Phylogeny and evolutionary history of Sapindaceae and *Dodonaea*

Thesis submitted by

Mark Gerard Harrington BSc (Hons) JCU

in May 2008

for the degree of Doctor of Philosophy

in the School of Marine & Tropical Biology

James Cook University

STATEMENT OF ACCESS

I, the undersigned, author of this work, understand that James Cook University will make this thesis available for use within the University Library and, via the Australian Digital Theses network, for use elsewhere.

I understand that, as an unpublished work, a thesis has significant protection under the Copyright Act and;

I do not wish to place any further restriction on access to this work.

_____ _____

Signature Date

STATEMENT OF SOURCES

DECLARATION

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

_____ _____

Signature Date

ELECTRONIC COPY

I, the undersigned, the author of this work, declare that the electronic copy of this thesis provided to the James Cook University Library is an accurate copy of the print thesis submitted, within the limits of the technology available.

_____ _____

Signature Date

STATEMENT ON THE CONTRIBUTION OF OTHERS

This project was funded with grants from the Australian Biological Resources Study, the Cooperative Research Centre for Tropical Rainforest Ecology and Management (Rainforest CRC) and James Cook University (JCU). I was supported by a Australian Postgraduate Award stipend and a completion scholarship from JCU.

Supervision was provided by Prof. Paul Gadek and Dr. Will Edwards (JCU) with editorial assistance from Christopher Quinn, Honorary Research Associate, Botanic Gardens Trust, Sydney.

Acknowledgements

I would like to thank my supervisors Prof. Paul Gadek and Dr. Will Edwards, and Dr. Judy West from the Australian National Herbarium, Christopher Quinn, and Ed Biffin from Adelaide University for further discussions and assistance.

I acknowledge Lynne Jones and Mathew Pye for advice on molecular techniques.

I extend my thanks to Peter Raven for exhorting his many friends to help get plant samples for this project. I thank the following people and institutions from around the world for supplying plant material – Chris Birkinshaw, George Yatskievych and Gregory Gust Missouri Botanical Garden and DNA Bank, Herbarium, Dr. Victor Barradas University National avtonoma of Mexico, Dr. Brinsley Burbidge St. George Village Botanical Garden, St. Croix Virgin Islands, Suhua shi, Sun Yat-sen University, Guangzhou, Cassius Kayombo National Herbarium of Tanzania, Pedro Acevedo United States National Herbarium, Smithsonian Institution, National Tropical Botanical Garden, Kauai, Hawaii, F. Billiet Living Plant Collections, National Botanic Garden of Belgium, Botanic Garden, University of Copenhagen, Dr. Jenny Xiang, North Carolina State University, Tony Roberts. Mary Gandini, Stuart Worboys, James Cook University, Judy West, Kirsten Cowley, Lyn Craven and Ed Biffin Centre for Plant Biodiversity Resaearch, Australian National Herbarium (CANB), Australian National Botanic Gardens living collection, Darren Crayn, Royal Botanical Gardens Sydney, Dr. Shyam Phartyal Forest Research Institute India, Cassius Kayombo National Herbarium of Tanzania, Amanda Shade Botanic Gardens and Parks Authority Perth Western Australia, Chien-I Huang, Royal Botanic Gardens Edinburgh, F. Billiet Living Plant Collections, National Botanic Garden of Belgium, Ching I Peng, Research Centre for Biodiversity. Academia Sinica, Taipei, Taiwan and Christine Stucley; and DNA samples – Shelly James Pacific Centre for Molecular Biodiversity, Bishop Museum, Hawaii Kew Botanical Gardens – DNA Bank, Dr. Clifford Morden Hawaii Plant DNA Library.

Finally, I can now get back to painting the house.

ABSTRACT

The Sapindaceae (soapberry family) is a large family of trees, shrubs and lianas comprising 133 genera (49 monotypic) and ca.1450 species, most with a tropical to subtropical distribution. The circumscription of Sapindaceae has varied, particularly with regard to inclusion of genera from the closely related, predominantly temperate families Aceraceae and Hippocastanaceae. Sequence data from two plastid genes, analysed separately and together using parsimony and Bayesian analysis support a broadly defined Sapindaceae incorporating Aceraceae, Hippocastanaceae and *Xanthoceras*. A division into four subfamilies is proposed: Sapindoideae, Hippocastanoideae, Dodonaeoideae, and Xanthoceroideae. Tribal groupings are critically evaluated in light of the analyses.

The evolutionary history of Sapindaceae is evaluated using a variety of Bayesian relaxed clock molecular estimates of divergence times, which either incorporate the dates implied by the fossil record, fossil constraints from outside Sapindaceae, and no fossil constraints. Analyses with fossil constraints from outside Sapindaceae imply a Pliocene-Miocene (6-28 Mya) origin for *Acer* and *Aesculus* rather than a Paleocene (ca 64 Mya) origin implied by the earliest attributed fossils of these genera, and an evolutionary rate for *Acer* and *Aesculus* consistent with the majority of other genera of Sapindaceae sampled. Including the fossil dated Paleocene origin of *Acer* and *Aesculus* as hard bound priors and using four constraints from outside Sapindaceae result in a potentially biologically implausible rapid change in the mean evolutionary rate on the stem branch leading to the split between *Acer* and *Aesculus* and their respective sister genera. These conflicting scenarios suggest the need for a substantial re-evaluation of our understanding of the tempo and mode of evolution of these lineages.

Intergeneric relationships within Dodonaeoideae, Hippcastanoideae and Xanthoceroideae were also assessed by firstly, generating secondary structure predictions for ITS and partial ETS sequences, and then using these predictions to assist alignment of the sequences. Secondly, the alignment was analysed using RNA specific models of sequence evolution that account for the variation in nucleotide evolution in the independent loops and covariating stems regions of the ribosomal spacers. The

phylogeny drawn from the analyses was compared with that from analyses using 'traditional' 4-state models and the plastid analyses.

To further our understanding of the origins of the Australian flora, and in particular plant adaptation and evolution in increasingly arid environments, the phylogeny and evolutionary history of Hopbushes (*Dodonaea* – Sapindaceae) and Pepperflowers (*Diplopeltis* – Sapindaceae) were evaluated based on nuclear ITS and partial ETS sequences and models of nucleotide evolution that incorporate secondary structure. The tempo and mode of evolution of these sister genera were evaluated using Bayesian relaxed clock molecular estimates of divergence times, Bayesian relative rates test, lineage through time plots and estimations of diversification rates (speciation minus extinction). The dry and temperate adapted genera of Sapindaceae (*Diplopeltis* species and *Dodonaea* including *Distichostemon* species) are relatively recent radiations in the Australian flora, and are most likely to be in response to increased aridity and seasonality from the late Miocene14 Mya to Recent. There is evidence of long distance dispersal from northern Australia to Madagascar in the early Pliocene.

