Caroline Island, Lorence 9548 (PTBG)	CNS QH6	+	+	+	+	+
Elaeocarpus largiflorens C.T.White subsp. *largiflorens*	XI (B)	Australia, D.M. Crayn 796 (NSW)	NSW 443	+	KJ675708	KJ658456	+	+
Elaeocarpus largiflorens subsp. *retinervis* B.Hyland & Coode	XI (B)	Australia, Y. Baba 357 (CNS)	CNS 1864	+	KJ675709	KJ658457	+	+
Elaeocarpus largiflorens subsp. *retinervis* B.Hyland & Coode	XI (B)	Australia, D.M. Crayn 503 (NSW)	NA	NA	DQ444684	NA	NA	NA
Elaeocarpus linsmithii Guymer	VII	Australia, Y. Baba *et al.* 833 (CNS)	CNS G00135	+	KJ675710	KJ658458	+	+
Elaeocarpus cf. *macrophyllus* Blume	*Elaeocarpus*	Borneo, S.N. Phoon *et al.* 261 (CNS, BRUN)	G 01176	+	+	+	+	+
Elaeocarpus marginatus Stapf ex Weibel	*acronodia*	Borneo, S.N. Phoon *et al.* 200 (CNS, BRUN)	G 01170	+	+	+	+	+
Elaeocarpus cf. *marginatus* Stapf ex Weibel	*acronodia*	Borneo, S.N. Phoon *et al.* 234 (CNS, BRUN)	G 01173	+	+	+	NA	NA
Elaeocarpus mastersii King	*acronodia*	Borneo, S.N. Phoon *et al.* 436 (CNS,	G 01183	+	+	+	+	+

		SAN)						
Elaeocarpus mastersii King	*acronodia*	Peninsular Malaysia, S.N. Phoon *et al.* 011 (CNS, KEP)	G 423	+	+	+	+	NA
Elaeocarpus mastersii King	*acronodia*	Peninsular Malaysia, S.N. Phoon *et al.* 030 (CNS, KEP)	G 442	+	+	+	NA	NA
Elaeocarpus mastersii King	*acronodia*	Singapore, S.N. Phoon *et al.* 059 (CNS)	G 465	+	+	+	NA	NA
Elaeocarpus mastersii King	*acronodia*	Peninsular Malaysia, S.N. Phoon *et al.* 037 (CNS, KEP)	G 488	+	+	+	+	+
Elaeocarpus multiflorus (Turcz.) Fern.-Vill.	*coilopetlaum*	Maluku, Ambon, S.N. Phoon *et al.* 138 (CNS)	G 01716	+	+	+	+	+
Elaeocarpus multinervosus R.Knuth	*polystachyus*	Borneo, S.N. Phoon *et al.* 180 (CNS, BRUN)	G 01169	+	+	+	+	+
Elaeocarpus multisectus Schltr.	III (A)	New Guinea, D.M. Crayn 561 (NSW)	NSW 312	+	KJ675711	KJ658459	+	+
Elaeocarpus mutabilis Weibel	*Elaeocarpus*	Borneo, S.N. Phoon *et al.* 159 (CNS, BRUN)	G 01167	+	+	+	+	+
Elaeocarpus myrtoides subsp. *vinkii* Coode	V (E)	Papua New Guinea, D.M. Crayn 539 (NSW)	NSW 313	+	KJ631309	+	NA	NA
Elaeocarpus nanus subsp. *congestifolius* (R.Knuth) Coode	*acronodia*	Borneo, C.D. Kilgour 887 (CNS)	G 407	+	+	+	+	NA
Elaeocarpus nanus Corner subsp. *nanus*	*acronodia*	Peninsular Malaysia, S.N. Phoon *et al.* 014 (CNS, KEP)	G 426	+	+	+	+	NA

Elaeocarpus nitentifolius Merr. & Chun	*acronodia*	China, C.C. Pang 46 (CNS)	G 01652	NA	+	+	+	+
Elaeocarpus nitentifolius Merr. & Chun	*acronodia*	China, living tree tag no.: 1617023	NA	HQ415429	NA	NA	NA	NA
Elaeocarpus nitidus Jack	*Elaeocarpus*	Singapore, S.N. Phoon *et al.* 056 (CNS)	G 463	+	NA	+	+	+
Elaeocarpus nouhuysii Koord.	VI (C)	New Guinea, D.M. Crayn 530 (NSW)	NSW 164	+	+	+	NA	NA
Elaeocarpus nouhuysii Koord.	VI (C)	New Guinea, D.M. Crayn 533 (NSW)	NSW 161	+	+	+	NA	+
Elaeocarpus obtusus Blume subsp. *obtusus*	*monocera* (*obtusus*)	Java, S.N. Phoon *et al.* 137 (CNS)	G 01191	+	+	+	+	+
Elaeocarpus obovatus G.Don	V (D)	Australia, M. Rossetto & D.M. Crayn *s.n.* (perhaps no voucher)	NSW 237	+	KJ675714	KJ658461	+	+
Elaeocarpus obovatus G.Don	V (D)	Australia, M. Rossetto & D.M. Crayn *s.n.* (perhaps no voucher)	NSW 266	+	NA	NA	NA	NA
Elaeocarpus ovigerus Brongn. & Gris	VI	New Caledonia, D.M. Crayn 763 (NSW)	NSW 73	NA	NA	KJ658462	NA	NA
Elaeocarpus palembanicus (Miq.) Corner	*coilopetalum*	Peninsular Malaysia, S.N. Phoon *et al.* 042 (CNS, KEP)	G 453	+	NA	NA	+	NA
Elaeocarpus petiolatus (Jack) Wall.	*coilopetalum*	Peninsular Malaysia, S.N. Phoon *et al.* 040 (CNS, KEP)	G 451	+	+	+	NA	NA
Elaeocarpus petiolatus (Jack) Wall.	*coilopetalum*	Singapore, S.N. Phoon *et al.* 058	G 464	+	+	+	+	NA

		(CNS)						
Elaeocarpus polydactylus Schltr.	V (D)	New Guinea, D.M. Crayn 577 (NSW)	NSW 170	+	KJ675715	KJ658463	+	+
Elaeocarpus polystachyus Wall. ex Müll. Berol.	*polystachyus*	Singapore, S.N. Phoon *et al.* 055 (CNS)	G 509	+	+	+	NA	NA
Elaeocarpus polystachyus Wall. ex Müll. Berol.	*polystachyus*	Peninsular Malaysia, S.N. Phoon *et al.* 120 (CNS, KEP)	G 01412	+	+	+	+	+
Elaeocarpus cf. *pseudopaniculatus* Corner	*coilopetalum*	Borneo, S.N. Phoon *et al.* 426 (CNS, SAN)	G 01181	+	+	+	+	+
Elaeocarpus ptilanthus Schltr.	V (A)	New Guinea, D.M. Crayn 554 (NSW)	NSW 155	+	YB	YB	+	+
Elaeocarpus pulchellus Brongn. & Gris	I	New Caledonia, D.M. Crayn 758 (NSW)	NSW 152	+	KJ675717	KJ658465	+	+
Elaeocarpus reticulatus Sm.	VII	Australia, J.M. Allen *s.n.* (herbarium accession no.: NSW 605722) (NSW)	NSW 7	+	DQ444683	KJ658466	+	NA
Elaeocarpus robustus Roxb.	*Elaeocarpus*	Peninsular Malaysia, S.N. Phoon *et al.* 009 (CNS, KEP)	G 421	+	NA	+	+	+
Elaeocarpus robustus Roxb.	*Elaeocarpus*	Indonesia, Java, Bogor Botanic Gardens, tree no. VI.C.175, S.N. Phoon *et al*. P 136 (CNS)	G 01715	NA	+	NA	NA	NA
Elaeocarpus rotundifolius Brongn. & Gris	I	New Caledonia, D.M. Crayn 761 (NSW)	NSW 72	+	KJ675718	NA	NA	NA
Elaeocarpus rugosus Roxb.	*monocera*	China, NA (Sequence no.:	NA	FS	NA	NA	NA	NA

	(*obtusus*)	China118)						
Elaeocarpus ruminatus F.Muell.	XI (A)	Australia, Y. Baba 446 (CNS)	CNS 1936	+	KJ675719	KJ658467	+	NA
Elaeocarpus ruminatus F.Muell.	XI (A)	Australia, D.M. Crayn 883 (NSW)	NSW 1737	+	NA	NA	NA	NA
Elaeocarpus cf. *sadikanensis* R.Knuth	*coilopetalum*	Borneo, S.N. Phoon *et al.* 432 (CNS, SAN)	G 01720	+	+	+	+	+
Elaeocarpus sarcanthus Schltr.	VIII (D)	New Guinea, D.M. Crayn 582 (NSW)	NSW 165	+	NA	+	+	NA
Elaeocarpus sayeri F.Muell.	VIII (C)	New Guinea, D.M. Crayn 557 (NSW)	NSW 167	+	NA	NA	NA	NA
Elaeocarpus sedentarius D.J.Maynard & Crayn	Unassigned	Australia, Y. Baba *et al.* 408 (CNS)	CNS QB6	+	KJ675720	KJ658468	+382	NA
Elaeocarpus sedentarius D.J.Maynard & Crayn	Unassigned	Australia, D.J. Maynard 02 (NSW)	NA	NA	DQ444676	NA	NA	NA
Elaeocarpus sericopetalus F.Muell.	XI (B)	Australia, D.M. Crayn 823 (NSW)	NSW 1856	+	DQ444692	KJ658469	NA	NA
Elaeocarpus sericopetalus F.Muell.	XI (B)	Australia, D.M. Crayn 887 (NSW)	NSW 1742	+	NA	+	NA	NA
Elaeocarpus seringii Montrouz.	I	New Caledonia, J. Munzinger 2852 (NOU)	NSW 1018	+	KJ675721	KJ658470	+	+
Elaeocarpus speciosus Brongn. & Gris	Unassigned	New Caledonia, Y. Pillon 115 (NOU)	NSW 1016	+	KJ675725	NA	NA	+
Elaeocarpus sphaericus K.Schum.	V (A)	India, NA (herbarium accession no.: NSW 710753) (NSW)	NSW 428	NA	KJ631299	+	+	+

Elaeocarpus sphaericus K.Schum.	V (A)	Hawaii, Flynn 7277 (NSW)	CNS QF6	+	+	KJ658424	+	+
Elaeocarpus sphaerocarpus H.T. Chang	Unassigned	China, NA (Sequence no.: China120)	NA	FS	NA	NA	NA	NA
Elaeocarpus stellaris L.S.Sm.	VI (B)	Australia, C. Costion 3531 (CNS)	CNS 2107	+	KJ675726	KJ658472	+	+
Elaeocarpus stellaris L.S.Sm.	VI (B)	Australia, Y. Baba *s.n.* (CNS)	CNS 2037	+	NA	NA	NA	NA
Elaeocarpus stipularis Blume	*Elaeocarpus*	Peninsular Malaysia, C.L. Lim FRI 65187 (CNS, KEP)	G 412	+	+	+	+	+
Elaeocarpus stipularis Blume	*Elaeocarpus*	Java, NA	G 413	+	+	+	+	NA
Elaeocarpus stipularis Blume	*Elaeocarpus*	Peninsular Malaysia, S.N. Phoon *et al.* 005 (CNS, KEP)	G 417	+	+	+	NA	NA
Elaeocarpus stipularis Blume	*Elaeocarpus*	Singapore, S.N. Phoon *et al.* 053 (CNS)	G 461	+	+	+	NA	NA
Elaeocarpus submonoceras subsp. *lasionyx* (Stapf ex Ridl.) Weibel	*Elaeocarpus*	Borneo, S.N. Phoon *et al.* 170 (CNS, BRUN)	G 01168	+	+	+	+	+
Elaeocarpus subserratus Baker	Unassigned	Madagascar, J.S. Miller 8762 (MO)	G 506	+	+	+	+	+
Elaeocarpus sylvestris (Lour.) Poir.	Unassigned	China, living tree tag no.: 0020036	NA	HQ415432	NA	NA	NA	NA
Elaeocarpus sylvestris var. *ellipticus* (Thunb. ex Murray) Hara	Unassigned	Japan, K. Aoki 011355 (KYO)	NSW 481	+	KJ675728	+	+	+
Elaeocarpus thelmae B.Hyland & Coode	XI (B)	Australia, Y. Baba *et al.* 792 (CNS)	CNS 2108	+	KJ675729	KJ658478	+	+
Elaeocarpus truncatus Weibel	*monocera*	Borneo, S.N. Phoon *et al.* 352 (CNS,	G 01188	+	+	+	+	+

	(*verticellatae*)	SAR)						
Elaeocarpus valetonii Hochr.	*Elaeocarpus*	Borneo, S.N. Phoon *et al.* 428 (CNS, SAN)	G 01182	+	+	+	+	+
Elaeocarpus varunua Buch.-Ham. ex Mast.	Unassigned	China, NA (Sequence no.: China112)	NA	FS	NA	NA	NA	NA
Elaeocarpus weibelianus Tirel	IV	New Caledonia, J. Munzinger 2833 (NOU)	NSW 1020	NA	KJ675730	KJ658479	+	+
Elaeocarpus williamsianus Guymer	VI (F)	Australia, D.M. Crayn 513 (NSW)	NSW 33	NA	DQ444693	NA	NA	NA
Elaeocarpus yateensis Guillaumin		New Caledonia, P.P. Lowry II 5616 (MO)	G 489	+	+	+	+	+
Elaeocarpus sp. Mt Bellenden Ker (L.J.Brass 18336) Qld Herbarium	Unassigned	Australia, Y. Baba 443 (CNS)	CNS 1935	+	KJ675722	KJ658475	+	NA
Elaeocarpus sp. Mt Misery (L.J.Webb+ 10905) Qld Herbarium	Unassigned	Australia, A. Ford 4312 (NSW)	NSW 126	+	KJ675723	+	+	+
Elaeocarpus sp. Mt Windsor Tableland (L.W.Jessup & GJM 1378) Qld Herbarium	Unassigned	Australia, Y. Baba & C.D. Kilgour 397 (CNS)	CNS 1884	+	KJ675724	KJ658477	+	+
Aceratium								
Aceratium concinnum (S.Moore) C.T.White	-	Australia, D.M. Crayn 858 (NSW)	NSW 145	+	DQ444678	KJ658480	+	+
Aceratium doggrellii C.T.White	-	Australia, MB 04 (provided by A. Ford, perhaps no voucher)	NSW 135	+	NA	KJ658481	+	+

Aceratium ferrugineum C.T.White	-	Australia, M. Harrington 296 (CNS)	NSW 59	NA	DQ444681	NA	NA	NA
Aceratium ledermannii Schltr.	-	New Guinea, D.M. Crayn 534 (NSW)	NSW 53	NA	DQ444677	NA	NA	NA
Aceratium megalospermum (F.Muell.) van Balgooy	-	Australia, D.M. Crayn 523 (NSW)	NSW 120	+	DQ444679	KJ658482	+	+
Aceratium sericoleopsis van Balgooy	-	Australia, D.M. Crayn 779 (NSW)	NSW 79	+	DQ444680	KJ658483	+	+
Aristotelia								
Aristotelia australasica F.Muell.	-	Australia, J.M. Allen *s.n.* (herbarium accession no.: NSW 605725) (NSW)	NSW 9	NA	DQ444661	NA	NA	NA
Aristotelia chilensis Stuntz	-	Australia, J.M. Allen *s.n.* (herbarium accession no.: NSW 605486) (NSW)	NSW 10	NA	DQ444660	NA	NA	NA
Aristotelia fruticosa Hook.*f*.	-	Australia, M.W. Chase 781 (K)	NSW 21	NA	DQ444662	KJ658515	NA	NA
Aristotelia peduncularis (Labill.) Hook.*f*.	-	Australia, L. Mulcahy *s.n.* (herbarium accession no.: NSW 606884) (NSW)	NSW 200	NA	DQ444659	KJ658516	NA	NA
Aristotelia serrata Oliv.	-	Australia, J.M. Allen *s.n.* (herbarium accession no.: NSW 605729) (NSW)	NSW 16	NA	DQ444663	KJ658517	NA	NA
Crinodendron								
Crinodendron hookerianum Gay	-	South America, J.M. Allen *s.n.* (herbarium accession no.: NSW 605484) (NSW)	NSW 11	+	DQ444666	KJ658511	+	NA

Crinodendron patagua Molina	-	South America, J.M. Allen *s.n.* (herbarium accession no.: NSW 605483) (NSW)	NSW 12	+	DQ444665	KJ658512	+	+
Dubouzetia								
Dubouzetia campanulata Pancher ex Brongn. & Gris	-	New Caledonia, D.M. Crayn 745 (NSW)	NSW 76	+	DQ444667	KJ658506	+	+
Dubouzetia caudiculata Sprague	-	New Caledonia, G. McPherson 3305 (NSW)	NA	NA	DQ444668	NA	NA	NA
Dubouzetia confusa Guillaumin & Virot	-	New Caledonia, T.J. Entwistle *s.n.* (NSW)	NSW 81	NA	KJ675736	NA	NA	NA
Dubouzetia elegans Brongn. & Gris	-	New Caledonia, J. Munzinger 2928 (NOU)	NSW 1022	NA	NA	KJ658508	NA	NA
Dubouzetia guillauminii Virot	-	New Caledonia, MCP 19401 (NOU)	NSW 2278	NA	KJ675737	KJ658509	+	NA
Dubouzetia kairoi Coode	-	New Guinea, D.M. Crayn 578 (NSW)	NSW 51	NA	DQ444670	NA	NA	NA
Dubouzetia saxatilis A.R.Bean & Jessup	-	Australia, D. Silke *s.n.* (NSW)	NSW 119	+474	DQ444669	KJ658510	NA	NA
Peripentadenia								
*Peripentadenia mearsii (*C.T.White) L.S.Sm.	-	Australia, P.I. Forster 29760 (BRI)	NSW 67	+	DQ444672	+	NA	NA
Peripentadenia phelpsii B.Hyland & Coode	-	Australia, D.M. Crayn 887 (NSW)	NSW 146	+	DQ444671	KJ658514	+	+