The cosmopolitan, polymorphic species *Dodonaea viscosa* (hop bush, varnish tree) has been the subject of taxonomic and ecological enquiry for over 150 years. ITS and partial ETS sequences for 50 samples from across its worldwide distribution were used to evaluate the evolutionary and biogeographic history of this species complex. *Dodonaea viscosa* is not an old lineage as has often been speculated based primarily on it vast distribution, but diverged from its most recent common ancestor and subsequently dispersed around the world within the last two million years. Results also indicate that there are at least two evolutionary lineages within *D. viscosa*.

Table of contents

List of tables

List of figures

Chapter 1: Introduction

The primary focus of this thesis is to develop an understanding of the phylogenetic, evolutionary and biogeographic patterns of a major Australian plant group: the virtually endemic Australian genus *Dodonaea*, which is the largest genus of Sapindaceae in Australia.

Sapindaceae are a reasonably large family of mainly trees, shrubs and lianas most with a primarily tropical or subtropical distribution. The greatest centre of diversity appears to be in the Southeast Asian region. Within Australia, Sapindaceae form a significant component of rainforest environments, with most of the 30 genera restricted to the northern tropical region, with a few species extending down the east coast within subtropical rainforests. The marked exception to this is the continent-wide distribution of the speciose *Dodonaea* and its allies, which grow as shrubs or small trees in a variety of arid, temperate and tropical habitats, and are only absent from dense rainforest and alpine communities.

The family is presently divided into two subfamilies based on whether there are one or more ovules per locule, with *Dodonaea* placed within the Dodonaeoideae (Radlkofer, 1933). This subfamily comprises 29 genera (14 monotypic) from five tribes, with a predominately austral distribution, suggesting it has had an ancient Gondwanan origin (Muller and Leenhouts, 1976). Very little is known about the complex pattern of relationships amongst the members of Dodonaeoideae (Muller and Leenhouts, 1976; Buijsen et al., 2003), and consequently there is little understanding of the diversity in this group, or the factors underlying it.

This study uses molecular sequence data for analyses of phylogenetic relationships to answer the conflict between the hypotheses of Radlkofer (1933), Muller and Leenhouts (1976) and West (1984) about the historical patterns of evolution amongst the taxa of the subfamily Dodonaeoideae, with particular focus on the tribe Dodonaeeae (includes *Diplopeltis*, *Dodonaea*, *Distichostemon* and *Loxodiscus*) and the genus *Dodonaea*.

Dodonaeeae includes two endemic Australian genera (*Diplopeltis* and *Distichostemon*), the monotypic *Loxodiscus* from New Caledonia, and *Dodonaea* that

are predominantly endemic to Australia. There are 64 *Dodonaea* species: 61 are found only in Australia (West, 1984; 1987), *D. polyandra* also occurs in New Guinea, *D. madagascariensis* only on the mountains in central Madagascar, while the polymorphic *D. viscosa* has a world wide distribution occurring between 33^{o} N and 44^{o} S, and being found in a variety of habitats from sea level to 3500 m altitude (Leenhouts, 1983). Many varieties and forms of *D. viscosa* have been described from within regional localities (Leenhouts, 1983).

Revisionary studies of *Dodonaea* undertaken by West (1984) identified six natural species groups, based on combinations of macromorphological characters. Several of these groups contained species in which adaptations to aridity were pronounced. However, West (1982b) was unable to determine whether those species that are now endemic to arid zones are evolved from previously more widespread species of the wetter, pre-arid environment of ancient Australia, or relatively recent invaders from surrounding temperate and tropical areas.

What is the significance of this exceptional adaptation in *Dodonaea* to aridity on the Australian continent, particularly occurring within this otherwise tropical family? We might attempt to interpret this by reference to the current understanding of the taxonomic affinity of the genus within the family, but this is poorly supported and has never been properly tested.

Understanding the ways plants have responded to changing climatic variables over time, why some adapt and change, or become widespread or rare, is critical to understanding and managing our unique environment in the future. Radiations can occur primarily either through extrinsic causes due to new environmental circumstances (chance dispersal to new environments and subsequent speciation into a wide diversity of ecological roles) or through intrinsic characters of the organisms themselves (the appearance of particular traits that allow the plant to utilise a niche in a novel manner) (Givnish, 1997; Hodges, 1997). Explicit tests of both these models can be difficult, either because of a lack of information on the historical biogeography of the group, or a lack of information to test whether diversification rates between lineages are actually associated with specific traits and are different from chance alone.

Phylogenetic analyses are essential for testing hypotheses of adaptive radiations because they identify monophyletic groups that provide data to identify radiations, to delineate the timing of the evolution of proposed key innovations, and to identify the sister taxa that are imperative in analyses of diversification (Losos and Miles, 2002). To understand the origin of the characters and traits associated with adaptations to aridity it is important to know firstly something of the phylogeny of plants found in these environments.

Knowledge of biogeographic distributions is vitally important to the conservation of Australia's biodiversity. In order to understand the basis of the present-day distribution patterns of the Australian flora, biogeographical analyses need to be undertaken for many groups of taxa (West, 1998). Very few biogeographical or temporal studies of the Australian flora have been undertaken (Crisp et al., 2004). This thesis will reconstruct the phylogeny of Sapindaceae and *Dodonaea* based on molecular sequence data, and use these phylogenies to construct and test hypotheses on the biogeographical and temporal evolution of the Australian flora.

The following specific questions will be addressed:

a) Is the Dodonaeoideae sensu Muller and Leenhouts (1976) monophyletic, and what are the relationships of Aceraceae and Hippocastanaceae to it?

b) Do the molecular phylogenies support the explicit infratribal phylogenetic hypotheses of Muller and Leenhouts (1976), and if not, what are the implications for our understanding of the evolution of the Australian elements of these tribes?

c) What is the origin of the predominately Australian genus *Dodonaea*?

d) Are the intergroup relationships of the six species groups of *Dodonaea* (West 1984), which the present morphological database cannot resolve with any certainty, resolved by the addition of molecular data?

e) Are arid habitat species of *Dodonaea* relicts of previously more widespread groups, or have they recently radiated into the arid zone from surrounding temperate or tropical areas?

f) Is there a correlation between the pantropical distribution of *D. viscosa*, and the evolution of particular genetic and/or morphological characters?

Chapter 2: Phylogenetic inference in Sapindaceae sensu lato using plastid *mat*K and *rbc*L DNA sequences

The Sapindaceae (soapberry family) is the largest family of the Order Sapindales. It is a family of trees, shrubs and lianas that comprises 133 genera (49 monotypic) and ca.1450 species, most with a tropical to subtropical distribution (Klaassen, 1999). Although some genera extend into temperate regions in Asia and North America, the greatest diversity is found in tropical Southeast Asia. Economically important members of the family include the tropical fruits lychee (*Litchi chinensis*), longan (*Dimocarpus longan*) and rambutan (*Nephelium lappaceum*).

Circumscription of the Sapindaceae has varied, particularly with regard to inclusion of genera from the closely related, predominantly temperate families Aceraceae and Hippocastanaceae. Takhtajan (1987), Cronquist (1988), and Dahlgren (1989) maintained Aceraceae and Hippocastanaceae as separate families. Broader concepts of Sapindaceae have frequently been adopted that have incorporated Aceraceae (Bentham and Hooker, 1862a), Hippocastanaceae (Hutchinson, 1926), or both (morphological basis - Heimsch, 1942; Muller and Leenhouts, 1976 Judd et al. 1994; secondary chemistry - Umadevi and Daniel, 1991; molecular basis - Gadek et al., 1996a; APG, 1998; Savolainen et al., 2000b; Thorne, 2000)

Radlkofer (1890; 1933) recognised two subfamilies within Sapindaceae sensu stricto (s.s.), Sapindoideae (= Eusapindaceae) having a single apotropous and upright or ascending ovule per locule, and Dodonaeoideae (= Dyssapindaceae) with two or rarely more apotropous and upright ovules per locule, or rarely one ovule which is epitropous and hanging (table 2.1). The Sapindoideae were further subdivided into two groups: Eusapindaceae nomophyllae (containing 2 tribes) with imparipinnate leaves, zygomorphic flowers with oblique disk, and folded cotyledons; and Eusapindaceae anomophyllae (7 tribes) with paripinnate leaves, actinomorphic flowers with annular disk, and curved cotyledons. Radlkofer (1890, 1933) considered the five tribes of the Dodonaeoideae could also be separated into two groups based on the presence or absence of a terminal leaflet (table 2.1). Radlkofer (1890) believed that elaborate structures such as possession of tendrils (modified inflorescence axes) for climbing, and

specialized pollen morphology of the eight genera of lianas of the Paullinieae (Sapindoideae), that contains almost one third of the family's species, represented the primitive form in the Sapindaceae, and that evolution in the family had proceeded via a series of reductions in morphological structures. He also considered that Dodonaeoideae were derived from Sapindoideae, and that the various tribes in the former were possibly derived from different tribes in the latter, which would render Dodonaeoideae polyphyletic. His tribal concepts remain virtually intact to this day, although a number of later authors have commented on his familial and subfamilial groupings.

Muller and Leenhouts (1976) assessed the phylogenetic relationships within the family using macromorphologic-typological phylogenetic series in two vegetative and 11 floral characters that were compared to a phylogenetic series derived from pollen morphology to support Radlkofer's concept of two subfamilies. However, they divided Sapindoideae into three groups (A-C) of equal rank that were further characterised by their geographical distributions and presence or absence of an aril, rather than accepting the two groups recognised by Radlkofer (table 2.1). They proposed almost a complete reversal of the phylogenetic scheme of Radlkofer (1890), recognising the morphological and palynological specialisations within some tribes of the Sapindoideae as representing the derived form in the family, and also suggesting that the predominantly austral distribution of Dodonaeoideae, along with the predominance of 'primitive' morphological and palynological forms, was evidence that this subfamily may contain the 'more primitive' lineages. Whereas Muller and Leenhouts (1976) made explicit and detailed hypotheses of the direction of evolution within and between the groups and tribes of Sapindoideae, their assessment of phylogenetic relationships within Dodonaeoideae remained ambiguous. The recent cladistic analysis of a morphological data set for a limited representation of the family (Judd et al., 1994) indicated that both Hippocastanaceae and Aceraceae were embedded with Sapindaceae s.s., and that neither subfamily was monophyletic.

Umadevi and Daniel (1991) proposed the division of Sapindaceae sensu lato (s.l.) into four subfamilies (table 2.1): Aceroideae (composed of just the members of the former Aceraceae), Sapindoideae (including Hippocastanaceae as a tribe), Dodonaeoideae (containing only *Dodonaea*), and Koelreuterioideae (containing the

remainder of the former Dodonaeoideae). Thorne (2000) divided the family into five subfamilies, separating Hippocastanoideae from Sapindoideae (table 2.1).

Plastid DNA sequences, especially the *rbcL* gene and more recently *matK*, have been used extensively to infer plant phylogenies: *rbcL* has been one of the most widely used regions of DNA for cladistic analysis, especially at the familial level (Chase and Albert, 1998; Soltis and Soltis, 1998; Savolainen et al., 2000a); *matK* has been useful in clarifying relationships at taxonomic levels at and below the family (Johnson and Soltis, 1995; Plunkett et al., 1997; Soltis and Soltis, 1998; Hilu and Alice, 1999). Combining sequence data from two or more loci has proved to be a powerful method of estimating phylogenies, providing improved resolution and internal support compared to single gene analyses (Soltis and Soltis, 1998; Soltis et al., 2000). Many recent studies have successfully combined *rbcL* and *matK* sequence data to resolve problems at the subfamilial and tribal levels in a range of angiosperm families (Soltis et al., 1996; Xiang et al., 1998b; Goldman, 2001).

This study is a continuation of the broader molecular investigation of the Sapindales (Gadek et al. 1996) that identified a strongly supported sister relationship of Aceraceae and Hippocastanaceae with Sapindaceae s.s. However, the restricted sampling in that study was not sufficient to differentiate the pattern of relationships within Sapindaceae s.s. The objectives of this study were to conduct phylogenetic analyses of Sapindaceae s.s., Hippocastanaceae and Aceraceae using DNA sequence data from *matK* and *rbcL* to evaluate: (1) family and subfamily concepts within Sapindaceae s.l., and 2) current tribal concepts and alliances based on morphology.