Platytheca								
Platytheca galioides Steetz	-	Australia, A.N. Rodd & G. Fensom 4973 (NSW)	NSW 40	NA	DQ444694	KJ658504	NA	NA
Sericolea								
Sericolea calophylla subsp. *grossiserrata* Coode	-	New Guinea, D.M. Crayn 550 (NSW)	NSW 78	+	DQ444675	KJ658484	+	+
Sericolea gaultheria Schltr.	-	New Guinea, D.M. Crayn 553 (NSW)	NSW 56	NA	DQ444674	NA	NA	NA
Sericolea micans Schltr. var. *micans*	-	New Guinea, D.M. Crayn 536 (NSW)	NSW 55	+	DQ444673	KJ658485	+	+
Sloanea								
Sloanea grandiflora Sm.	-	South America, Miller & Hauk 9409 (MO)	G 505	+	+	NA	NA	NA
Sloanea langii F.Muell.	-	Australia, P.I. Forster 30070 (BRI)	NSW 104	NA	DQ444655	KJ658519	NA	NA
Sloanea lepida Tirel	-	New Caledonia, J. Munzinger 3778 (NOU)	NSW 2277	NA	KJ675738	KJ658520	NA	NA
Sloanea macbrydei F.Muell.	-	Australia, A. Ford & B. Hewett 4295 (CNS)	NSW 136	NA	KJ675739	KJ658521	NA	NA
Sloanea montana (Labill.) A.C.Sm.	-	New Caledonia, D.M. Crayn 765 (NSW)	NSW 225	NA	KJ675740	KJ658522	NA	NA
Sloanea picapica Standl.	-	Central America, Stevens & Montiel J. 28053 (MO)	G 512	+	+	NA	NA	NA

Sloanea raynaliana Tirel	-	New Caledonia, Munzinger & McPherson 743 (MO)	G 507	+	+	NA	NA	NA
Sloanea rhodantha (Baker) Capuron	-	Madagascar, Razanatsima 301 (MO)	G 511	+	+	NA	NA	NA
Sloanea sogerensis Baker *f.*	-	Papua New Guinea, D.M. Crayn 532 (NSW)	NSW 54	NA	DQ444657	KJ658523	NA	NA
Sloanea tomentosa Rehder & E.H.Wilson	-	Thailand, R. Pooma *s.n.* (CNS)	NSW 474	+	+	NA	NA	NA
Sloanea woollsii F.Muell.	-	Australia, D.M. Crayn 780 (NSW)	NSW 77	NA	DQ444654	KJ658524	NA	NA
Tetratheca								
Tetratheca affinis Endl.	-	Australia, Cranfield & Ward 126 (PERTH)	NSW 65	NA	NA	KJ658486	NA	NA
Tetratheca aphylla F.Muell. subsp. *aphylla*	-	Australia, NA	NA	NA	AY237265	NA	NA	NA
Tetratheca aphylla subsp. *megacarpa* R.Butcher	-	Australia, R. Butcher RB 908 (PERTH)	NA	NA	AY237268	YB	NA	NA
Tetratheca bauerifolia F.Muell. ex Schuchardt	-	Australia, T. Downing TD 38 (MEL)	NSW 68	NA	EF095748	KJ658487	NA	NA
Tetratheca ciliata Lindl.	-	Australia, T. Downing TD 33 (MEL)	NSW 69	NA	DQ444698	KJ658488	NA	NA
Tetratheca confertifolia Steetz	-	Australia, D.M. Crayn 722 (NSW)	NSW 80	NA	KJ675731	KJ658489	NA	NA
Tetratheca efoliata F.Muell.	-	Australia, NA (Perth 06208231) (PERTH)	NSW 108	NA	KJ675732	KJ658490	NA	NA

Tetratheca ericifolia Sm.	-	Australia, J. Howell *s.n.* (herbarium accession no.: NSW 619997) (NSW)	NSW 36	NA	EF095746	KJ658491	NA	NA
Tetratheca filiformis Benth.	-	Australia, R. Butcher RB 966 (PERTH)	NSW 102	NA	DQ444695	KJ658492	NA	NA
Tetratheca harperi F.Muell.	-	Australia, NA	NA	NA	AY237277	NA	NA	NA
Tetratheca hirsuta Lindl.	-	Australia, R. Butcher RB 915 (PERTH)	NSW 94	NA	EF095742	KJ658493	NA	NA
Tetratheca hispidissima Steetz	-	Australia, R. Butcher RB 964 (PERTH)	NSW 101	NA	KJ675733	KJ658494	NA	NA
Tetratheca juncea Sm.	-	Australia, M. Rossetto & D.M. Crayn 524 (NSW)	NSW 37	NA	DQ444696	KJ658495	NA	NA
Tetratheca nephelioides R.Butcher	-	Australia, R. Butcher & J.A. Wege RB 909 (PERTH)	NA	NA	AY237271	NA	NA	NA
Tetratheca nuda Lindl.	-	Australia, D.M. Crayn *et al.* 731 (NSW)	NSW 62	NA	KJ675734	KJ658496	NA	NA
Tetratheca parvifolia Joy Thomps.	-	Australia, R. Butcher RB 916 (PERTH)	NSW 95	NA	DQ444697	KJ658497	NA	NA
Tetratheca paynterae subsp. *cremnobata* R.Butcher	-	Australia, R. Butcher et al. RB 902 (PERTH)	NA	NA	AY237273	NA	NA	NA
Tetratheca pilifera Lindl.	-	Australia, R. Butcher RB 922 (PERTH)	NSW 96	NA	EF095745	KJ658498	NA	NA
Tetratheca pubescens Turcz.	-	Australia, T. Downing TD 39 (MEL)	NSW 70	NA	DQ444699	KJ658499	NA	NA

Tetratheca retrorsa Joy Thomps.	-	Australia, R. Butcher RB 929 (MEL)	NSW 98	NA	EF095743	KJ658500	NA	NA
Tetratheca rupicola Joy Thomps.	-	Australia, J. Bradford 871 (MO)	NA	NA	AF299192	NA	NA	NA
Tetratheca shiressii Blakely	-	Australia, D.M. Crayn & M. Rossetto 604 (NSW)	NA	NA	EF095747	NA	NA	NA
Tetratheca stenocarpa J.H. Willis	-	Australia, T. Downing TD 53 (MEL)	NSW 71	NA	DQ444700	KJ658501	NA	NA
Tetratheca thymifolia Sm.	-	Australia, D.M. Crayn 602 (NSW)	NSW 83	NA	KJ675735	KJ658502	NA	NA
Tetratheca virgata Steetz	-	Australia, R. Butcher RB 928 (PERTH)	NSW 97	NA	EF095749	KJ658503	NA	NA
Tremandra								
Tremandra diffusa DC. subsp. *diffusa*	-	Australia, R. Butcher RB 961 (PERTH)	NSW 100	NA	DQ444701	NA	NA	NA
*Tremandra stelliger*a R.Br.	-	Australia, D.M. Crayn 706 (NSW)	NSW 57	NA	DQ444702	KJ658505	NA	NA
Vallea								
Vallea stipularis var. *pyrifolia* F.Ballard	-	South America, M.W. Chase 654 (K)	NSW 20	NA	DQ444664	KJ658525	NA	NA
Brunelliaceae								
Brunellia colombiana Cuatrec.	-	South America, J.C. Bradford 753 (MO)	NA	NA	AF299181, AF299234	NA	NA	NA
Cephalotaceae								
Cephalotus follicularis Labill.	-	Australia, T.D. Macfarlane 2549 (MO);	NA	NA	AF299193, AF299246	NA	EU264228	EU264228

		NA, 2001.0116 (UCBG)						
Cunoniaceae								
Ackama rosifolia A.Cunn.	-	New Zealand, Bradford 909 (MO)	NA	NA	AF299162, AF299215	NA	NA	NA
Cunonia balansae Brongn. & Gris	-	New Caledonia, Bradford 617 (MO)	NA	NA	AF299155, AF299208	NA	NA	NA
Cunonia capensis L.	-	NA, 96.0024 (UCBG)	NA	NA	NA	NA	EU264229	EU264229
Davidsonia pruriens var. *jeryseyana* F.M.Bailey	-	Australia, Bradford 887 (MO)	NA	NA	AF299185, AF299238	NA	NA	NA
Oxalidaceae								
Averrhoa carambola L.	-	NA, C. Morton 2010-30 (CM)	NA	NA	NA	NA	EU264230	EU264230

Appendix 2.2 Morphological character states scored for parsimony reconstruction. Coding of characters for mapping onto the molecular phylogenetic trees was as follows: straight = 0; curved = 1; entire = 0; ruminate = 1. CNS (Australian Tropical Herbarium), K (Royal Botanic Gardens, Kew), NA infrageneric grouping not applicable.

Taxon	Group	Embryo shape	Endosperm ornamentation	Reference	Voucher
Elaeocarpus					
Elaeocarpus acrantherus Merr.	*coilopetalum*	curved	entire	Coode (1996d)	
Elaeocarpus alaternoides Brongn. & Gris	*dicera*	straight	entire	Tirel (1983)	
Elaeocarpus angustifolius Blume	V (A)	straight	entire	Coode (1984, 2010)	
Elaeocarpus angustipes R.Knuth	*coilopetalum*	curved	entire	Coode (unpublished)	
Elaeocarpus arnhemicus F.Muell.	V (D)	straight	entire	Coode (1978, 1984)	
Elaeocarpus bancroftii F.Muell.	VI (B)	straight	entire	Coode (1984)	
Elaeocarpus barbulatus R.Knuth	*Elaeocarpus*	straight	entire	Coode (2001c)	
Elaeocarpus beccarii A.DC. subsp. *beccarii*	*Elaeocarpus*	straight	entire	Coode (1994)	
Elaeocarpus bifidus Hook. & Arn.	Unassigned	straight	entire	Coode (unpublished)	
Elaeocarpus brachypodus Guillaumin	VI	straight	entire	Tirel (1983)	
Elaeocarpus brunnescens R.Knuth	*Elaeocarpus*	straight	entire	Coode (unpublished)	
Elaeocarpus bullatus Tirel	Unassigned	straight	entire	Tirel (1983)	
Elaeocarpus carolinae B.Hyland & Coode	VII	curved	entire	Coode (1984)	
Elaeocarpus carolinensis Koidz.	Unassigned	straight	entire	Coode (unpublished)	
Elaeocarpus chinensis Hook.f. ex Benth.	Unassigned	curved	entire	Tang & Phengklai (2007)	Y. Tsiang 3016 (K)
Elaeocarpus chrysophyllus Merr.	*acronodia*	curved	ruminate	Coode (1996b, unpublished)	
Elaeocarpus clementis var. *borneensis* (Ridl.) Coode	*polystachyus*	curved	ruminate	Coode (1996c, unpublished)	

Elaeocarpus clementis var. *clemensiae* (R.Knuth) Coode	*polystachyus*	curved	ruminate	Coode (1996c, unpublished)	
Elaeocarpus clementis Merr. var. *clementis*	*polystachyus*	curved	ruminate	Coode (1996c, unpublished)	
Elaeocarpus coumbouiensis Guillaumin	Unassigned	straight	entire	Coode (unpublished)	
Elaeocarpus coorangooloo J.F.Bailey & C.T.White	VI (E)	straight	entire	Coode (1984)	
Elaeocarpus cristatus Coode	*monocera* (*obtusus*)	straight	entire	Coode (1998, 2007, unpublished)	
Elaeocarpus culminicola Warb.	VII	curved	entire	Coode (1984)	
Elaeocarpus cupreus Merr.	*polystachyus*	curved	ruminate	Coode (1996c, unpublished)	
Elaeocarpus dentatus (J.R. & G.Forst.) Vahl	V (D)	straight	entire	Coode (1984)	
Elaeocarpus dognyensis Guillaumin	Unassigned	straight	entire	Tirel (1983)	
Elaeocarpus dolichobotrys Merr.	*Elaeocarpus*	straight	entire	Coode (unpublished)	
Elaeocarpus dongnaiensis Pierre	*Elaeocarpus*	straight	entire	Weibel (1968)	
Elaeocarpus dubius A.DC.	*coilopetalum*	curved	entire	Weibel (unpublished); Tang & Phengklai (2007)	W.T. Tsang 26754 (K)
Elaeocarpus elliffii B.Hyland & Coode	XI (B)	curved	ruminate	Coode (1984)	
Elaeocarpus eumundi F.M.Bailey	VII	curved	entire	Coode (1984)	
Elaeocarpus euneurus Stapf ex Ridl.	*acronodia*	curved	ruminate	Coode (1996b, unpublished)	
Elaeocarpus fairchildii Merr.	VI (D)	straight	entire	Coode (1978, unpublished)	
Elaeocarpus ferrugineus (Jack) Steud.	*acronodia*	curved	ruminate	Coode (1996b, unpublished)	
Elaeocarpus ferruginiflorus C.T.White	XI (B)	curved	ruminate	Coode (1984)	
Elaeocarpus floribundus Blume	*Elaeocarpus*	straight	entire	Coode (unpublished)	
Elaeocarpus foveolatus F.Muell.	XI (B)	curved	ruminate	Coode (1984)	
Elaeocarpus geminiflorus Brongn. & Gris	Unassigned	straight	entire	Tirel (1983)	

Elaeocarpus glaber Blume	*Elaeocarpus*	straight	entire	Coode (unpublished)
Elaeocarpus gordonii Tirel	Unassigned	straight	entire	Coode (unpublished); Tirel (1984)
Elaeocarpus grahamii F.Muell.	VII	curved	entire	Coode (1998, unpublished)
Elaeocarpus grandiflorus Sm.	*monocera* (*obtusus*)	straight	entire	Coode (1984)
Elaeocarpus grandis F.Muell.	V (A)	straight	entire	Coode (1984)
Elaeocarpus griffithii (Wight) A.Gray	*coilopetalum*	curved	entire	Coode (unpublished)
Elaeocarpus gummatus Guillaumin	Unassigned	straight	entire	Tirel (1983)
Elaeocarpus holopetalus F.Muell.	X	straight	entire	Coode (1984)
Elaeocarpus hookerianus Raoul	V (D)	straight	entire	Coode (1984)
Elaeocarpus hortensis var. *neocaledonica* Tirel	VI	straight	entire	Tirel (1983)
Elaeocarpus hylobroma Y.Baba & Crayn	V	straight	entire	Baba & Crayn (2012)
Elaeocarpus cf. *inopinatus* Coode	*Elaeocarpus*	straight	entire	Coode (1994, unpublished)
Elaeocarpus jacobsii Coode	*acronodia*	curved	ruminate	Coode (1996b, unpublished)
Elaeocarpus johnsonii F.Muell. ex C.T.White	IV	straight	entire	Coode (1984)
Elaeocarpus jugahanus Coode	*coilopetalum*	curved	entire?	Coode (1998, unpublished)
Elaeocarpus kerstingianus Schltr.	*dicera*	straight	entire	Weibel (1968)
Elaeocarpus kirtonii F.Muell. ex F.M.Baile	VII	curved	entire	Coode (1984)
Elaeocarpus knuthii Merr. subsp. *knuthii*	*acronodia*	curved	ruminate	Coode (1996b, unpublished)
Elaeocarpus kusanoi Koidz.	*coilopetalum*	curved?	ruminate?	Coode (unpublished)
Elaeocarpus largiflorens C.T.White subsp. *largiflorens*	XI (B)	curved	ruminate	Coode (1984)

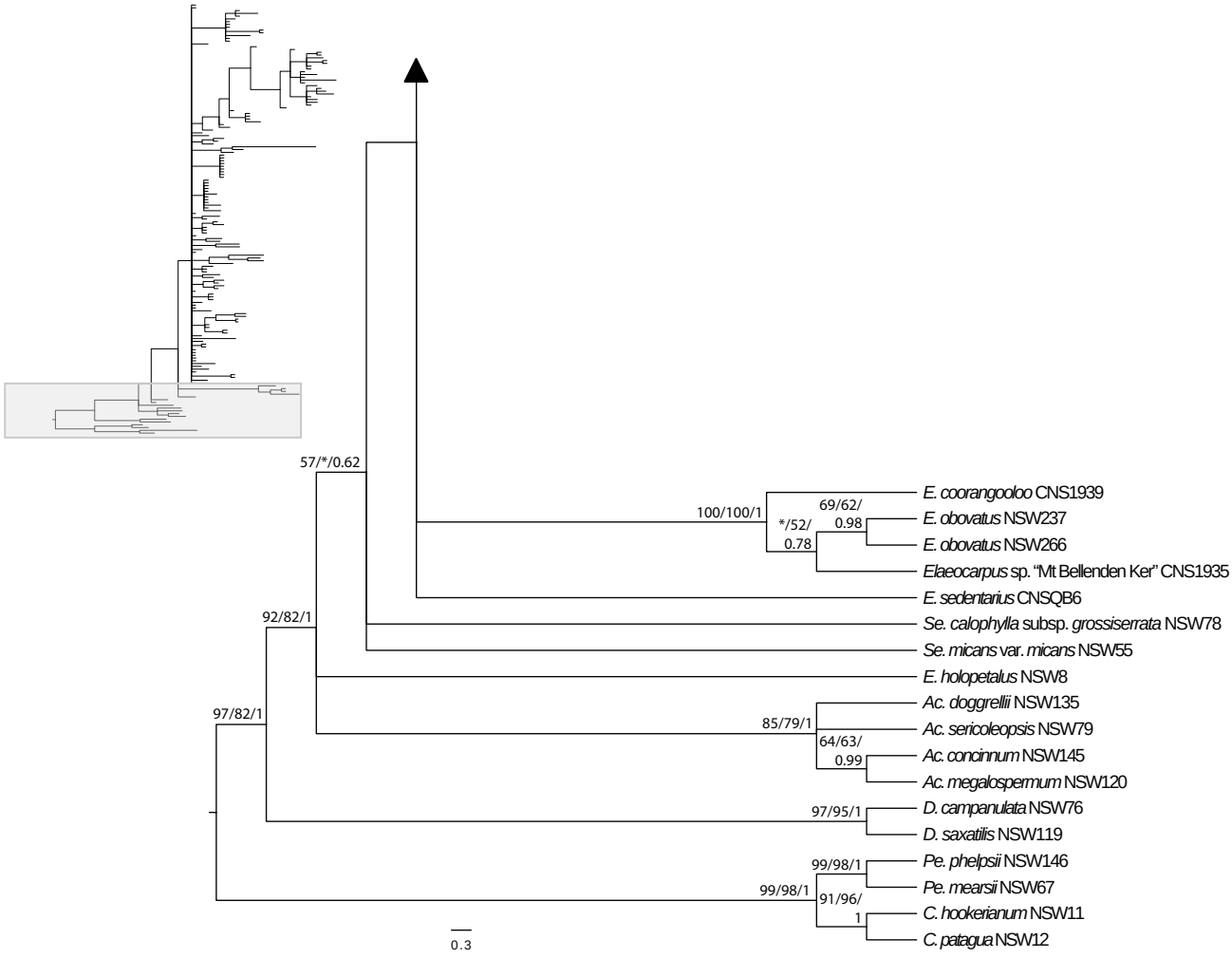
Elaeocarpus largiflorens subsp. *retinervis* B.Hyland & Coode	XI (B)	curved	ruminate	Coode (1984)	
Elaeocarpus linsmithii Guymer	VII	unknown	unknown	Coode (1984)	
Elaeocarpus cf. *macrophyllus* Blume	*Elaeocarpus*	straight	entire	Coode (2001c, unpublished)	
Elaeocarpus marginatus Stapf ex Weibel	*acronodia*	curved	ruminate	Coode (1996b, unpublished)	
Elaeocarpus mastersii King	*acronodia*	curved	ruminate	Coode (1996b, unpublished)	
Elaeocarpus multiflorus (Turcz.) Fern.-Vill.	*coilopetalum*	curved	ruminate?	Coode (2001d)	
Elaeocarpus multinervosus R.Knuth	*polystachyus*	curved	ruminate	Coode (1996c, unpublished)	
Elaeocarpus multisectus Schltr.	III (A)	straight	entire	Coode (1978)	
Elaeocarpus mutabilis Weibel	*Elaeocarpus*	straight	entire	Coode (unpublished)	
Elaeocarpus myrtoides subsp. *vinkii* Coode	V (E)	straight	entire	Coode (1978)	
Elaeocarpus nanus subsp. *congestifolius* (R.Knuth) Coode	*acronodia*	curved	ruminate	Coode (1996b, unpublished)	
Elaeocarpus nanus Corner subsp. *nanus*	*acronodia*	curved	ruminate	Coode (1996b, unpublished)	
Elaeocarpus nitentifolius Merr. & Chun	*acronodia*	curved	ruminate	Tang & Phengklai (2007); Weibel (1968)	Tsang & Fung 692 (K)
Elaeocarpus nitidus Jack	*Elaeocarpus*	straight	entire	Coode (2001c, unpublished)	
Elaeocarpus nouhuysii Koord.	VI (C)	straight	entire	Coode (1978)	
Elaeocarpus obovatus G.Don	V (D)	straight	entire	Coode (1984)	
Elaeocarpus obtusus Blume subsp. *obtusus*	*monocera* (*obtusus*)	straight	entire	Coode (1984, 1998, unpublished)	
Elaeocarpus palembanicus (Miq.) Corner	*coilopetalum*	curved	entire	Coode (unpublished)	

Elaeocarpus petiolatus (Jack) Wall.	*coilopetalum*	curved	ruminate	Coode (1998, unpublished)	
Elaeocarpus polydactylus Schltr.	V (D)	straight	entire	Coode (1978)	
Elaeocarpus polystachyus Wall. ex Müll.Berol.	*polystachyus*	curved	ruminate	Coode (1996c, unpublished)	
Elaeocarpus cf. *pseudopaniculatus* Corner	*coilopetalum*	curved	entire?	Coode (1998, unpublished)	
Elaeocarpus ptilanthus Schltr.	V (A)	straight	entire	Coode (1978, 2010)	
Elaeocarpus pulchellus Brongn. & Gris	I	straight	entire	Tirel (1983)	
Elaeocarpus reticulatus Sm.	VII	curved	entire	Coode (1984)	
Elaeocarpus robustus Roxb.	*Elaeocarpus*	straight	entire	Coode (1996a, unpublished)	Pierre 1870 (K)
Elaeocarpus rotundifolius Brongn. & Gris	I	straight	entire	Tirel (1983)	
Elaeocarpus ruminatus F.Muell.	XI (A)	curved	ruminate	Coode (1984)	
Elaeocaprus cf. *sadikanensis* R.Knuth	*coilopetalum*	curved?	entire?	Coode (1998, unpublished)	
Elaeocarpus sarcanthus Schltr.	VIII (D)	curved	entire	Coode (1978)	
Elaeocarpus sedentarius D.J.Maynard & Crayn	Unassigned	straight	entire	Maynard *et al.* (2008)	
Elaeocarpus sericopetalus F.Muell.	XI (B)	curved	ruminate	Coode (1984)	
Elaeocarpus seringii Montrouz.	I	straight	entire	Tirel (1983)	
Elaeocarpus speciosus Brongn. & Gris	Unassigned	straight	entire	Tirel (1983)	
Elaeocarpus sphaericus K.Schum.	V (A)	straight	entire	Coode (1978, 1984, 2010)	
Elaeocarpus stellaris L.S.Sm.	VI (B)	straight	entire	Coode (1984)	
Elaeocarpus stipularis Blume	*Elaeocarpus*	straight	entire	Coode (2001c)	
Elaeocarpus submonoceras subsp. *lasionyx* (Stapf *ex* Ridl.) Weibel	*Elaeocarpus*	straight	entire	Coode (1994, unpublished)	

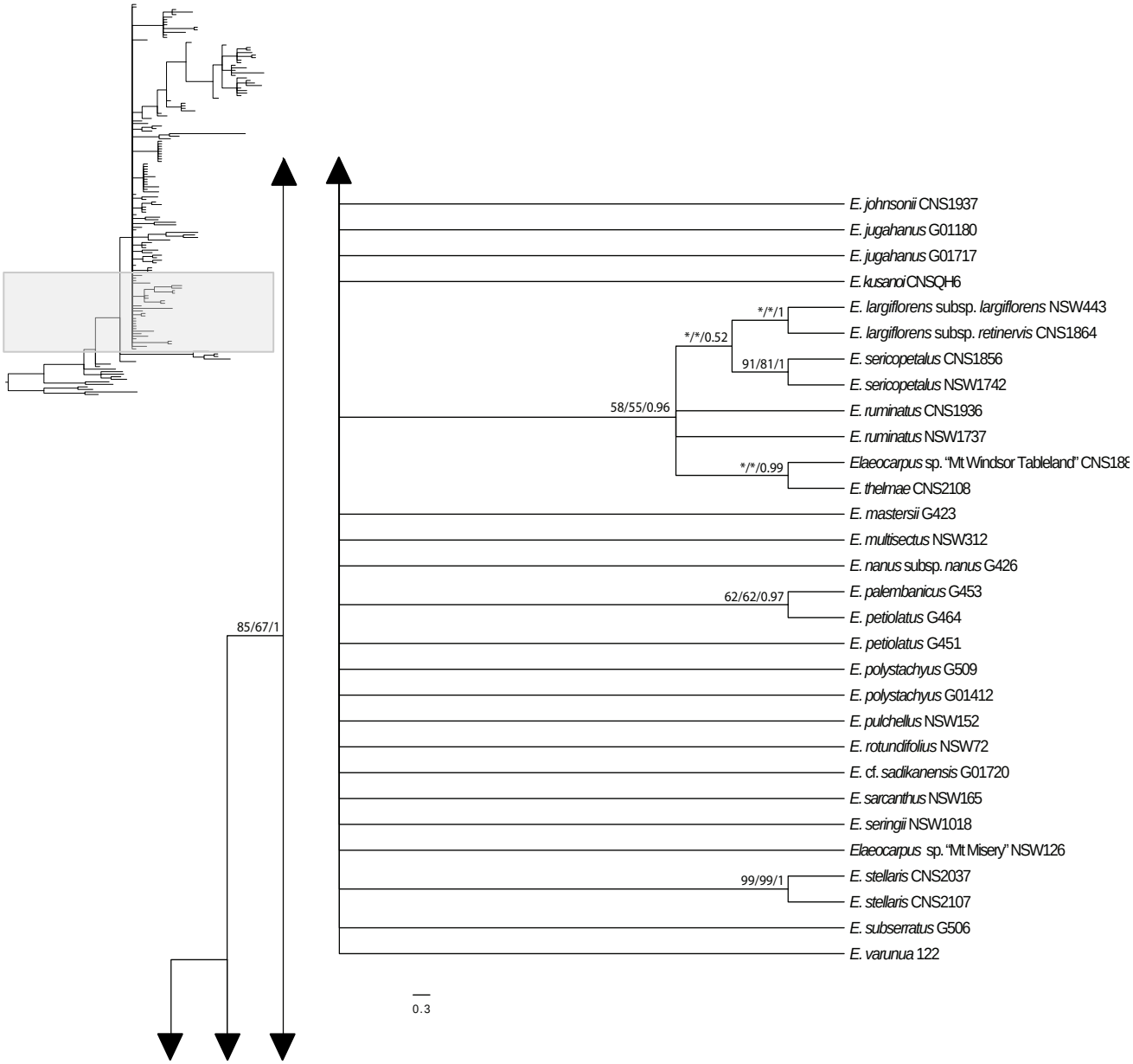
Elaeocarpus subserratus Baker	Unassigned	curved	entire	Tirel (1985)	Miller 8762 (K), Service Forester de Madagascar SF 12922 (K)
Elaeocarpus sylvestris var. *ellipticus* (Thunb. ex Murray) Hara	Unassigned	straight	entire	Tang & Phengklai (2007)	E.H. Wilson s.n. (22 November 1918) (K)
Elaeocarpus thelmae B.Hyland & Coode	XI (B)	curved	ruminate	Coode (1984)	
Elaeocarpus truncatus Weibel	*monocera* (*verticellatae*)	straight	entire	Coode (1998)	
Elaeocarpus valetonii Hochr.	*Elaeocarpus*	straight	entire	Coode (1996a, unpublished)	
Elaeocarpus weibelianus Tirel	IV	straight	entire	Tirel (1983)	
Elaeocarpus yateensis Guillaumin	Unassigned	straight	entire	Tirel (1983)	
Elaeocarpus sp. Mt Bellenden Ker (L.J.Brass 18336) Qld Herbarium	Unassigned	straight	entire	NA	B. Gray 2482 (CNS)
Elaeocarpus sp. Mt Misery (L.J.Webb+ 10905) Qld Herbarium	Unassigned	straight	entire	NA	B. Gray 5294 (CNS)
Elaeocarpus sp. Mt Windsor Tableland (L.W.Jessup & GJM 1378) Qld Herbarium	Unassigned	curved	ruminate	NA	M. Godwin C 3030 (CNS), B. Hyland 5541 (CNS)
Aceratium					
Aceratium concinnum (S.Moore) C.T.White	NA	straight	entire	Coode (1978, 2004)	
Aceratium doggrellii C.T.White	NA	straight	entire	Coode (1978, 2004)	
Aceratium megalospermum (F.Muell.) van Balgooy	NA	straight	entire	Coode (1978, 2004)	
Aceratium sericoleopsis van Balgooy	NA	straight	entire	Coode (1978, 2004)	
Crinodendron					
Crinodendron hookerianum Gay	NA	straight	entire	Coode (2004)	

Crinodendron patagua Molina	NA	straight	entire	Coode (2004)
Dubouzetia				
Dubouzetia campanulata Pancher ex Brongn. & Gris	NA	straight	entire	Coode (1978, 2004), Tirel (1983)
Dubouzetia guillauminii Virot	NA	straight	entire	Coode (1978, 2004), Tirel (1983)
Dubouzetia saxatilis A.R.Bean & Jessup	NA	straight	entire	Coode (1978, 2004), Tirel (1983)
Peripentadenia				
Peripentadenia mearsii (C.T.White) L.S.Sm.	NA	straight	entire	Coode (2004)
Peripentadenia phelpsii B.Hyland & Coode	NA	straight	entire	Coode (2004)
Sericolea				
Sericolea calophylla subsp. *grossiserrata* Coode	NA	curved	entire	Coode (1978, 2004)
Sericolea micans Schltr. var. *micans*	NA	curved	entire	Coode (1978, 2004)

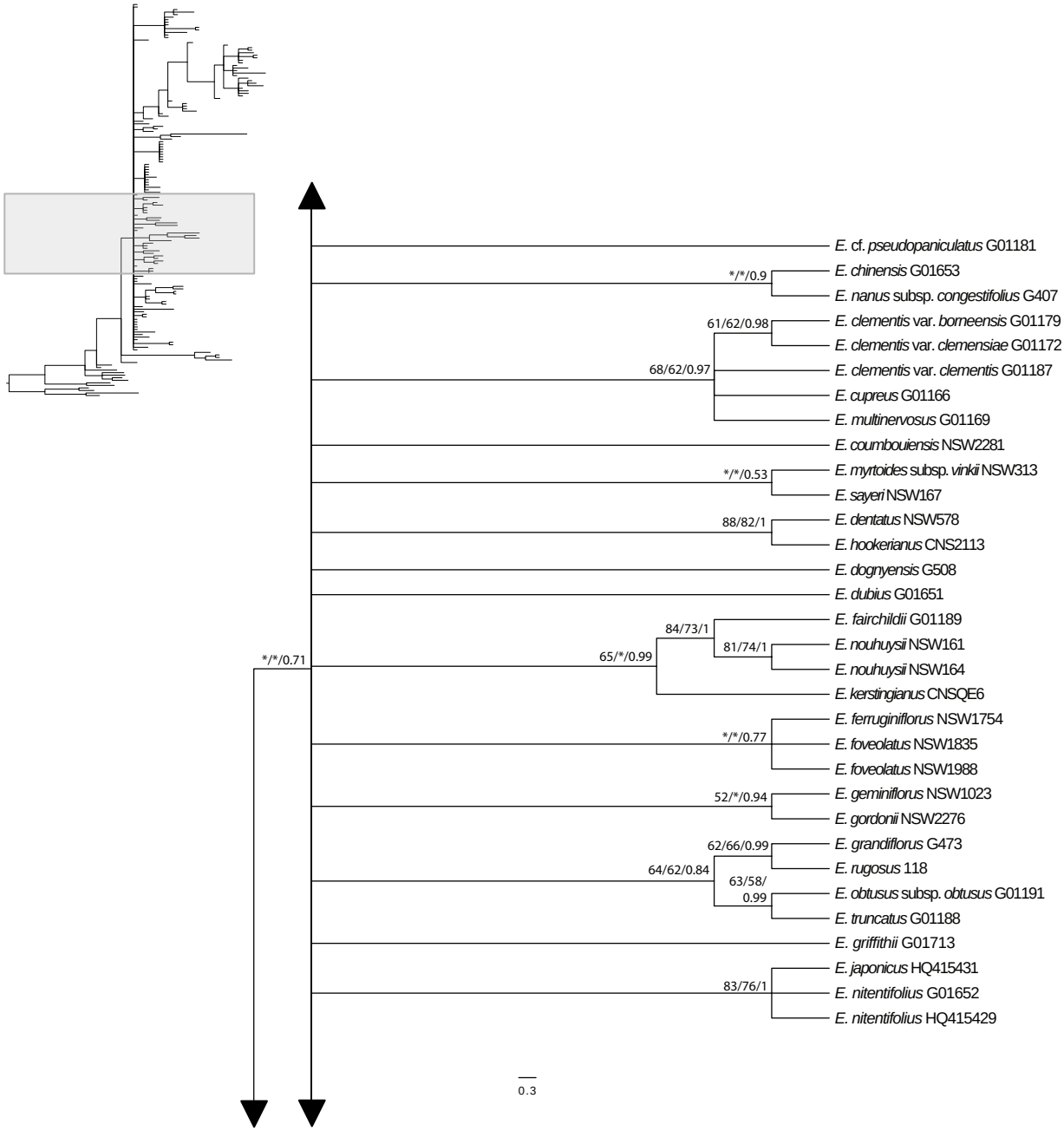
Appendix 2.3 Maximum clade credibility tree from Bayesian analysis of the *psbA-trnH* intergenic spacer. DNA extraction numbers or GenBank numbers are presented after taxon names. Branch support values are to the left of the nodes in the following order: MP bootstrap/ ML bootstrap/ Bayesian posterior probability. Genus abbreviations are *Aceratium* (Ac), *Crinodendron* (C), *Dubouzetia* (D), *Elaeocarpus* (E), *Peripentadenia* (Pe) and *Sericolea* (Se). * Branch collapses in the MP or ML 50 % majority rule consensus tree. The insert shows branch lengths proportional to the amount of nucleotide change and the section of the tree which has been enlarged.