Table 2.1: Enumeration schemes for subfamilies and tribes assigned to Sapindaceae. The numerical sequence of Radlkofer (1890, 1933) and Muller and Leenhouts (1976) classification systems is intended to generally reflect phylogeny with the lowest numbered tribe considered to be the more primitive. Nomophyllae = tips of leaf completely developed (in compound leaves a small terminal leaflet is present), anomophyllae = tip of leaf (apart from simple leaves) incompletely developed (normally the actual terminal leaflet is missing).

Radlkofer (1933)	Muller and Leenhouts (1976)	Umadevi and Daniel (1991)	Thorne (2000)
Subfam. I. Eusapindaceae	**Subfam II. Sapindoideae**	**Sapindoideae**	**Sapindoideae**
Eusapindaceae nomophyllae	Group C	(incl. Hippocastanaceae)	
1. Paullinieae	13. Paullinieae		
2. Thouinieae	12. Thouinieae		
Eusapindaceae anomophyllae	Group A		
3. Sapindeae	8. Sapindeae		
4. Aphanieae	7. Lepisantheae (incl. Aphanieae)		
5. Lepisantheae	6. Melicocceae		
6. Melicocceae	Group B		
7. Schleichereae	10. Schleichereae		
8. Nephelieae	11. Nephelieae		
9. Cupanieae	9. Cupanieae		
Subfam. II. Dyssapindaceae	**Subfam I. Dodonaeoideae**		
Dyssapindaceae nomophyllae			
10. Koelreuterieae	5. Koelreuterieae	**Koelreuterioideae**	**Koelreuterioideae**
11. Cossignieae	4. Cossinieae	**Koelreuterioideae**	**Koelreuterioideae**
12. Dodonaeeae	1. Dodonaeeae	**Dodonaeoideae (*Dodonaea*)**	**Dodonaeoideae (*Dodonaea*)**
Dyssapindaceae anomophyllae			
13. Doratoxyleae	2. Doratoxyleae	**Koelreuterioideae**	**Koelreuterioideae**
14. Harpullieae	3. Harpullieae (possibly incl. Hippocastanaceae)	**Koelreuterioideae**	**Koelreuterioideae**
(Aceraceae)	(3a? Aceraceae)	**Aceroideae**	**Aceroideae**
(Hippocastanaceae)			**Hippocastanoideae**

Materials and Methods

Sampling

Plastid sequence data were acquired from 144 taxa representing 95 genera for this study. Sequences were drawn from other laboratories (Mark Chase – Molecular Systematics Section, Jodrell Laboratory, Royal Botanic Gardens, Kew, U.K. 44 *rbcL* sequences) and individuals Jenny Xiang (*matK* sequences for representatives of Hippocastanaceae) and Karen Edwards (68 *matK* sequences). Outgroup sequences were generally retrieved from GenBank, and I generated a further 30 *rbcL* and 27 *matK* ingroup sequences. The sampling strategy was intended to broadly survey the entire family, including both subfamilies and generally following the existing taxonomy of Muller and Leenhouts (1976; including name changes that follow recent revisions). The sampling of Sapindaceae s.s. includes approximately 70% of recognised genera, with multiple representatives from all tribes, including 25/29 of the genera included in Dodonaeoideae, and 52/103 of Sapindoideae, and both genera of Aceraceae and Hippocastanaceae. While sampling generally covers the entire geographical distribution of the Sapindaceae s.s., and includes nearly all genera with >10 species (except *Cupania*, *Urvillea*, *Thinouia*, *Toulicia* and *Thouinia* from tropical America, *Placodiscus* from tropical West Africa, and *Tinia* and *Tinopsis* from Madagascar), there is less representation from the multitude of small or monotypic genera from tropical America and Madagascar. Subfamily and tribal placements of all included species, GenBank accession numbers, voucher details, and references for previously published sequences are listed in appendix 2.1. Previous phylogenetic studies (Gadek et al. 1996; Savolainen et al., 2000) have been inconclusive as to the sister relationship of a strongly supported monophyletic "Sapindaceae s.l. clade", that includes representatives of Sapindaceae s.s., Aceraceae, and Hippocastanaceae, from amongst the other eight families of the Sapindales (APG, 2003). Outgroups from multiple combinations of Sapindales family representatives (see discussion) were used to root the analyses.

For all taxa, total genomic DNA was isolated from fresh, silica dried, or herbarium material using the 2x CTAB protocol (Doyle and Doyle, 1990). Double-stranded copies of *matK* and *rbcL* were amplified from total DNA using the polymerase chain reaction (PCR). Amplification products and methods for sequencing *matK*, including primers

used, were the same as previously outlined in Edwards and Gadek (2001); *rbcL* was amplified and sequenced as described by Fay et al. (1998). Additional primers required to sequence *matK* for representatives from Dodonaeoideae and Anacardiaceae are included in table 2.2.

Nucleotide sequences for *matK* and some *rbcL* were visually checked, with gaps inserted where necessary for alignment, and consensus sequences constructed using ABI Prism Sequence Navigator (Perkin Elmer), whereas the remaining *rbcL* sequences were edited and aligned using the programs Sequence Navigator and Auto Assembler (Applied Biosystems, Inc.); the sequences were produced in two different labs, hence the use of different software.

Table 2.2: List of new *matK* primers used in this study.

Dodonaeoideae	4180, 5'TCACTAAGTGTGAAACGTT3'
	100F 5'GCRYTATGTATCATTTGATA3'
	900F 5'TATGSTTGTWAYAGSATCCT3'
Anacardiaceae	4244, 5'TTAGGGCATCCCATYAGTAAGG3'

Combined Sequence Data

To assess whether phylogeny reconstruction would benefit from a combined analysis of the plastid data a joint *rbcL-matK* database was constructed for a subset of taxa (48 plus six outgroups), with the further inclusion of eight composite genera for which different species from each data set were combined to represent a single placeholder taxon (Hoot et al., 1997). The prerequisite for the creation of composite genera was the requirement for monophyly (Kellogg and Linder, 1995). Species that were combined to produce each composite genus are given in table 2.3. The combinability of the *rbcL* and *matK* data sets was evaluated by implementing the partition homogeneity test (Farris et al., 1995) to assess incongruence of the matrices. The test implemented in PAUP* 4.0b10 (Swofford, 2002) included 500 homogeneity replicates, each involving 10 tree bisection reconnection (TBR) branch swapping replicates, and the maximum number of trees saved was restricted to 100 per replicate, with uninformative characters excluded (Lee 2001).