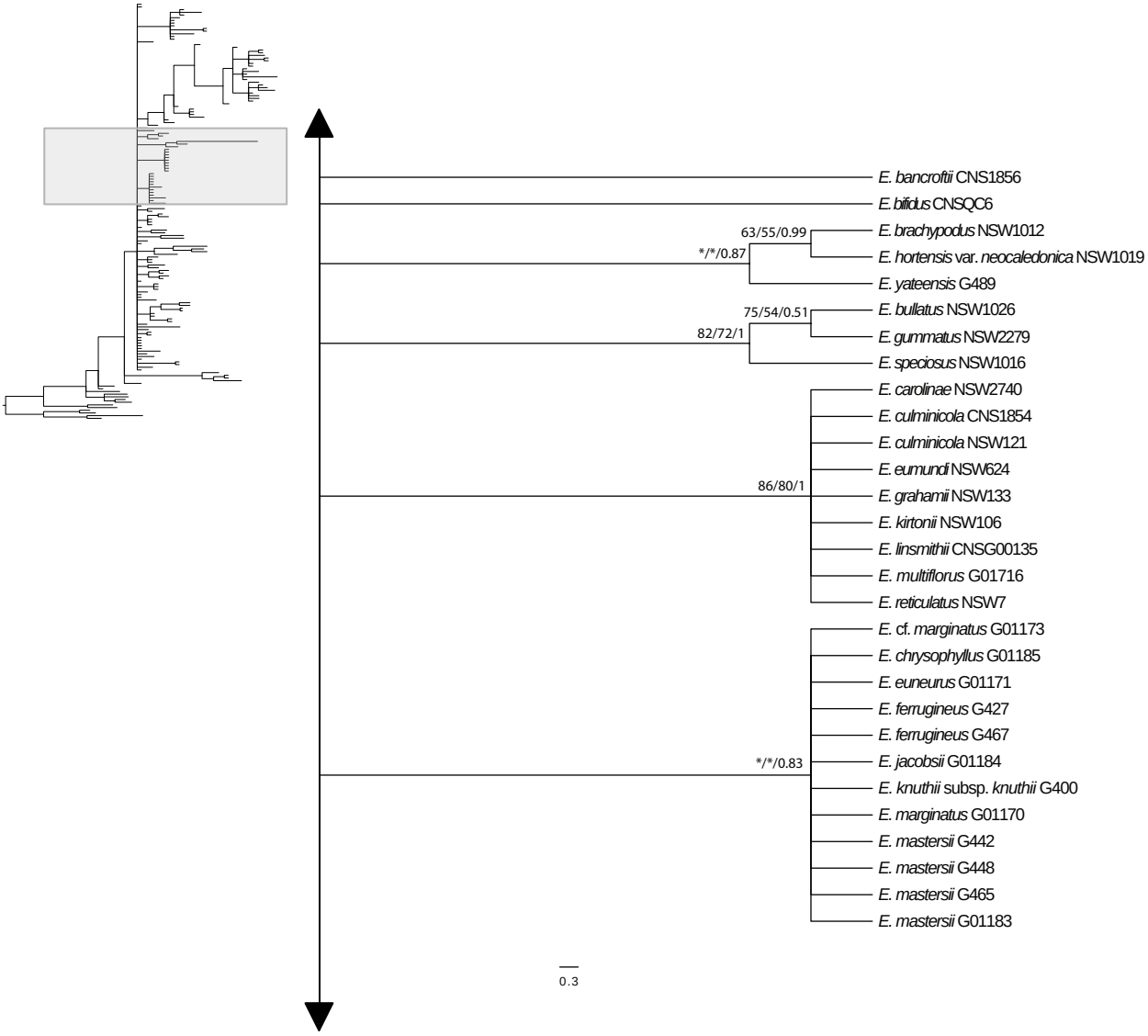
Appendix 2.3 (continued).



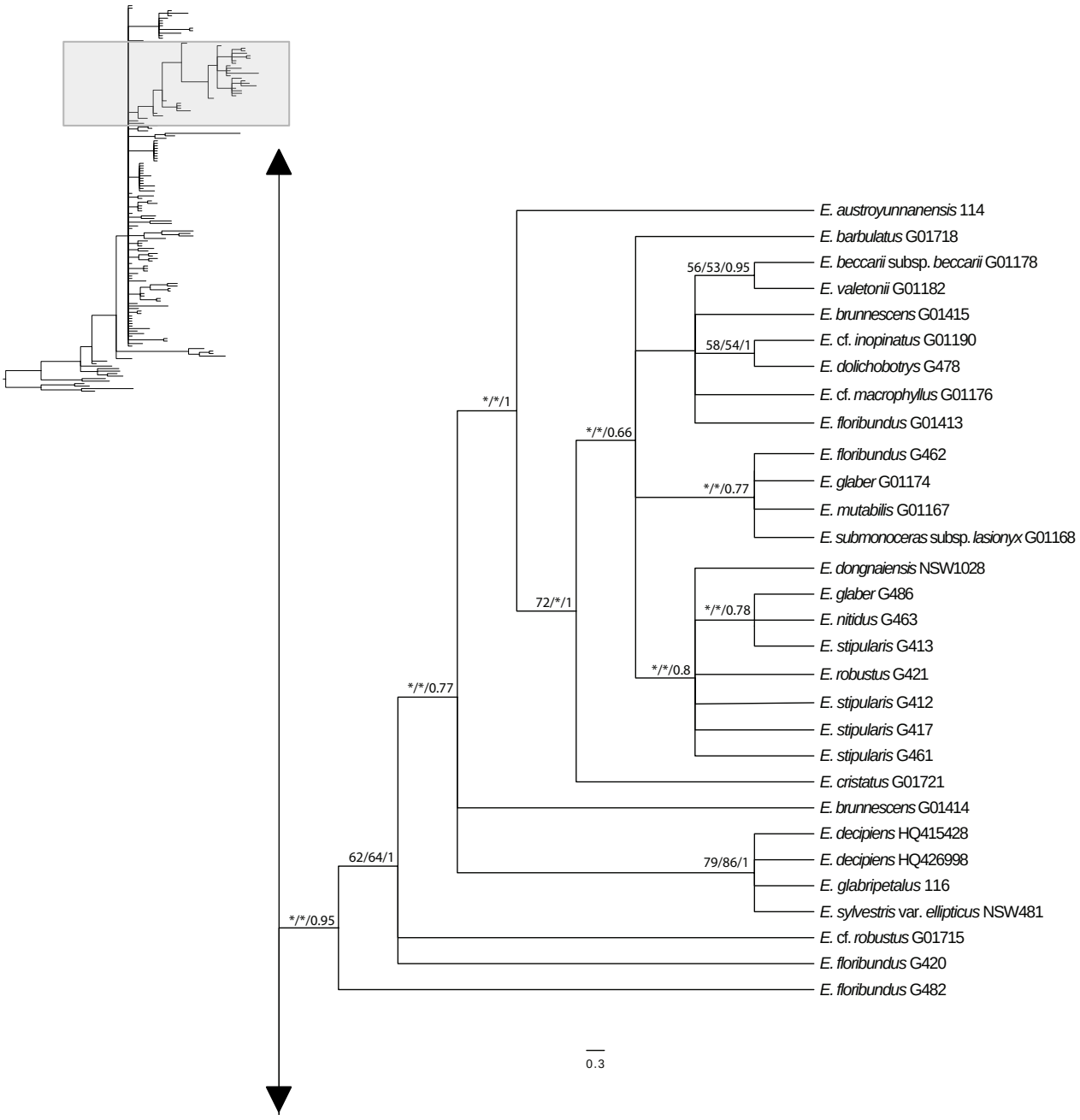
Appendix 2.3 (continued).



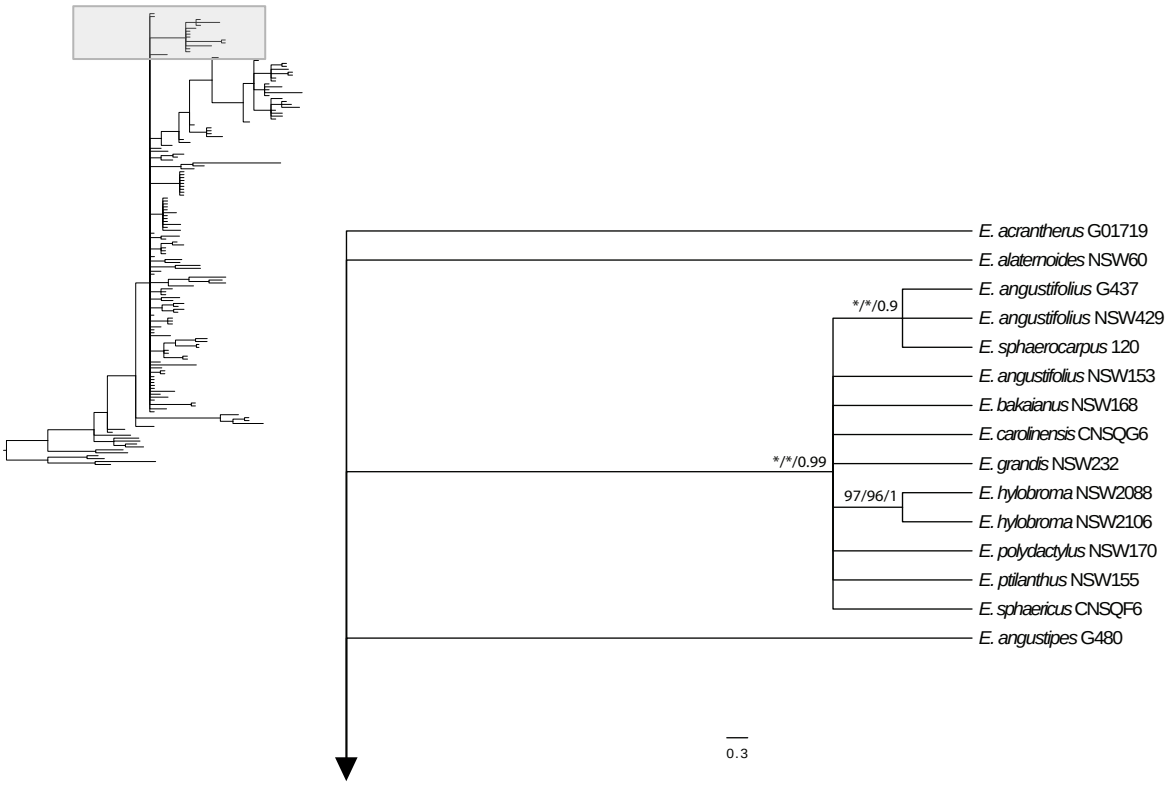
Appendix 2.3 (continued).



Appendix 2.3 (continued).

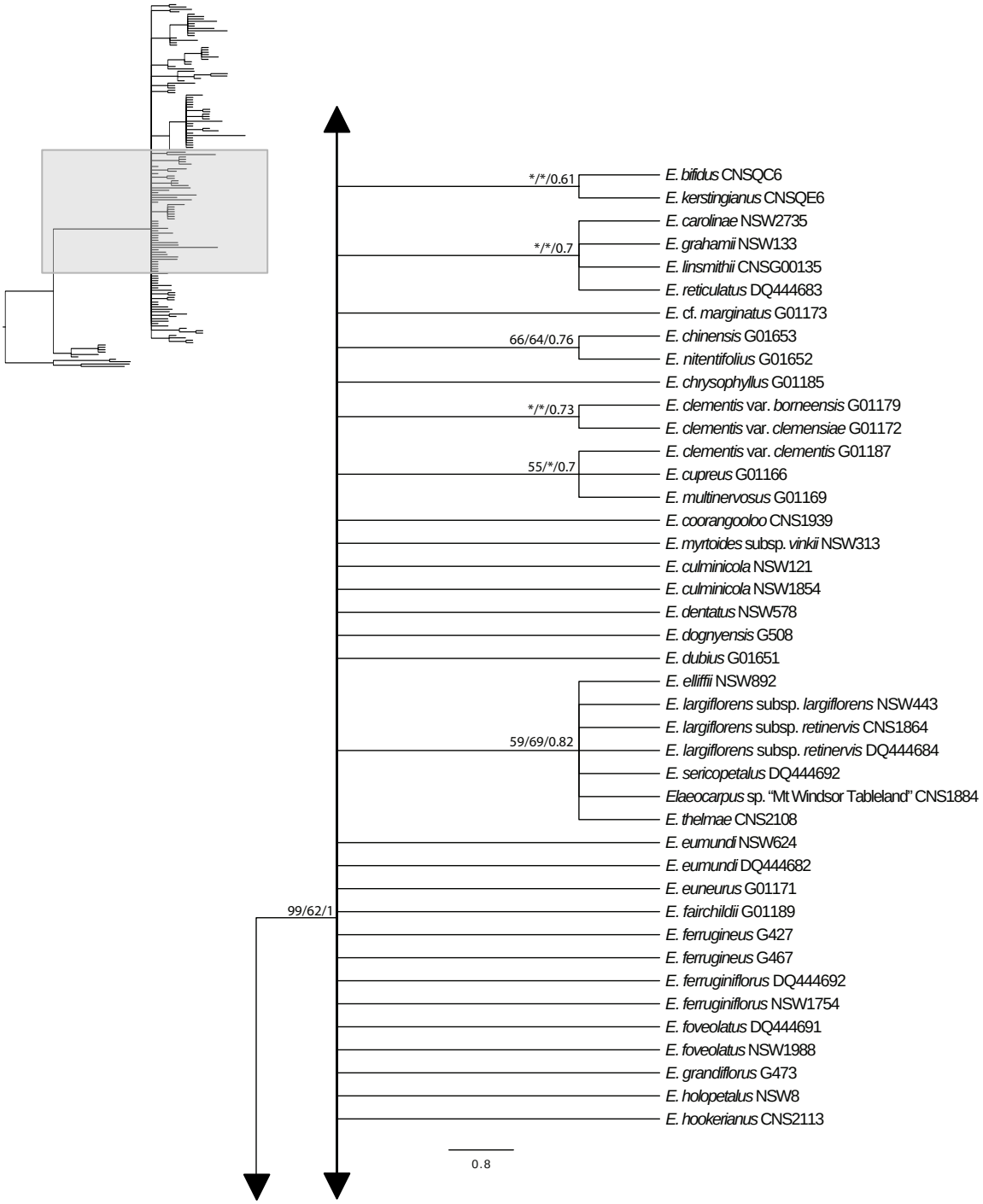


Appendix 2.3 (continued).

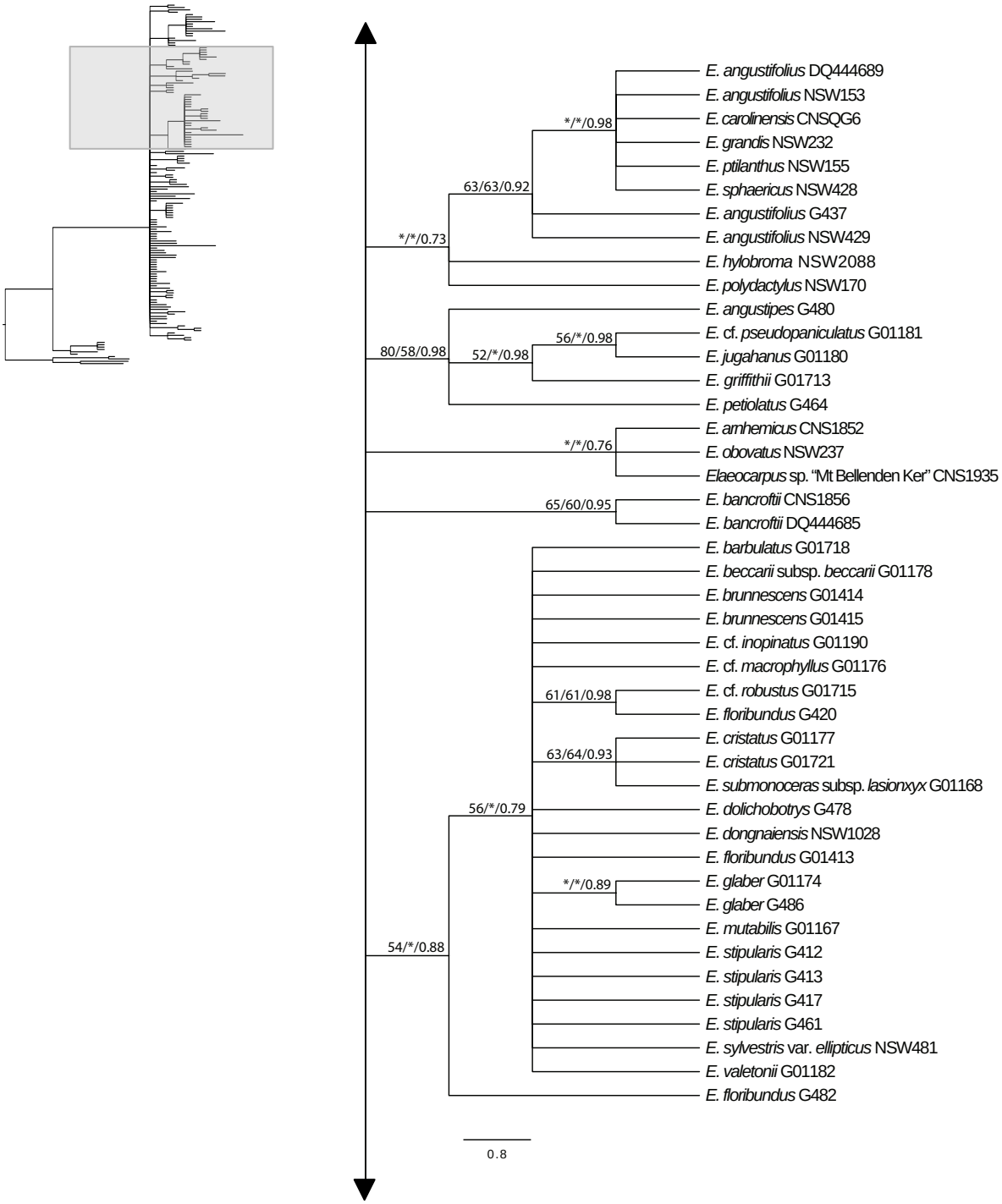


Appendix 2.4 Maximum clade credibility tree from Bayesian analysis of the *trnL-trnF* region. DNA extraction numbers or GenBank numbers are presented after the taxon names. Branch support values are to the left of the nodes in the following order: MP bootstrap/ ML bootstrap/ Bayesian posterior probability. Genus abbreviations are *Aceratium* (Ac), *Crinodendron* (C), *Dubouzetia* (D), *Elaeocarpus* (E), *Peripentadenia* (Pe) and *Sericolea* (Se). * Branch collapses in the MP or ML 50 % majority rule consensus tree. The insert shows branch lengths proportional to the amount of nucleotide change and the section of the tree which has been enlarged.

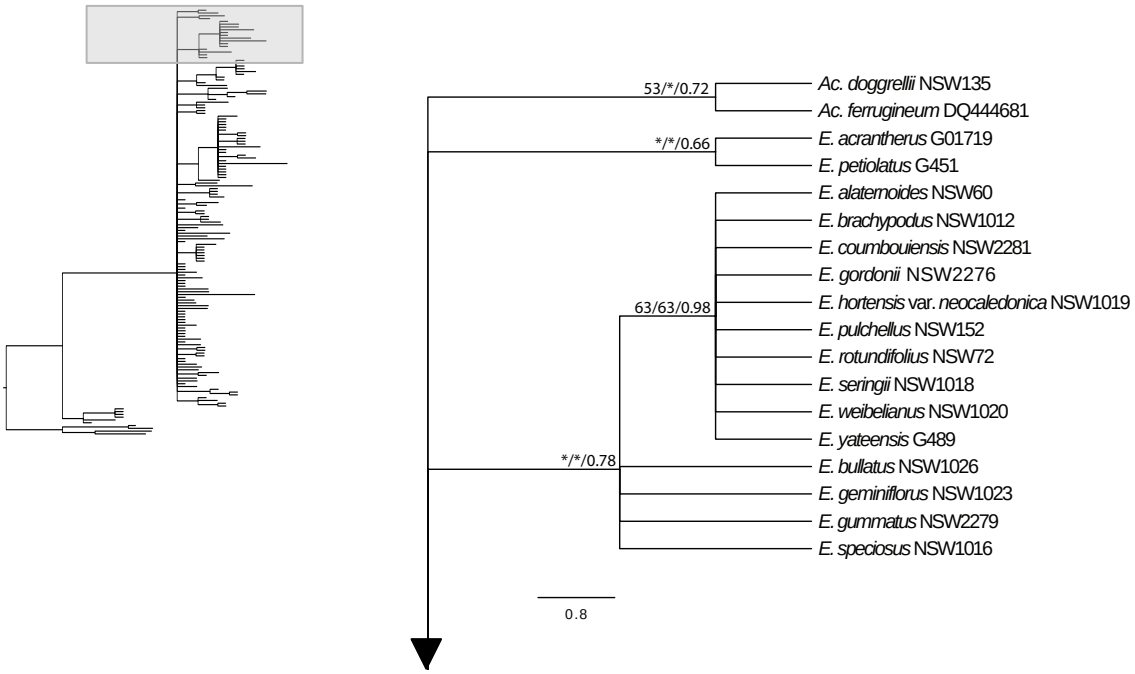
Appendix 2.4 (continued).



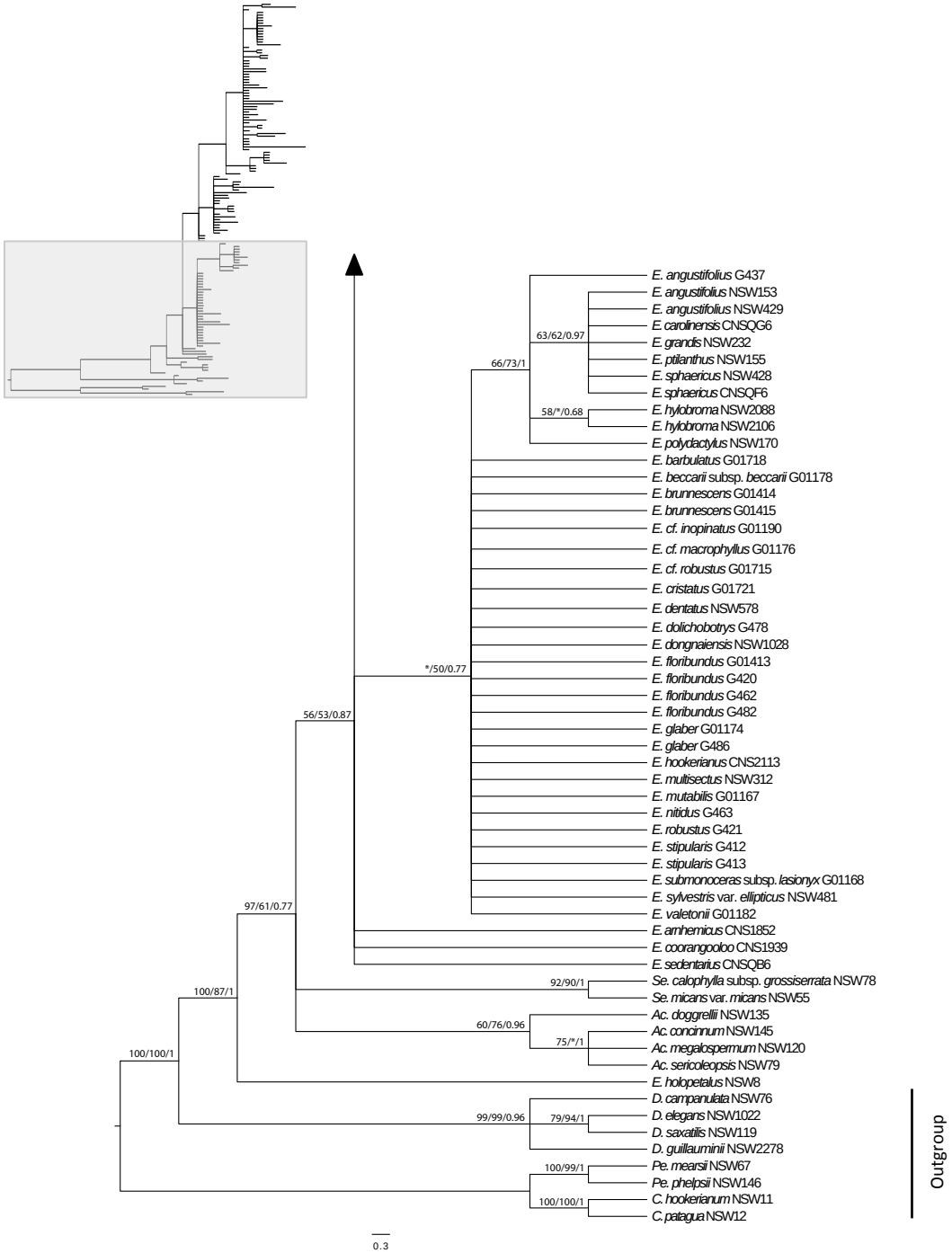
Appendix 2.4 (continued).



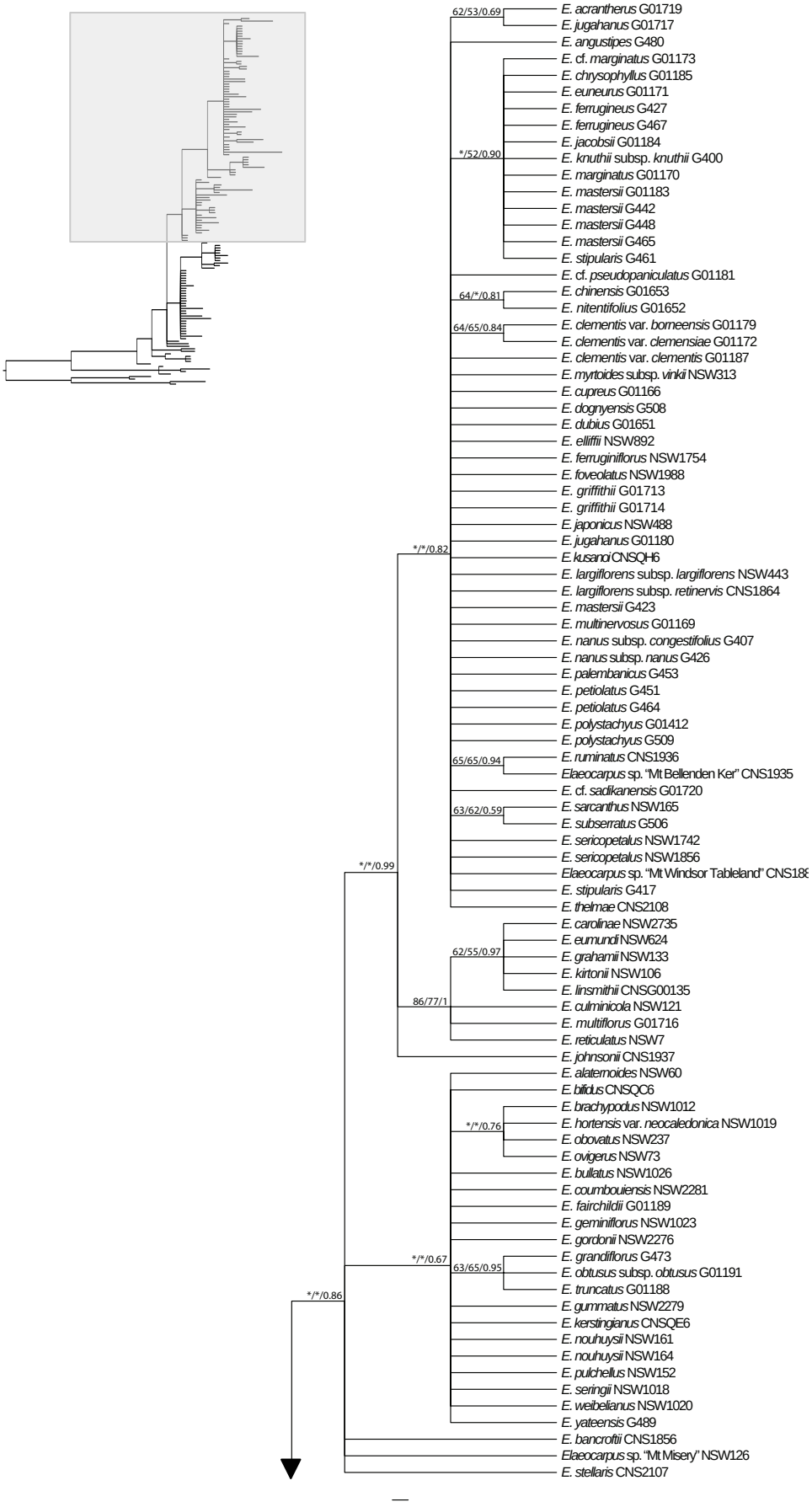
Appendix 2.4 (continued).



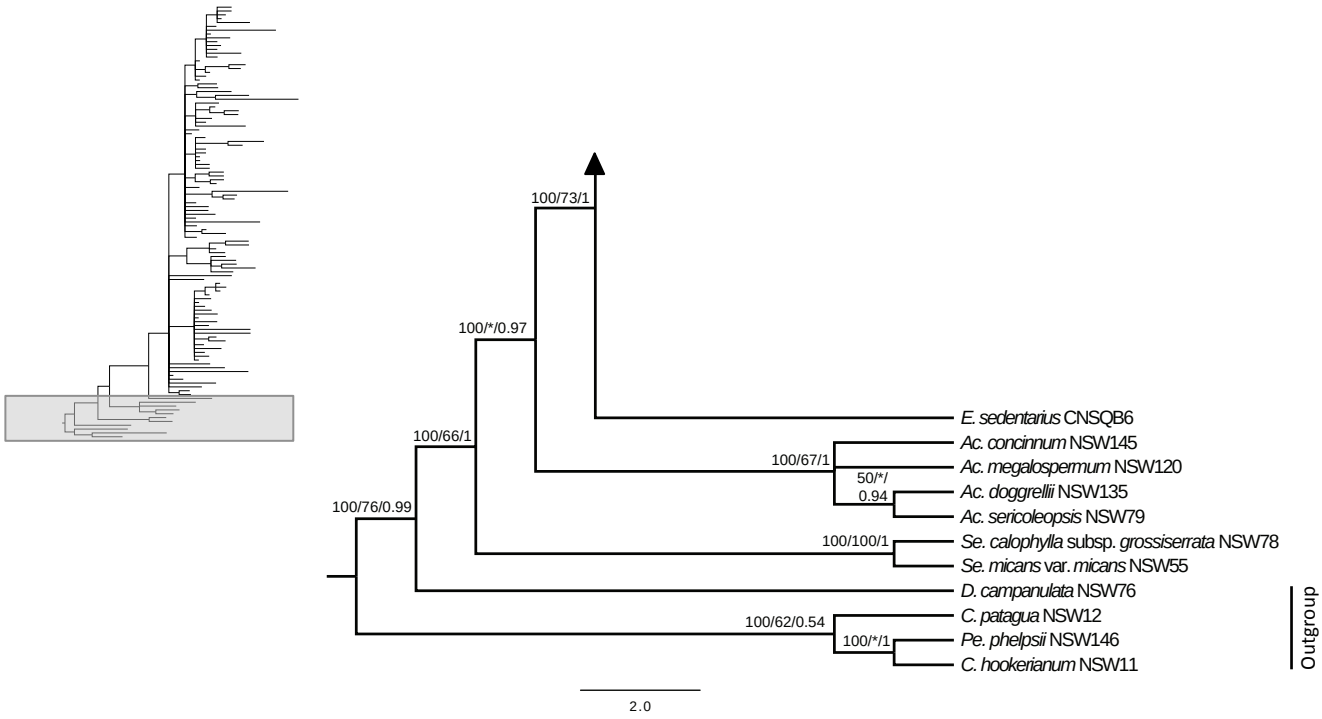
Appendix 2.5 Maximum clade credibility tree from Bayesian analysis of the *trnV-ndhC* region. DNA extraction numbers or GenBank numbers are presented after the taxon names. Branch support values are to the left of the nodes in the following order: MP bootstrap/ ML bootstrap/ Bayesian posterior probability. Genus abbreviations are *Aceratium* (Ac), *Crinodendron* (C), *Dubouzetia* (D), *Elaeocarpus* (E), *Peripentadenia* (Pe) and *Sericolea* (Se). * Branch collapses in the MP or ML 50 % majority rule consensus tree. The insert shows branch lengths proportional to the amount of nucleotide change and the section of the tree which has been enlarged.



Appendix 2.5 (continued).

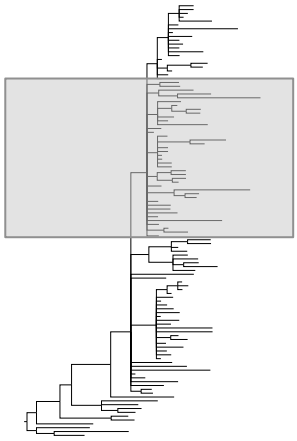


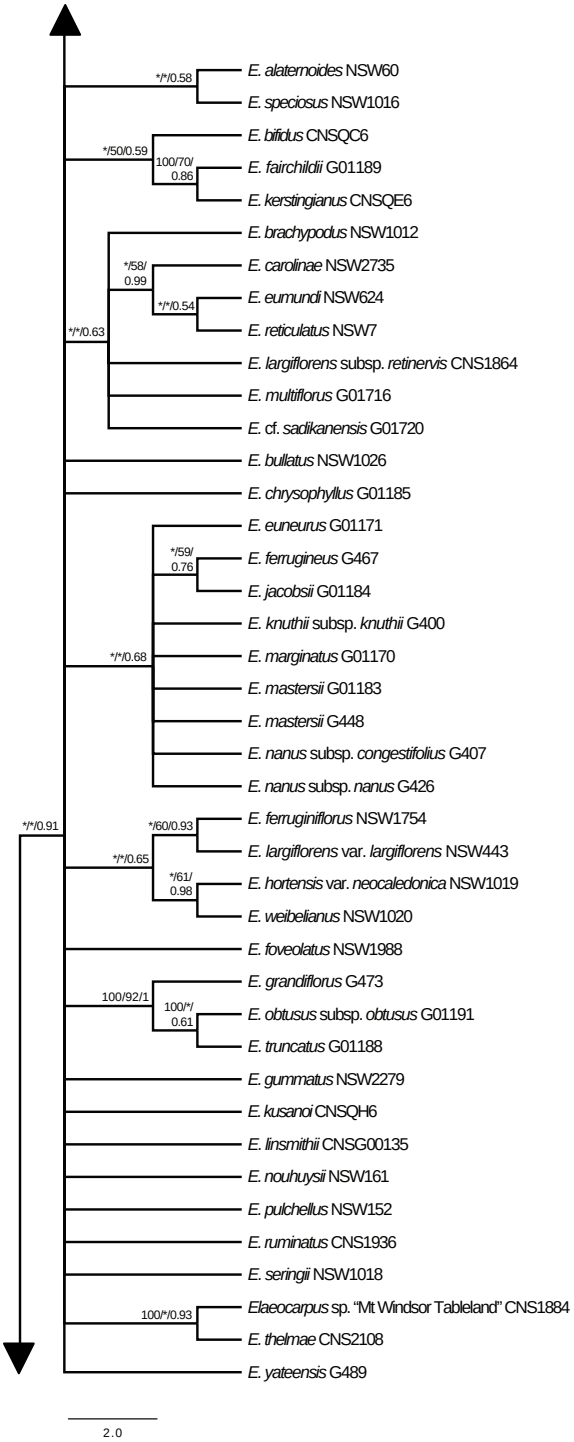
Appendix 2.6 Maximum clade credibility tree from Bayesian analysis of the *trnV-ndhC* region. DNA extraction numbers or GenBank numbers are presented after the taxon names. Branch support values are to the left of the nodes in the following order: MP bootstrap/ ML bootstrap/ Bayesian posterior probability. Genus abbreviations are *Aceratium* (Ac), *Crinodendron* (C), *Dubouzetia* (D), *Elaeocarpus* (E), *Peripentadenia* (Pe) and *Sericolea* (Se). * Branch collapses in the MP or ML 50 % majority rule consensus tree. The insert shows branch lengths proportional to the amount of nucleotide change and the section of the tree which has been enlarged.