Table 2.3: Species combined as placeholder taxa.

Genus	Species – *matK*	Species – *rbcL*	References for monophyly
Aesculus	*pavia*	*hippocastanum*	Forest et al. (2001) – morphological characters Xiang (1998a) – *matK* and ITS sequence data
Allophylus	*javensis*	*natalensis*	Acevedo-Rodriguez (1993) – pollen morphology and macromorphological characters
Arytera	*litoralis*	*divaricate*	Turner (1995) – macromorphological and leaf anatomical characters
Billia	*hippocastanum*	*sp.*	Forest et al. (2001) – morphological characters
Dodonaea	*triquetra*	*viscosa*	M. Harrington (unpubl. data – chapter 5) – ITS and ETS sequence data
Guioa	*semiglauca*	*lasioneura*	van Welzen (1989) – macromorphological and leaf anatomical characters
Harpullia	*arborea*	*ramiflora*	Buijsen (1995; 2003) – leaf anatomical and morphological characters Muller (1985) – pollen morphology Leenhouts (1985) – morphological characters
Paullinia	*venosa*	*pinnata*	Acevedo-Rodriguez (1993) – pollen morphology and macromorphological characters

Phylogenetic Analyses

Parsimony analyses on the separate *matK*, *rbcL*, and combined datasets were performed with PAUP* using the heuristic search option with 1,000 random taxon additions. The MulTrees options and TBR branch swapping were employed. Analysis of the separate *rbcL* and *matK* datasets was over two stages because the number of trees found in an initial search exhausted available memory: the initial search saved twenty trees in each random taxon addition replicate, while the second search used TBR branch swapping on all these saved trees with Maxtrees set to 20,000. No limit was set for the combined *rbcL-matK* datasets. Relative support for clades in all analyses was inferred using 500 bootstrap (Felsenstein, 1985) replicates, each involving 10 TBR branch swapping replicates and MulTrees on saving only 100 trees per replicate (Salamin et al., 2003) to enable the analyses to go to completion. Bremer support/decay values (Bremer, 1988; Donoghue et al., 1992) for each clade were inferred using PAUP* and AutoDecay 4.0.2 PPC (Eriksson, 1998) with 10 replicates of random taxon addition and TBR branch swapping for each constraint tree, and saving 100 trees each replicate. All characters were equally weighted, and potentially informative insertions/deletions (indels) were

scored as presence/absence characters for inclusion in the analyses, and also later optimized onto phylogenetic trees. Datasets are on the accompanying CD.

To compare the results from the maximum parsimony analysis with other analytical methods, a Bayesian estimation of posterior probability (PP) for each node was conducted in MrBayes ver. 3.1.2 (Huelsenbeck and Ronquist, 2001). An advantage of current Bayesian methods over maximum likelihood approaches is that it allows partitioning of the datasets and the running of separate models of nucleotide evolution with independent parameters across the partitions. Simulation studies have shown that there may be incorrect branch length estimation by maximum likelihood methods when heterogeneous data partitions are analysed by a single evolutionary model (Kolaczkowski and Thornton, 2004). Bayesian inference of phylogeny also allows uncertainty in the model parameters to be incorporated into the phylogenetic reconstruction (Nylander et al., 2004). Nucleotide substitution models were selected by the corrected Akaike Information Criterion (AIKc) using MrAIC.pl 1.4 (Nylander, 2004), with the likelihood scores under different models estimated using PHYML (Guindon and Gascuel, 2003b). For both the *rbcL* and *matK* genes, the general-time-reversible (GTR) model, with variable rates and gamma distribution was selected. In MrBayes each data set was partitioned by codon position to allow for variable rates to be calculated across each position (MrBayes commands set partition=by codon prset ratepr=variable Lset nst=6 rates=invgamma). Two replicate analyses were conducted for each dataset starting with a randomly selected tree, with the Markov-chain Monte Carlo searching over four parallel tree-search paths, for one million generations, saving one tree every 100 generations. In the combined analysis, variation was partitioned among genes, with all parameters free to vary independently within each partition, and each gene was allowed to evolve at a different rate. To determine statistically significant PP support for individual nodes, 95% majority rule consensus trees were generated in PAUP* from the trees remaining after 1000 burn-in trees were excluded from each analyses (Huelsenbeck and Ronquist, 2001; Lewis, 2001). The MrBayes command sumt was used to generate a consensus tree with branch lengths (measured in expected substitutions per site and with values based only on those trees containing the relevant bipartition).

In evaluating the results of analyses, bootstrap values >50%, decay values ≥1 and PP values ≥95% are interpreted as indicating a degree of support from the molecular data for elements in the topology.

Results

Sequence Data

The length of the *rbcL* sequence was 1428 base pairs (bp) for all 86 taxa (including two outgroups from Malvales and 14 representative genera from seven other families of Sapindales) sampled, but the first 28 bp containing the annealing site of the 5' PCR primer were deleted prior to analysis. The aligned *matK* matrix consisted of 1704 positions and 117 taxa (includes 12 outgroups). There is a frame shift at the 3' end of the *matK* gene that results in variability in the position of the stop codon in a number of taxa due to a number of large (7-24 bp) indels, and as these also created difficulties in alignment with outgroup taxa, positions 1645-1688 were excluded from the analysis. In the *matK* data matrix 23 indels were identified, of which 12 were potentially phylogenetically informative (figs. 2.2-3: indels a-l) and 11 were autapomorphic.

In the *rbcL* analysis, 461 of the 1400 characters were variable, of which 255 (18.2 %) were potentially informative. Of the 1557 characters from *matK*, 899 were variable and 618 (39.7%) parsimony informative. The combined dataset of 2952 characters for 62 taxa included 10 scored indels; 1196 positions were variable and 683 (23.1%) parsimony informative.

Figure 2.1 (next page): Consensus tree with branch lengths obtained from the Bayesian *rbcL* analysis, with all PP values ≥95% on the branches shown with ●. Bootstrap support (>50%) and decay values (≥1) only displayed on branches discussed in chapter. Tribal assignments follow Muller and Leenhouts (1976): Cos – Cossinieae, Dod – Dodonaeeae, Dor – Doratoxyleae, Har – Harpullieae, Koel – Koelreuterieae, Mel – Melicocceae, Lep – Lepisantheae, Sap – Sapindeae, Cup – Cupanieae, Sch – Schleichereae, Nep – Nephelieae, Tho – Thouinieae, Pau – Paullinieae, Acer - Aceraceae.; subfamilial assignments: D, Dodonaeoideae; S, Sapindoideae. Major clades labelled I-III; subclades A-D; see text.