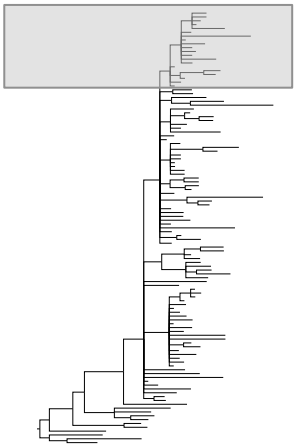
Appendix 2.6 (continued).

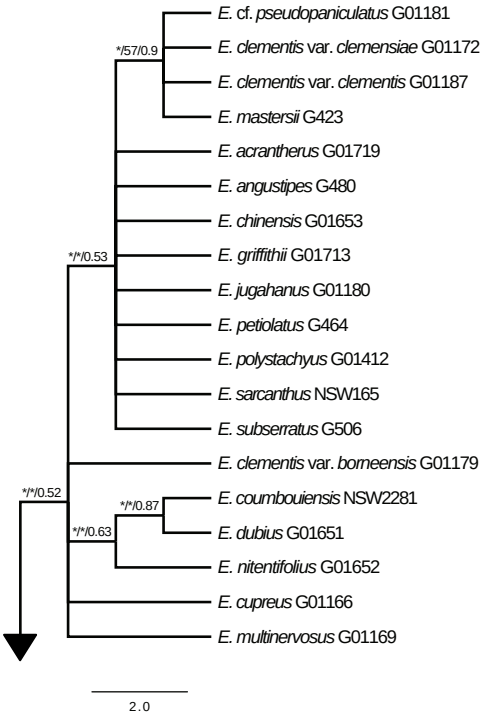
Appendix 2.6 (continued).





Appendix 2.6 (continued).





Appendix 3.1 Species distribution data.

Area codes are Madagascar (A), continental Asia (B), West Malesia (C), Central Malesia (D), East Malesia (E), Australia (F), New Zealand (G) and the Pacific Islands (H).

No.	Taxa	Area	Source
1	*Aceratium concinnum* NSW145	F	GBIF (www.gbif.org, accessed on 11 April 2014)
2	*Aceratium doggrellii* NSW135	F	GBIF (www.gbif.org, accessed on 11 April 2014)
3	*Aceratium megalospermum* NSW120	F	GBIF (www.gbif.org, accessed on 11 April 2014)
4	*Aceratium sericoleopsis* NSW79	F	GBIF (www.gbif.org, accessed on 11 April 2014)
5	*Elaeocarpus acrantherus* G01719	C	Coode (1998)
6	*Elaeocarpus alaternoides* NSW60	H	Tirel (1983)
7	*Elaeocarpus angustifolius* G437	BCDE FH	Tirel (1983); Coode (1984)
8	*Elaeocarpus angustipes* G480	C	Coode (unpublished)
9	*Elaeocarpus arnhemicus* CNS1852	EF	Coode (1984, unpublished)
10	*Elaeocarpus bancroftii* CNS1856	F	Coode (1984)
11	*Elaeocarpus barbulatus* G01718	C	Coode (2001c)
12	*Elaeocarpus beccarii* subsp. *beccarii* G01178	C	Coode (unpublished)
13	*Elaeocarpus bifidus* CNSQC6	H	GBIF (www.gbif.org, accessed on 11 April 2014)
14	*Elaeocarpus brachypodus* NSW1012	H	Tirel (1983)
15	*Elaeocarpus brunnescens* G01414	CD	Coode (unpublished)
16	*Elaeocarpus bullatus* NSW1026	H	Tirel (1983)
17	*Elaeocarpus carolinae* NSW2735	F	Coode (1984)
18	*Elaeocarpus carolinensis* CNSQG6	H	GBIF (www.gbif.org, accessed on 11 April 2014)
19	*Elaeocarpus* cf. *inopinatus* G01190	C	Coode (unpublished)
20	*Elaeocarpus* cf. *macrophyllus* G01176	C	Coode (unpublished)
21	*Elaeocarpus* cf. *pseudopaniculatus* G01181	C	Coode (unpublished)
22	*Elaeocarpus chinensis* G01653	B	Tang & Phengklai (2007)
23	*Elaeocarpus chrysophyllus* G01185	C	Coode (unpublished)
24	*Elaeocarpus clementis* var. *borneensis* G01179	C	Coode (1996c)
25	*Elaeocarpus clementis* var. *clemensiae* G01172	C	Coode (1996c)
26	*Elaeocarpus clementis* var. *clementis* G01187	C	Coode (1996c)
27	*Elaeocarpus coorangooloo* CNS1939	F	Coode (1984)
28	*Elaeocarpus coumbouiensis* NSW2281	H	Tirel (1983)
30	*Elaeocarpus cristatus* G01721	C	Coode (unpublished)
31	*Elaeocarpus culminicola* NSW121	DEF	Coode (1984)
32	*Elaeocarpus cupreus* G01166	C	Coode (1996c)
33	*Elaeocarpus dentatus* NSW578	G	Coode (1984)
34	*Elaeocarpus dognyensis* G508	H	Tirel (1983)
35	*Elaeocarpus dolichobotrys* G478	C	Coode (unpublished)

36	*Elaeocarpus dongnaiensis* NSW1028	B	Gagnepain *et al*. (1907 – 1912)
37	*Elaeocarpus dubius* G01651	B	Tang & Phengklai (2007)
38	*Elaeocarpus elliffii* NSW892	F	Coode (1984)
39	*Elaeocarpus eumundi* NSW624	F	Coode (1984)
40	*Elaeocarpus euneurus* G01171	C	Coode (unpublished)
41	*Elaeocarpus fairchildii* G01189	E	Coode (unpublished)
42	*Elaeocarpus ferrugineus* G467	C	Coode (unpublished)
43	*Elaeocarpus ferruginiflorus* NSW1754	F	Coode (1984)
44	*Elaeocarpus floribundus* G420	CD	Coode (unpublished)
45	*Elaeocarpus foveolatus* NSW1988	F	Coode (1984)
46	*Elaeocarpus geminiflorus* NSW1023	H	Tirel (1983)
47	*Elaeocarpus glaber* G01174	CD	Coode (unpublished)
48	*Elaeocarpus gordonii* NSW2276	H	Tirel (1983)
49	*Elaeocarpus grahamii* NSW133	F	Coode (1984)
50	*Elaeocarpus grandiflorus* G473	CD	Coode (unpublished)
51	*Elaeocarpus griffithii* G01713	C	Coode (unpublished)
52	*Elaeocarpus gummatus* NSW2279	H	Tirel (1983)
53	*Elaeocarpus holopetalus* NSW8	F	Coode (1984)
54	*Elaeocarpus hookerianus* CNS2113	G	Coode (1984)
55	*Elaeocarpus hortensis* var. *neocaledonica* NSW1019	H	Tirel (1983)
56	*Elaeocarpus hylobroma* NSW2088	F	Baba & Crayn (2012)
57	*Elaeocarpus jacobsii* G01184	C	Coode (unpublished)
58	*Elaeocarpus johnsonii* CNS1937	F	Coode (1984)
59	*Elaeocarpus jugahanus* G01180	C	Coode (unpublished)
60	*Elaeocarpus kerstingianus* CNSQE6	E	GBIF (www.gbif.org, accessed on 11 April 2014)
61	*Elaeocarpus kirtonii* NSW106	F	Coode (1984)
62	*Elaeocarpus knuthii* subsp. *knuthii* G400	C	Coode (unpublished)
63	*Elaeocarpus kusanoi* CNSQH6	E	GBIF (www.gbif.org, accessed on 11 April 2014)
64	*Elaeocarpus largiflorens* subsp. *retinervis* CNS1864	F	Coode (1984)
65	*Elaeocarpus largiflorens* var. *largiflorens* NSW443	F	Coode (1984)
66	*Elaeocarpus linsmithii* CNSG00135	F	Coode (1984)
67	*Elaeocarpus marginatus* G01170	C	Coode (unpublished)
68	*Elaeocarpus mastersii* G465	C	Coode (unpublished)
69	*Elaeocarpus multiflorus* G01716	BDE	Coode (unpublished); Tang & Phengklai (2007)
70	*Elaeocarpus multinervosus* G01169	C	Coode (1996c)
71	*Elaeocarpus multisectus* NSW312	E	Coode (unpublished)
72	*Elaeocarpus mutabilis* G01167	C	Coode (unpublished)
29	*Elaeocarpus myrtoides* subsp. *vinkii* NSW313	E	Coode (unpublished)
73	*Elaeocarpus nanus* subsp. *congestifolius* G407	C	Coode (unpublished)
74	*Elaeocarpus nanus* subsp. *nanus* G426	C	Coode (unpublished)
75	*Elaeocarpus nitentifolius* G01652	B	Tang & Phengklai (2007)

76	*Elaeocarpus nitidus* G463	BCD	Coode (2001c)
77	*Elaeocarpus nouhuysii* NSW161	E	Coode (unpublished)
78	*Elaeocarpus obovatus* NSW237	F	Coode (1984)
79	*Elaeocarpus obtusus* subsp. *obtusus* G01191	CD	Coode (unpublished)
80	*Elaeocarpus palembanicus* G453	C	Coode (unpublished)
81	*Elaeocarpus petiolatus* G464	BCD	Tang & Phengklai (2007); Coode (unpublished)
82	*Elaeocarpus polydactylus* NSW170	E	Coode (unpublished)
83	*Elaeocarpus polystachyus* G01412	C	Coode (1996c)
84	*Elaeocarpus ptilanthus* NSW155	E	Coode (unpublished)
85	*Elaeocarpus pulchellus* NSW152	H	Tirel (1983)
86	*Elaeocarpus reticulatus* NSW7	F	Coode (1984)
87	*Elaeocarpus robustus* G421	BC	Coode (1996a, unpublished)
88	*Elaeocarpus rotundifolius* NSW72	H	Tirel (1983)
89	*Elaeocarpus ruminatus* CNS1936	F	Coode (1984)
90	*Elaeocarpus* cf. *sadikanensis* G01720	C	Coode (unpublished)
91	*Elaeocarpus sarcanthus* NSW165	E	Coode (unpublished)
92	*Elaeocarpus sedentarius* CNSQB6	F	Maynard *et al*. (2008)
93	*Elaeocarpus sericopetalus* NSW1856	F	Coode (1984)
94	*Elaeocarpus seringii* NSW1018	H	Tirel (1983)
95	*Elaeocarpus* sp. Mt Bellenden Ker CNS1935	F	Coode (1984)
96	*Elaeocarpus* sp. Mt Misery NSW126	F	Coode (1984)
97	*Elaeocarpus* sp. Mt Windsor Tableland CNS1884	F	Coode (1984)
98	*Elaeocarpus speciosus* NSW1016	H	Tirel (1983)
99	*Elaeocarpus stellaris* CNS2107	F	Coode (1984)
100	*Elaeocarpus stipularis* G417	BCD	Coode (2001c)
101	*Elaeocarpus submonoceras* subsp. *lasionyx* G01168	CD	Coode (unpublished)
102	*Elaeocarpus subserratus* G506	A	Tirel (1985)
103	*Elaeocarpus sylvestris* var. *ellipticus* NSW481	B	Tang & Phengklai (2007)
104	*Elaeocarpus thelmae* CNS2108	F	Coode (1984)
105	*Elaeocarpus truncatus* G01188	C	Coode (unpublished)
106	*Elaeocarpus valetonii* G01182	C	Coode (unpublished)
107	*Elaeocarpus weibelianus* NSW1020	H	Tirel (1983)
108	*Elaeocarpus yateensis* G489	H	Tirel (1983)
109	*Sericolea calophylla* subsp. *grossiserrata* NSW78	E	Coode (1978)
110	*Sericolea micans* var. *micans* NSW55	E	Coode (1978)

Appendix 4.1 Herbarium specimens examined for morphometric analysis. Species: EC (*Elaeocarpus cupreus*), ECb (*Elaeocarpus clementis* var. *borneensis*), ECs (*Elaeocarpus clementis* var. *clemensiae*), ECc (*Elaeocarpus clementis* var. *clementis*), ECk (*Elaeocarpus clementis* var. *kostermansii*), EI (*E. integripetalus*), EM (*Elaeocarpus multinervosus*), EP (*Elaeocarpus polyanthus*) and ES (*Elaeocarpus polystachyus*); Herbaria: BO (Herbarium Bogoriense, Bogor), BRUN (Brunei Forestry Centre), K (Royal Botanic Gardens, Kew), KEP (Forest Research Institute Malaysia, Kepong), L (Nationaal Herbarium Nederland, Leiden University Branch, Leiden), SAN (Forest Research Centre, Sandakan), SAR (Department of Forestry, Kuching), SING (Singapore Herbarium); Characters scored: [a]specimen scored for vegetative and floral characters, [b]specimen scored for vegetative and fruit characters; [c]specimen scored for vegetative characters only; Floristic provinces: Borneo (B), Malay Peninsula (MP), Sumatra (S); Floristic subprovinces: Lowland Flora in Borneo (B), East Coast Sabah Floristic Subprovince (ECSS), *Kerangas* Forest (K), Mountain Flora in Borneo (MFB), Mountain Flora in Malay Peninsula (MFMP), Lowland Flora in Malay Peninsula (MP), Seasonally Flooded Riverside (R), Riau Pocket (RP), Sumatra (S), Seasonal Asiatic-Australasian Intrusion in North Borneo (SAAI), Seasonal Asiatic Intrusion (SAI); NA information not available.

Taxa	Collection Number	Collector	Collection Date	Herbarium	Province	Subprovince	Locality
EC	Native Collector 1702	Native Collector	NA	K[a] (Isotype)	B	NA	Borneo. Sarawak.
EC	Haviland *s.n.* (K000708112)	Haviland	10 Oct 1889	K[a] (Holotype)	B	NA	Borneo. Sarawak, Penkulu Ampat.
EC	Jaheri 645	Jaheri	1896 – 1897	BO[b]	B	NA	Borneo. Kalimantan, Sungai Magne, expedition, Nieuwenhuis.
EC	BRUN 553	Ashton PS	17 Sep 1957	SING[a]	B	B	Borneo. Brunei, Andulau Forest Reserve.