0.1 substitutions/site
Serjania communis Pau S
Paullinia venosa Pau S
Cardiospermum grandiflorum Pau S
Allophylus javensis Tho S
Allophylus cobbe Tho S
Bridgesia incisifolia Tho S
Diatenopteryx sorbifolia Tho S
Guindilia trinervis Tho S
Matayba tenax Cup S
Toechima tenax Cup S
Guioa semiglauca Cup S
Mischocarpus littoralis Cup S
Cupaniopsis anacardioides Cup S
Jagera serrata Cup S
Arytera littoralis Cup S
Diploglottis campbelli Cup S
Dimocarpus longan Nep S
Litchi sinensis Nep S
Blighia sapida Cup S
Nephelium mutabile Nep S
Talisia nervosa Mel S
Tristiropsis acutangula Mel S
Alectryon connatus Nep S
Alectryon excelsus Nep S
Sapindus saponaria Sap S
Erioglossum rubiginosum Lep S
Hornea mauritannia Sap S
Atalaya alata Sap S
Lepisanthes alata Lep S
Hebecoccus ferugineus Lep S
Tristira triptera Mel S
Paranephelium macrophyllum Cup S
Schleichera oleosa Sch S
Koelreuteria elegans Koel D
Koelreuteria paniculata Koel D
Erythrophysa aesculina Koel D
Erythrophysa transvaalensis Koel D
Delavaya yunnanensis Har D
Ungnadia speciosa Har D
Arfeuillea arborescens Har D
Conchopetalum brachysepalum Har D
Magonia pubescens Har D
Majidea fosteri Har D
Diplopeltis huegelii Dod D
Llagunoa mollis Cos D
Diplopeltis stuartii Dod D
Distichostemon hispidulus Dod D
Loxodiscus coriaceus Dod D
Cossinia pinnata Cos D
Dodonaea triquetra Dod D
Dodonaea viscosa NZ Dod D
Harpullia arborea Har D
Llagunoa nitida Har D
Sinoradlkofera minor Koel D
Exothea diphylla Dor D
Exothea paniculata Dor D
Ganophyllum falcatum Dor D
Hypelate trifoliata Dor D
Zanha africana Dor D
Filicium decipiens Dor D
Filicium longifolium Dor D
Hippobromus pauciflora Dor D
Averrhoidium dalyi Dor D
Diplokeleba floribunda Cup S
Aesculus pavia Har D
Billia hippocastanum Har D
Handeliodendron bodinieri Har D
Acer saccharum Acer
Dipteronia sinensis Acer
Xanthoceras sorbifolium Har D
Flindersia australis
Acronychia acidula
Rutaceae
Simarouba glauca
Quassia amara
Simbaroubaceae
Azadirachta indica
Cedrela odorata
Meliaceae
Bursera fagarioides
Commiphora habessinica
Burseraceae
Buchanania latifolia
Anacardium sp.
Anacardiaceae
Kirkia wilmsii
Kirkiaceae
Nitraria retusa
Peganum harmala
Nitrariaceae
Biebersteinia orphanidis
Biebersteinaceae
Bombax
Bixa orellana
MALVALES
A
B
C
D
I
II
III