EC	Kostermans 9507	Kostermans A	24 May 1954	BO[a]	B	B	Borneo. Sabah (East Borneo), Bangko Tdg., mouth of Mahakam river.
EC	FRI 14776	Suppiah T	18 Feb 1971	KEP[c]	MP	MP	Malay Peninsula. Pahang, 6 mile south of Kampung Aur.
EC	S 24639	Sibat L	25 Jun 1966	KEP[a]	B	RP	Borneo. Sarawak, 4th division, Bintulu, Nyabau Catchment area.
EC	Clemens 20820	Clemens J & MS	24 - 19 Oct 1929	K[a] (Isotype)	B	RP	Borneo. Sarawak, Kuching, Mount Matang.
EC	de Wilde 21152	de Wilde & de Wilde-Duyfjes	1 Jul 1991	K[a]	S	S	Sumatra. Sekundur Forest Reserve, east side of Gunung Leuser National Park, Langket.
EC	S 29171	Othman H	13 Sep 1969	KEP[a], SAR[a]	B	B	Borneo. Sarawak, Kapit, Ulu Balleh, Sungai Mengiong.
EC	S 45023	Banyeng N & Paie I	14 Oct 1982	KEP[a], SAN[a], SAR[a]	B	B	Borneo. Sarawak, base camp to Bukit Sadok.
EC	Meijer 2506	Meijer W	13 Dec 1953	BO[a,b]	B	ECSS	Borneo. Kalimantan, Tarakan, Sesanip.
EC	Kostermans 7061	Kostermans A	28 May 1952	BO[a]	B	B	Borneo. Sarawak, Tdg. Bangko region, Sungai Mahakam.
EC	Anderson 8468	Anderson JAR	25 Jul 1957	BO[a], SAR[a]	B	RP	Borneo. Sarawak, Kuching, Selang Forest Reserve.
EC	Kostermans 9714	Kostermans A	13 Jul 1954	BO[a]	B	B	Borneo. Kalimantan, Balikpapan, Sungai Mentawir region.
EC	S 27224	Paie I	25 Aug 1968	BO[a]	B	RP	Borneo. Sarawak, Bintulu, Ulu Segan.
EC	Amdjah 1070	Amdjah	Dec 1912	BO[b]	B	B	Borneo. North Kalimantan, Samenggaris river.
EC	Wiriadinata 651	Wiriadinata H	28 Jun 1975	BO[b]	B	B	Borneo. East Kalimantan, Longbagun, Camp

							Tikah.
EC	W 695	Ambri & Arifin	14 Mar 1991	BO[a]	B	B	Borneo. East Kalimantan, Wanariset research area.
EC	FRI 15900	Whitmore TC	18 Feb 1971	KEP[b], SING[b]	MP	MP	Malay Peninsula. Pahang, Kampung Aur.
EC	SFN 40932	Sinclair J & Kiah S	22 Sep 1955	SING[a]	MP	RP	Malay Peninsula. Terengganu, 27¾ mile Jerangau road.
EC	SFN 32414	Kiah S	17 Mar 1937	BO[b], SING[b]	MP	MP	Malay Peninsula. Johor, Sungai Kayu.
EC	Meijer 2568	Meijer W	16 Dec 1953	SING[a]	B	ECSS	Borneo. Kalimantan, Tarakan, Sesanip-Djuata.
EC	BRUN 674	Ashton PS	27 Sep 1957	BRUN[a], SAR[a], SING[a]	B	B	Borneo. Brunei, Bukit Teraja.
EC	Haviland 2230	Haviland GD	24 Feb 1893	SING[a]	B	RP	Borneo. Sarawak, Kuching.
EC	Haviland 49	Haviland GD	NA	SING[a]	B	RP	Borneo. Sarawak, Pengkalan Ampat.
EC	S 4813	Rehal Dollah	22 Nov 1956	SAR[a], SING[a]	B	NA	Borneo. Sarawak, Selang.
EC	KEP 108999	Suppiah T	20 Jun 1967	SING[b]	MP	MP	Malay Peninsula. Pahang, Lepar Forest Reserve.
EC	S 45654	Dayang Awa & Paie I	16 Apr 1983	KEP[a], SAN[a], SAR[a]	B	RP	Borneo. Sarawak, Tebakang area, Bukit Alak.
ECb	Creagh *s.n.*	Creagh	NA	K[b] (Holotype)	B	NA	Borneo. Sabah, east coast.
ECb	Kostermans 4502	Kostermans A	30 Aug 1950	BO[c]	B	B	Borneo. Kalimantan, north Balikpapan Sungai Wain Region.
ECb	ARs & R 396	Ruskandi A & Rugayah	17 Feb 2001	BO[c]	B	B	Borneo. East Kalimantan, Bukit Bangkirai.

ECb	Kostermans 6328	Kostermans A	11 Apr 1952	BO[c]	B	B	Borneo. Kalimantan, W Samarinda, Loa Djanan.
ECb	Kostermans 42/35	Kostermans A	9 Sep 1950	BO[c]	B	B	Borneo. Kalimantan, north Balikpapan, Sungai Wain region.
ECb	Kloss 19195	Bodeu-Kloss C	25 Aug 1927	SING[c]	B	ECSS	Borneo. Sabah, Sandakan, Bettotau?.
ECb	SAN 107702	Leopold & Sigin	4 Dec 1984	KEP[c]	B	ECSS	Borneo. Sabah, Sandakan, Ulu Sungai Kun-Kun.
ECb	SAN 88887	Madani L	12 Aug 1978	KEP[c]	B	ECSS	Borneo. Sabah, Sandakan, Mile 112 Telupid Ranau Road.
ECb	Ramos 1373	Ramos M	Sep - Dec 1920	BO[a]	B	ECSS	Borneo. Sabah, Sandakan and vacinity.
ECb	SAN 73830	Leopold, Gaty & Dewo	9 Sep 1971	SING[c]	B	SAAI	Borneo. Sabah, Ranau, Wonod gravel pit.
ECb	SAN 114344	Amin & Jarius	14 Aug 1986	KEP[a]	B	SAAI	Borneo. Sabah, Kota Belud, Kinasopian.
ECb	SAN 85043	Cockburn PF	26 Aug 1976	KEP[a]	B	SAAI	Borneo. Sabah, Lahad Datu, true left bank of Sungai Danum Segama.
ECb	SAN 116230	Amin & Jarius	23 Sep 1986	KEP[a], SAN[a]	B	ECSS	Borneo. Sabah, Ranau, Marambai.
ECb	SAN 79440	Aban G & Saikeh	20 Mar 1974	KEP[b], SAR[b]	B	ECSS	Borneo. Sabah, Sandakan, 8 mile Telupid.
ECb	SAN 108403	Maidil *et al*.	14 Mar 1985	KEP[b], SAN[b], SAR[b]	B	ECSS	Borneo. Sabah, Telupid, Bukit Tawai Keramuak.
ECb	SAN 79420	Aban G & Saikeh	19 Mar 1974	KEP[b], SAN[b], SAR[b]	B	ECSS	Borneo. Sabah, Sandakan, Telupid, Kampung Wonod.
ECb	SAN 149014	Sugau JB	18 Nov 2006	KEP[b], SAN[b]	B	MFB	Borneo. Sabah, Pensiangan, Sapulut Forest Reserve.
ECb	SAN 67044	Pikko M	25 Sep 1984	KEP[a], SAN[a]	B	ECSS	Borneo. Sabah, Labuk Sugut, Sungai Ulu Logan.

ECb	SAN 26632	Singh J	18 Oct 1962	KEP[a], SAN[a]	B	K	Borneo. Sabah, Sandakan, Kebuan China Sibuga Forest Reserve.
ECb	SAN 32610	Jawanting A	19 Nov 1962	SAN[b]	B	MFB	Borneo. Sabah, Sandakan, KF Loong's logging area, Camp Mamahat.
ECb	PK 2739	Kessler PJA & Arifin Z	22 Feb 2000	SAN[a], BO[a]	B	B	Borneo. South Kalimantan, Tabalong, Tanjung.
ECb	Endert 4998	Endert FH	16 Nov 1925	BO[b]	B	B	Borneo. Kalimantan, west Kutei.
ECb	SAN 62060	Lasan P	17 Oct 1967	SAN[a], BO[a]	B	ECSS	Borneo. Sabah, Sandakan, Jalan Kabili Sepilok Forest Reserve.
ECb	Ramos 1479	Ramos M	Sep–Dec 1920	BO[a]	B	ECSS	Borneo. Sabah, Sandakan.
ECb	Agama 9249	Agama J	14 Sep 1938	SING[a]	B	ECSS	Borneo. Sabah, Sandakan, Kabili Forest Reserve.
ECb	SAN 79490	Dewol S	22 May 1978	SAN[a], SING[a,b]	B	ECSS	Borneo. Sabah, Sandakan, 58 mile Telupid.
ECb	SAN 44632	Lajengah JK	5 Nov 1965	SAN[a], SAR[a], SING[a]	B	MFB	Borneo. Sabah, Ranau, Lohan Forest Reserve.
ECb	Elmer 20280	Elmer ADE	Oct–Dec 1921	BO[a], SING[a]	B	ECSS	Borneo. Sabah, Sandakan.
ECb	Onggib 10581	Onggib	27 Nov 1938	KEP[b], SING[b]	B	ECSS	Borneo. Sabah, Sandakan, Kabili Sepilok Forest Reserve.
ECb	A 1612	Kadir	7 Mar 1950	BO[b], SAN[b], SING[b]	B	ECSS	Borneo. Sabah, Sandakan, Kabili Sepilok Forest Reserve.
ECb	A 2654	Agama J	30 Sep 1949	BO[a], SAN[a], SING[a]	B	ECSS	Borneo. Sabah, Sandakan, Kabili Sepilok Forest Reserve.
ECb	SFN 26475	Carr CE	10 Mar 1933	SING[b]	B	MFB	Borneo. Sabah, Sungai Mahandui.

ECb	A 914	Kadir	5 Dec 1948	BO[b], SING[b]	B	ECSS	Borneo. Sabah, Sandakan, Sepilok, Sungai Arang.
ECb	SAN 50504	Ahmad	24 Sep 1966	SAN[a], SING[a]	B	SAAI	Borneo. Sabah, Beaufort, Beaufort Hill.
ECb	Wood A 2979	Wood GHS	6 Apr 1954	KEP[b], SING[b]	B	ECSS	Borneo. Sabah, Sandakan, Kabili Sepilok Forest Reserve.
ECb	SAN 16936	Wood GHS	24 Feb 1956	KEP[b], SING[b]	B	ECSS	Borneo. Sabah, Sandakan, Kabili Sepilok Forest Reserve.
ECb	A 1856	Patrick	9 Dec 1951	SAN[b], SING[b]	B	ECSS	Borneo. Sabah, Sandakan, Kabili Sepilok Forest Reserve.
ECb	Clemens 32039	Clemens J & Clemens MS	10 Mar 1933	K[b]	B	MFB	Borneo. Sabah, Kinabalu, Penibukan.
ECc	Church 546	Church AC *et al*.	4 Nov 1993	SING[c]	B	B	Borneo. West Kalimantan, Sintang, Bukit Baka National Park, east of camp along bank of Sungai Ella and environs.
ECc	Mahyar 1160	Mahyar UW *et al*.	26 Apr 1994	SING[a]	B	B	Borneo. Central Kalimantan, Sintang HPH km 84, along new subsidiary logging road to southeast.
ECc	SAN 127888	Sumbing J	23 Aug 1989	SAN[a]	B	MFB	Borneo. Sunsuron KM 54 Jalan Tambunan/Penampang.
ECc	S 36823	Chai P *et al*.	17 Sep 1975	KEP[c]	B	RP	Borneo. Sarawak, 1st/2nd Division boundary, Gunung Buri, Ulu Simunjan, Ulu Sungai Empili.
ECc	Richards 1489	Richards	27 Aug 1932	K[a] (Holotype)	B	RP	Borneo. Sarawak, 4th Division, Mount Dulit trail, near Long Kapa.
ECc	SAN 69911	Patrick *et al.*	5 Oct 1984	KEP[a,b], SAN[b], SAR[a,b]	B	SAAI	Borneo. Sabah, Kota Kinabatangan, Koyah, Ladang Pendirosa.
ECc	SAN 91354	Fedilis K & Sumbing	10 Nov 1979	KEP[a], SAN[a],	B	ECSS	Borneo. Sabah, Kalabakan, 13 mile Hap Seng

		J		SAR[a]			logged area.
ECc	A 3854	Charington C	14 Feb 1955	KEP[b]	B	ECSS	Borneo. Sabah, Sandakan, Kabili Sepilok Forest Reserve.
ECc	S 41357	Othman *et al.*	11 Nov 1979	KEP[b], SAN[b], SAR[b]	B	B	Borneo. Sarawak, Kapit, Ulu Balleh, Mengiong.
ECc	S 45972	Yii PC & Banyeng N	30 Apr 1983	KEP[a], SAN[a], SAR[a]	B	K	Borneo. Sarawak, 17 mile Bau-Lundu road, Gunung Undan.
ECc	S 46002	Yii PC & Jegong	2 May 1983	KEP[b], SAN[b], SAR[b]	B	K	Borneo. Sarawak, 27 km Bau-Lundu road, Gunung Raya.
ECc	SAN 127734	Sumbing J	14 Aug 1989	SAN[a]	B	MFB	Borneo. Sabah, Tambunan-Penampang road, 48 km Togudon-Tungol.
ECc	Moulton 17	Moulton JC	30 Oct 1914	SING[a]	B	RP	Borneo. Sarawak, Baram, Lio-Matu.
ECc	Mahyar 1152	Mahyar UW *et al.*	26 Apr 1994	BO[b], SING[b]	B	B	Borneo. Central Kalimantan, Sintang HPH 84-87 km.
ECc	Mahyar 798	Mahyar UW *et al.*	10 Apr 1994	BO[b], SING[b]	B	B	Borneo. Central Kalimantan, Sintang HPH 70 km.
ECc	Church 596	Church AC *et al.*	6 Nov 1993	BO[b], KEP[a], SING[b]	B	B	Borneo. West Kalimantan, Sintang, Bukit Baka National Park.
ECk	Kostermans 13101	Kostermans A	16 Sep 1956	L[b]	B	B	Borneo. Kalimantan, west Kutei, Gunung Palimasan.
ECk	Kostermans 13040	Kostermans A	14 Sep 1956	L[a,b]	B	B	Borneo. Kalimantan, west Kutei, Gunung Palimasan.
ECs	S 21889	Sibat L	22 Jul 1964	KEP[a]	B	NA	Borneo. Sarawak, Anap, Bukit Mersing.
ECs	Clemens 51312	Clemens J & MS	1 Jan 1934	K[a] (Isotype)	B	MFB	Borneo. Sabah, Kinabalu, Bungal.

ECs	BRUN 130	Ashton PS	6 Jul 1957	SING[a]	B	R	Borneo. Brunei, River Ingei.
ECs	BRUN 237	Ashton PS	12 Jul 1957	SING[a]	B	R	Borneo. Brunei, Ulu Belait at K. Penipir.
ECs	Coode 6927	Coode MJE *et al*.	5 Dec 1991	SING[c]	B	RP	Borneo. Brunei, Belait, Bukit Teraja.
ECs	Coode 6576	Coode MJE *et al*.	6 Apr 1990	SING[c]	B	RP	Borneo. Brunei, Tempurong, Temburong River, just upstream from Wong Nguan rapids.
ECs	Coode 6906	Coode MJE *et al*.	4 Dec 1991	KEP[c]	B	RP	Borneo. Brunei, Belait, Bukit Teraja, path above Rest house.
ECs	Simpson 2122	Simpson DA & Marsh M	18 Oct 1991	KEP[c]	B	RP	Borneo. Brunei, Belait, Labi, Bukit Teraja, ridge running north from summit.
ECs	BRUN 16462	Hussain O *et al*.	22 Dec 1994	BRUN[a]	B	RP	Borneo. Brunei, Tutong, Tanjong Maya, Bukit Udal, Bukit Kukup.
ECs	BRUN 17082	Ariffin K *et al*.	7 Oct 1995	BRUN[a]	B	RP	Borneo. Brunei, Belait, Sungai Liang, Andulau Forest Reserve, compartment 5, *ex-situ* area.
ECs	BRUN 21725	Ariffin K *et al*.	13 Jul 2006	BO[a], BRUN[a]	B	RP	Borneo. Brunei, Tutong, Ukong, Kampong Long Mayan, Jalan Kecil Long Mayan-Merangking (between border on Tutong & Belait District).
ECs	Og-B 27	Ogata K *et al*.	12 Nov 1995	BRUN[a]	B	RP	Borneo. Brunei, Belait, Labi, Bukit Teraja.
ECs	BRUN 17049	Ariffin K *et al.*	23 Aug 1995	BRUN[a], SAN[a], SING[a]	B	R	Borneo. Brunei, Belait, Melilas, Sungai Belait.
ECs	SAN 115403	Amin A	13 Apr 1988	KEP[a], SAN[a], SAR[a]	B	R	Borneo. Sabah, Melalia Fr. Sipitang.
ECs	SAN 106923	Fidilis & Omar	12 Oct 1984	KEP[a], SAR[a]	B	B	Borneo. Sabah, Sapulut, Sungai Saburan.

ECs	SAN 124307	Bousi J *et al.*	20 Aug 1989	KEP[a], SAN[a]	B	ECSS	Borneo. Sabah, Sandakan, Segaliud Lokan Forest Reserve.
ECs	S 40310	George R	20 Sep 1978	KEP[a], SAR[a]	B	RP	Borneo. Sarawak, Lambir National Park.
ECs	S 56617	Othman I *et al.*	21 Mar 1989	KEP[a], SAN[a]	B	RP	Borneo. Sarawak, Lundu, Bedaun.
ECs	Coode 7777	Coode MJE *et al.*	9 Jan 1994	BRUN[b], KEP[b]	B	RP	Borneo. Brunei, Belait, Labi, Sungai Rampayoh.
ECs	SAN 67095	Amin A *et al.*	18 Sep 1984	KEP[a], SAN[a]	B	ECSS	Borneo. Sabah, Labuk Sugud, Sungai Tongod.
ECs	BRUN 17086	Ariffin K	11 Oct 1995	BRUN[a], KEP[a]	B	RP	Borneo. Brunei, Belait, Andulau Forest Reserve.
ECs	SAN 79201	Minjulu F	22 Oct 1975	KEP[b], SAN[b]	B	ECSS	Borneo. Sabah, Tawau, Ulu Kuamut Hap Seng logging area.
ECs	S 38905	Martin PJ	5 May 1977	KEP[b]	B	MFB	Borneo. Sarawak, Gunung Mulu National Park.
ECs	S 38206	Martin PJ	5 Oct 1976	KEP[b], SAR[b]	B	MFB	Borneo. Sarawak, Gunung Mulu National Park.
ECs	BRUN 20261	Watu A *et al.*	15 Mar 2003	BRUN[b]	B	RP	Borneo. Brunei, Belait, Bukit Sawat, Kampung Laidlakang.
ECs	BRUN 17151	SATC Students (BND 1992 – 96) *et al.*	17 Oct 1995	BRUN[a]	B	RP	Borneo. Brunei, Belait, Andulau Forest Reserve.
ECs	WKM 298	Wong KM	22 Jul 1988	BRUN[a,b], SAN[a]	B	RP	Borneo. Brunei, Temburong, Kuala Belalong.
ECs	WKM 2005	Wong KM	18 Sep 1990	BRUN[a]	B	RP	Borneo. Brunei, Belait, Sungai Liang Arboretum.
ECs	BRUN 18631	Ariffin K *et al.*	26 Jul 1997	BRUN[b]	B	RP	Borneo. Brunei, Tutong, Ukong, Kampong Rampau.
ECs	Og-B 20	Ogata K *et al.*	12 Nov 1995	BRUN[a]	B	RP	Borneo. Brunei, Labi, Bukit Teraja.