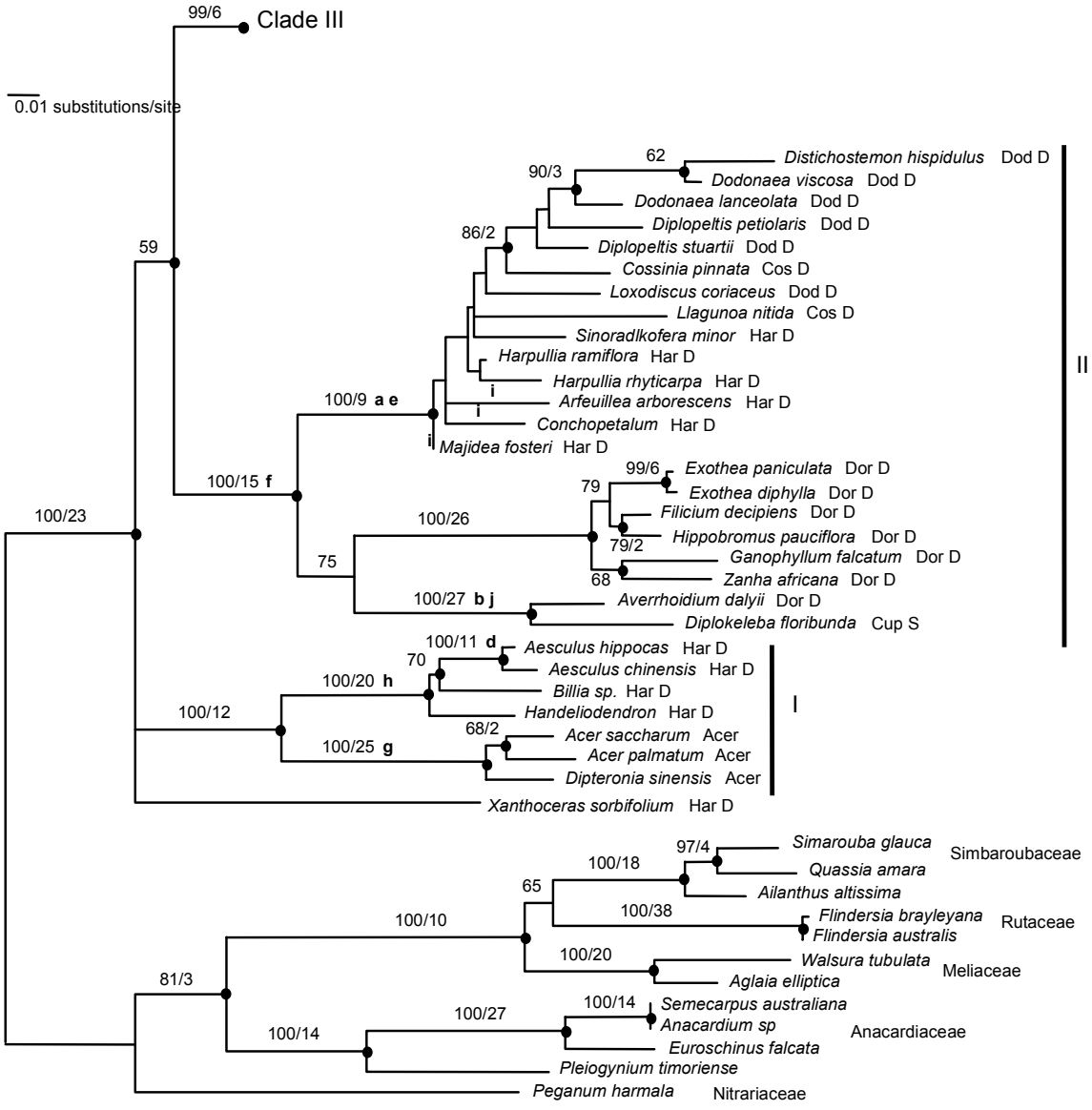
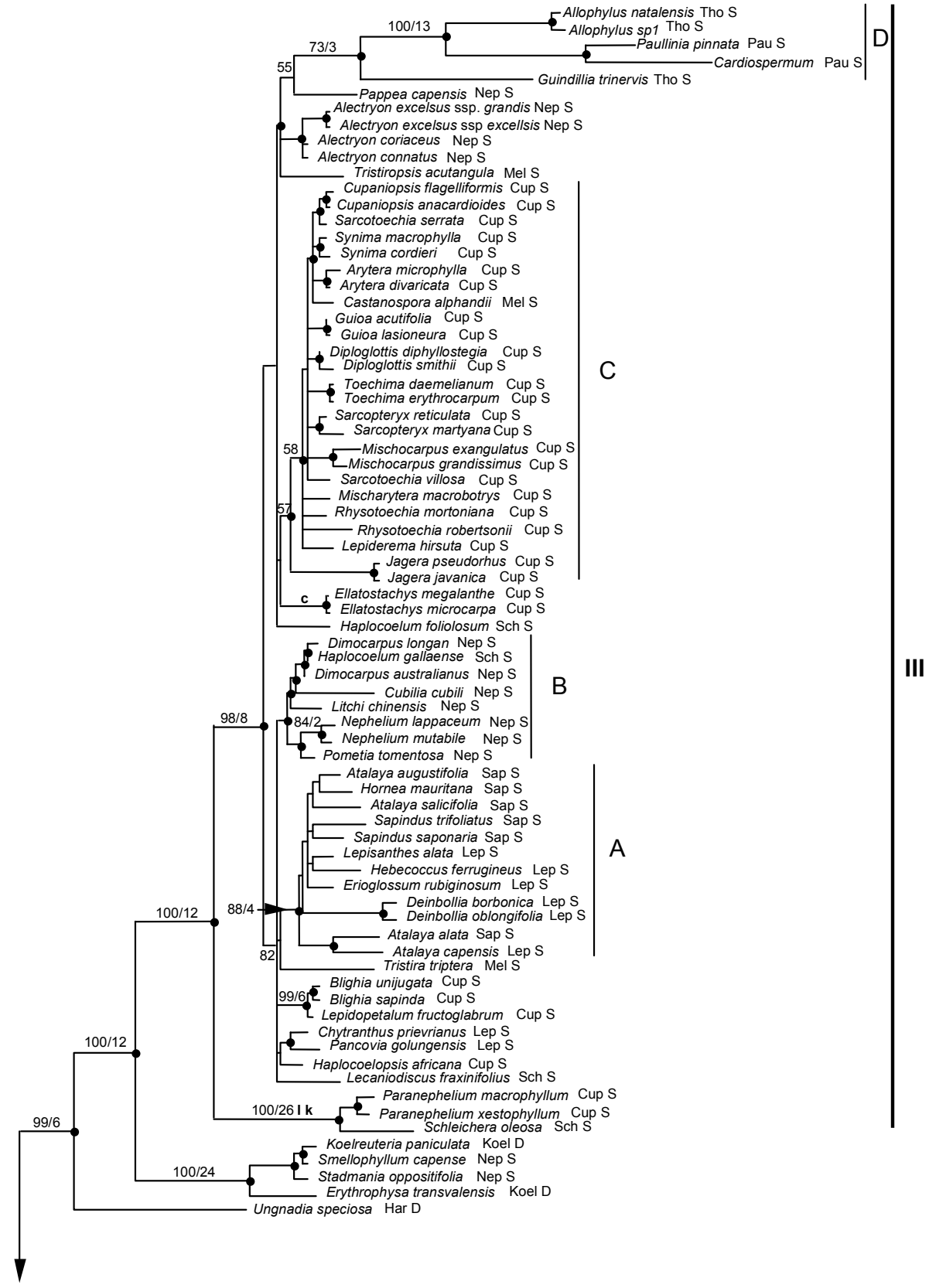


Figure 2.2 (continued over page): Consensus tree with branch lengths obtained from the Bayesian *matK* analysis. Support values and other labelling as in fig. 2.1. Branches on which informative insertion/deletion events a-j have arisen are marked.