ECs	Og-B 513	Ogata K *et al.*	26 Jul 1997	BRUN[a,b]	B	RP	Borneo. Brunei, Tutong, Ukong, Kampong Rampau.
ECs	WKM 3171	Wong KM *et al.*	16 Feb 2011	BRUN[b]	B	RP	Borneo. Brunei, Labi, Bukit Teraja.
ECs	BRUN 17047	Ariffin K	23 Aug 1995	BRUN[b], SAN[b], SING[b]	B	K	Borneo. Brunei, Belait, Ulu Sungai Belait.
ECs	CWL 413	Chew WL	17 Jun 1962	SING[b]	B	MFB	Borneo. Sarawak, Gunung Mulu National Park, Gunung Mulu.
ECs	SAN 72626	Leopold M. & Patrick	20 Feb 1971	SAN[b], SING[b]	B	ECSS	Borneo. Sabah, Sandakan, Segaliud Lokan Forest Reserve.
ECs	CSC 4614	Chin et al.	12 Nov 1995	SING[b]	B	RP	Borneo. Brunei, Labi, Bukit Teraja.
ECs	SAN 88367	Fedilis K & Sumbing J	18 Mar 1978	SAN[b], SING[b]	B	ECSS	Borneo. Sabah, Tawau, Imbak, Impong road.
ECs	BRUN 15564	Niga N *et al.*	17 Nov 1994	BRUN[a], SAN[a], SING[a]	B	B	Borneo. Brunei, Belait, Kampung Merangking.
ECs	Coode 6883	Coode MJE *et al.*	3 Dec 1991	BRUN[a], SING[a]	B	RP	Borneo. Brunei, Labi, Bukit Teraja.
ECs	Dransfield 6868	Dransfield J *et al.*	11 Nov 1990	SAN[a]	B	K	Borneo. Brunei, Labi, Bukit Teraja.
ECs	Coode 6918	Coode MJE *et al.*	5 Dec 1991	BRUN[a], SAN[a], SAR[a], SING[a]	B	RP	Borneo. Brunei, Labi, Bukit Teraja.
ECs	Coode 7123	Coode MJE *et al.*	22 Dec 1991	BRUN[b], KEP[b], SING[b]	B	RP	Borneo. Brunei, Belait, Sungai Liang, Sungai Lumut.
ECs	Coode 7790	Coode MJE *et al.*	10 Jan 1994	BRUN[b], KEP[b]	B	RP	Borneo. Brunei, Belait, Labi, Sungai Rampayoh.
EI	Teysmann 1088 H.B.	Teysmann JE		BO[a] (Holotype)	S	S	Sumatra. West Sumatra, Payakumbuh [Paya Kombo].

EM	Beccari 2698	Beccari O	1865 - 1868	K[b] (Isotype)	B	NA	Borneo. Sarawak.
EM	S 39965	Lee B	2 Sep 1978	KEP[c]	B	B	Borneo. Sarawak, 7th Division, Belaga, Linau-Balui, Sungai Jelini.
EM	S 33193	Chai P *et al*.	16 Sep 1973	KEP[c]	B	B	Borneo. Sarawak, 7th Division, Ulu Kapit, Pelagus Protected Forest.
EM	S 42883	George R *et al*.	13 Sep 1980	KEP[c]	B	B	Borneo. Sarawak, 5th Division, Limbang, Ulu Medamit, Sungai Ensungei, Tanjung Long Amok.
EM	S 19056	Asah U	3 Apr 1964	KEP[c]	B	B	Borneo. Sarawak, Ulu Mujong above north Temalad.
EM	SAN 20871	Charington C	3 Apr 1960	SING[c]	B	ECSS	Borneo. Sabah, Sandakan, Sepilok Forest Reserve, S.P. 17, compartment 8.
EM	Dan 3012	Dan B	3 Jun 1961	SING[a]	B	K	Borneo. Sarawak, Miri, Lambir Hills Forest Reserve.
EM	S 46107	Paie I	11 May 1983	KEP[a]	B	K	Borneo. Sarawak, 1st Division, Lundu, Kampung Rasau, Sungai Boeng.
EM	S 24825	Sibat L	7 Apr 1966	BO[c]	B	RP	Borneo. Sarawak, Miri, 6.5 miles Bakam Road.
EM	SAN 80466	Dewol S & Talib	20 Mar 1976	SAN[a]	B	SAAI	Borneo. Sabah, Papar, Ulu Kimanis.
EM	S 19917	Othman H	13 Apr 1964	KEP[a], SAN[a], SAR[a]	B	B	Borneo. Sarawak, Kapit, Balleh, Ulu Mujong, north Semperaja.
EM	S 47126	Mohtar A *et al.*	23 Oct 1983	KEP[b]	B	RP	Borneo. Sarawak, Lambir National Park, Sungai Bakam.

EM	S 39098	Paie I	15 Jul 1977	KEP[b], SAN[b], SAR[b]	B	RP	Borneo. Sarawak, Miri, Niah, Ulu Sungai Sah.
EM	S 24088	Morshidi A	3 Mar 1966	KEP[b], SAN[b], SAR[b]	B	RP	Borneo. Sarawak, Miri, Lambir National Park.
EM	Coode 7013	Coode MJE *et al.*	13 Dec 1991	BRUN[a]	B	RP	Borneo. Brunei, Tutong, Rambai, Bukit Bahak.
EM	BRUN 22174	Mohd Yusop AR *et al.*	17 Jan 2008	BRUN[b]	B	RP	Borneo. Brunei, Belait, Labi, Labi Hill Forest Reserve, Bukit Telingan.
EM	BRUN 21898	Watu A *et al*.	5 Feb 2008	BRUN[b]	B	RP	Borneo. Brunei, Belait, Labi, Labi Hill Forest Reserve, Ulu Sungai Bauali.
EM	BRUN 18245	Ariffin K *et al*.	10 Apr 1997	BRUN[b]	B	RP	Borneo. Brunei, Belait, Merangking, Bukit Sawat.
EM	BRUN 19113	Noor Azam AR *et al.*	27 Oct 1998	BRUN[a]	B	RP	Borneo. Brunei, Belait, Sukang, Buau, Biadong-Buau road.
EM	BRUN 18203	Said IM	31 May 1997	BRUN[a]	B	RP	Borneo. Brunei, Temburong, Amo, Kampung Batang Duri Park.
EM	Haviland 908	Haviland GD	13 Dec 1892	K[b]	B	RP	Borneo. Sarawak.
EM	Synge 490	Synge PM	16 Sep 1932	SING[a]	B	K	Borneo. Sarawak, Mount Dulit, Ulu Koyan.
EM	SFN 35741	Daud & Tachun	11 Aug 1938	SING[a]	B	NA	Borneo. Sarawak, Sungai Sebunut.
EM	S 24837	Benang B	7 Apr 1966	SAN[b], SAR[b], SING[b]	B	RP	Borneo. Sarawak, Miri, 6½ mile Bakam road.
EM	Haviland *s.n.*	Haviland GD	1892	SING[b]	B	RP	Borneo. Sarawak, Kuching.
EM	Coode 7951	Coode MJE *et al.*	30 Jan 1994	BRUN[a], SAN[a], SAR[a], SING[a]	B	RP	Borneo. Brunei, Tempurong, Batu Apoi, Selapon.

EM	BRUN 15706	Joffre AA *et al.*	20 Jul 1994	BRUN[b], SING[b]	B	B	Borneo. Brunei, Belait, Merangking-Buau road.
EP	Zen 2292	Zen	2 Nov 1960	SING[c]	B	RP	Borneo. Sarawak, Kuching, Arboretum tree no. 2292.
EP	S 40583	Paie I	6 Nov 1982	KEP[c]	B	RP	Borneo. Sarawak, Kuching, 12th mile Serian Road, Semengoh Arboretum, Block no. 22.
EP	S 352	Omar	4 Jan 1924	K[a] (Holotype)	B	RP	Borneo. Sarawak, Kuching, 12th Mile Penrissen Road (future Semengoh Arboretum).
EP	S 25517	Anderson JAR	3 Jun 1966	SAN[b]	B	RP	Borneo. Sarawak, Kuching, Semengoh Arboretum.
EP	S 24332	Banyeng N & Sibat L	30 Aug 1966	SAN[b]	B	RP	Borneo. Sarawak, Kuching, 12 mile Penrissen road [possibly Semengoh Arboretum].
EP	Ghazalli 13408	Ghazalli	2 Mar 1961	SING[b]	B	RP	Borneo. Sarawak, Kuching, Semengoh Arboretum.
EP	S 27991	Wright E	6 Sep 1968	BO[a], SAN[a], SAR[a], SING[a]	B	RP	Borneo. Sarawak, Bintulu, Similajau-Labang.
EP	S 36493	Ismawi O *et al.*	9 Jun 1977	KEP[a], SAR[a]	B	RP	Borneo. Sarawak, Kuching, 12.5 mile Penrissen road [possibly Semengoh Arboretum].
EP	Omar 352	Omar	4 Jan 1924	K[a], KEP[a]	B	RP	Borneo. Sarawak, Kuching, 12 mile Penrissen road [possibly Semengoh Arboretum].
ES	FRI 17832	Suppiah T	23 Sep 1970	KEP[a]	MP	MFMP	Malay Peninsula. Johor, ridge top Gunung Belumut.
ES	Goodenough *s.n.* (SING 0017325)	Goodenough JS	1892	SING[a]	MP	MP	Malay Peninsula. Singapore, Changi.
ES	Burkill 718	Burkill IH	2 Feb 1915	SING[b]	MP	MP	Malay Peninsula. Singapore, Jurong road, 12th

							mile.
ES	FRI 17752	Suppiah T	19 Sep 1970	SING[a]	MP	MP	Malay Peninsula. Rengam Forest Reserve, Gunong Lambak Microwave Station.
ES	Ngadiman 34523	Ngadiman	10 Nov 1937	SING[c]	MP	MP	Malay Peninsula. Singapore, Bukit Timah.
ES	Sinclair *s.n.* (BO 1684368)	Sinclair J	29 Jan 1949	BO[c]	MP	MP	Malay Peninsula. Singapore, Bukit Timah Forest Reserve, north view hut.
ES	Sinclair *s.n.* (BO 1684352)	Sinclair J	27 Nov 1953	BO[a]	MP	MP	Malay Peninsula. Singapore, Bukit Timah Forest Reserve, south view hut, edge of tank.
ES	FRI 2780	Kochummen KM	15 Sep 1969	SING[a]	MP	RP	Malay Peninsula. Johor, Mersing Endau Road, 10th mile.
ES	FRI 7680	Cockburn PF	27 Feb 1968	SING[a]	MP	RP	Malay Peninsula. Johor, east Johor coast, Tanjung Sedili Kechil.
ES	HS 1048	Hardial S & Samsuri	4 Mar 1973	SING[a]	MP	RP	Malay Peninsula. Johor, Jemaluang.
ES	K 365	Kadim & Noor M	19 Jul 1959	SING[a]	MP	RP	Malay Peninsula. Johor, Endau, Kampong Hubong.
ES	KEP 98940	Rahim I	4 Sep 1966	KEP[a]	MP	RP	Malay Peninsula. Terengganu, Dungun, Bkt Bauk Forest Reserve, compartment 17B.
ES	CWL 239	Chew WL	11 Oct 1961	SING[c]	MP	MP	Malay Peninsula. Singapore, Bukit Timah Forest Reserve.
ES	FRI 717	Whitmore TC	5 Sep 1966	KEP[b], SING[a,b]	MP	SAI	Malay Peninsula. Perak, Slim Hills Forest Reserve.
ES	AKK 349	Ando M *et al.*	Aug–Sep 1968	KEP[b]	MP	MP	Malay Peninsula. Selangor, Semangko Forest Reserve.

ES	KL 4199	Thomas P & Teo LE	26 Jan 1993	KEP[b]	MP	RP	Malay Peninsula. Johor, Mersing.
ES	FRI 2540	Kochummen KM	2 Feb 1970	KEP[b], SAR[b]	MP	RP	Malay Peninsula. Kelantan, Ulu Sat Forest Reserve.
ES	KL 4411	Wiant C & Teo LE	21 Sep 1994	KEP[a,b]	MP	RP	Malay Peninsula. Johor, Mersing.
ES	FRI 52922	Wilkie P *et al*	23 Feb 2007	KEP[b], SAN[b]	B	RP	Malay Peninsula. Johor, Kota Tinggi, Panti Forest Reserve.
ES	AKK 150	Ando M *et al.*	Aug–Sep 1968	KEP[b]	B	RP	Malay Peninsula. Selangor, Semangko Forest Reserve.
ES	Burkill 4334	Burkill IH	22 Mar 1919	BO[b]	MP	MP	Malay Peninsula. Singapore, Punggol.
ES	Goodenough *s.n.*	Goodenough JS	1892	SING[a]	MP	MP	Malay Peninsula. Singapore, Changi.
ES	Maxwell 81-74	Maxwell JF	15 Apr 1981	SING[b]	MP	MP	Malay Peninsula. Singapore, Bukit Timah Nature Reserve.
ES	SA 1486	Samsuri A	23 Nov 1978	SING[a]	MP	MP	Malay Peninsula. Singapore, Bukit Timah Nature Reserve.
ES	SFN 36366	Henderson MR	10 Mar 1939	BO[b], SING[b]	MP	MP	Malay Peninsula. Singapore, Bukit Timah Nature Reserve.
ES	Ridley 4947a	Ridley HN	23 Feb 1890	SING[a]	MP	MP	Malay Peninsula. Singapore, Bukit Timah Nature Reserve.
ES	SFN 36367	Henderson MR	10 Mar 1939	BO[b], SING[b]	MP	MP	Malay Peninsula. Singapore, Bukit Timah Nature Reserve.
ES	SFN 36371	Ngadiman	21 Mar 1939	BO[b], SING[b]	MP	MP	Malay Peninsula. Singapore, Bukit Timah Nature Reserve.
ES	Hardial 634	Hardial	23 Oct 1967	SING[a]	MP	MP	Malay Peninsula. Singapore, Bukit Timah Nature

							Reserve.
ES	SA 1324	Samsuri A	7 Dec 1976	SING[b]	MP	MP	Malay Peninsula. Singapore, Bukit Timah Nature Reserve.
ES	Anonymous *s.n.*	Anonymous	Jul 1885	SING[a]	MP	MP	Malay Peninsula. Singapore.
ES	SFN 36189	Ngadiman	6 Feb 1939	SING[b]	MP	MP	Malay Peninsula. Singapore, Bukit Timah Nature Reserve.
ES	Hullet? 189	Hullet RW?		SING[a]	MP	MP	Malay Peninsula. Singapore.
ES	SFN 36352	Ngadiman	27 Feb 1939	BO[b], SING[b]	MP	MP	Malay Peninsula. Singapore, Bukit Timah Nature Reserve.
ES	Goodenough *s.n.*	Goodenough JS	28 Mar 1893	SING[b]	MP	MP	Malay Peninsula. Singapore, Sungai Murai.
ES	FRI 0060	Whitmore TC	25 Feb 1966	SAN[b], SING[b]	MP	RP	Malay Peninsula. Pahang, Jengka Forest Reserve.
ES	Ridley 104	Ridley HN	10 Jan 1889	SING[b]	MP	MP	Malay Peninsula. Singapore, Jurong West.
ES	Ridley 4947	Ridley HN	1892	SING[a]	MP	MP	Malay Peninsula. Singapore, Tuas.
ES	SFN 33249	Corner EJH	30 Aug 1937	BO[a], KEP[a], SING[a]	MP	MP	Malay Peninsula. Singapore, Bukit Timah Nature Reserve.
ES	Osman 23827	Osman	24 Dec 1930	SING[a]	MP	MP	Malay Peninsula. Selangor, Kuala Kubu, Bukit Kutu.
ES	CF 3102	Yeop AR	18 Mar 1918	SING[b]	MP	RP	Malay Peninsula. Pahang, Bukit Goh Forest Reserve.
ES	FRI 11360	Suppiah T	8 Jul 1969	KEP[a], SING[a,b]	MP	MFMP	Malay Peninsula. Negeri Sembilan, Jelebu Forest Reserve.
ES	FRI 7774	Cockburn PF	3 Mar 1968	SING[b]	MP	RP	Malay Peninsula. Johor, Kota Tinggi, Gunung

							Panti Timur.
ES	FRI 2616	Kochummen KM	14 Nov 1968	SING[b]	MP	RP	Malay Peninsula. Terengganu, Bukit Bauk Forest Reserve.
ES	MS 2224	Mohd Shah & Sanusi	25 Sep 1970	SING[a]	MP	MFMP	Malay Peninsula. Johor, Kluang, Gunung Belumut.
ES	FRI 15143	Whitmore TC	13 Feb 1970	KEP[b], SAN[b], SING[b]	MP	MP	Malay Peninsula. Selangor, road to Gap of Fraser's Hill.
ES	SA 1353	Samsuri A	11 Jan 1977	SING[a]	MP	MP	Malay Peninsula. Singapore, Bukit Timah Nature Reserve.
ES	Burkill 7611	Burkill IH	14 Jan 1922	SING[a]	MP	MP	Malay Peninsula. Singapore, Punggol.