Clades I and II previous page

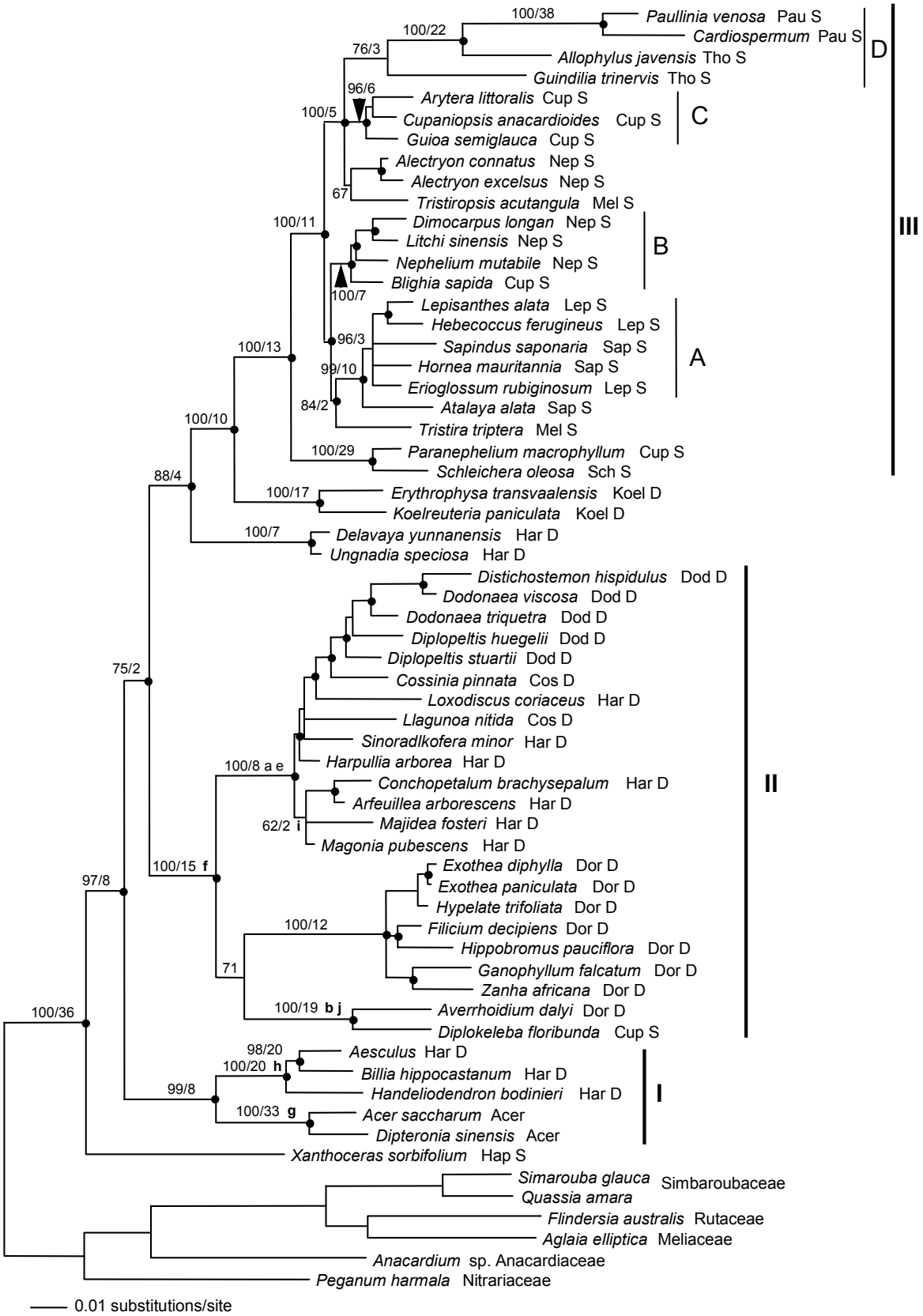

100/38
Paullinia venosa Pau S
100/22
Cardiospermum Pau S
D
76/3
Allophylus javensis Tho S
Guindilia trinervis Tho S
96/6
Arytera littoralis Cup S
100/5
Cupaniopsis anacardioides Cup S
C
Guioa semiglauca Cup S
Alectryon connatus Nep S
Alectryon excelsus Nep S
67
Tristiropsis acutangula Mel S
100/11
Dimocarpus longan Nep S
Litchi sinensis Nep S
B
Nephelium mutabile Nep S
Blighia sapida Cup S
100/7
Lepisanthes alata Lep S
Hebecoccus ferugineus Lep S
96/3
Sapindus saponaria Sap S
100/13
A
99/10
Hornea mauritannia Sap S
Erioglossum rubiginosum Lep S
84/2
Atalaya alata Sap S
100/10
Tristira triptera Mel S
100/29
Paranephelium macrophyllum Cup S
Schleichera oleosa Sch S
III
88/4
100/17
Erythrophysa transvaalensis Koel D
Koelreuteria paniculata Koel D
100/7
Delavaya yunnanensis Har D
Ungnadia speciosa Har D
Distichostemon hispidulus Dod D
Dodonaea viscosa Dod D
Dodonaea triquetra Dod D
Diplopeltis huegelii Dod D
75/2
Diplopeltis stuartii Dod D
Cossinia pinnata Cos D
Loxodiscus coriaceus Har D
Llagunoa nitida Cos D
Sinoradlkofera minor Har D
Harpullia arborea Har D
100/8 a e
Conchopetalum brachysepalum Har D
II
Arfeuillea arborescens Har D
62/2 i
Majidea fosteri Har D
Magonia pubescens Har D
100/15 f
Exothea diphylla Dor D
97/8
Exothea paniculata Dor D
Hypelate trifoliata Dor D
100/12
Filicium decipiens Dor D
Hippobromus pauciflora Dor D
71
Ganophyllum falcatum Dor D
Zanha africana Dor D
100/36
100/19 b j
Averrhoidium dalyi Dor D
Diplokeleba floribunda Cup S
98/20
Aesculus Har D
100/20 h
Billia hippocastanum Har D
I
99/8
Handeliodendron bodinieri Har D
100/33 g
Acer saccharum Acer
Dipteronia sinensis Acer
Xanthoceras sorbifolium Hap S
Simarouba glauca Simbaroubaceae
Quassia amara
Flindersia australis Rutaceae
Aglaia elliptica Meliaceae
Anacardium sp. Anacardiaceae
Peganum harmala Nitrariaceae
0.01 substitutions/site

Figure 2.3 (previous page): Consensus tree with branch lengths obtained from the Bayesian analyses of the combined *rbcL* and *matK* data. Support values and other labelling as in fig. 2.1. Branches on which informative insertion/deletion events a-j (but not c) have arisen are marked.

Phylogenetic Analyses

The heuristic analysis of the *rbcL* data returned 20,000 equally parsimonious trees of 1169 steps, in a single island, with a consistency index (CI) excluding uninformative characters of 0.34, and a retention index (RI) of 0.68. The consensus tree with branch lengths obtained from the Bayesian analysis, with bootstrap support (>50%), decay (≥1), and PP (≥95%) values on the branches, is shown in fig. 2.1. Three major clades that are retrieved in all other analyses (see below) are labelled I, II and III, in fig. 2.1. Much of the topology receives only weak support (<70% bootstrap), but there is moderate support for *Xanthoceras sorbifolium* as sister to all ingroup representatives (71% bootstrap, +4 decay, 98% PP), and also moderate support for the three major clades (clade I: 64%, +1, 97%; clade II: 79%, +2, 100%; clade III: 75%, +3, 99%). There is strong support for *Delavaya yunnanensis* as sister to *Ungnadia speciosa* (100%, +4, 100%), and for the two species of *Erythrophysa* as sisters to the two species of *Koelreuteria* (100%, +8, 100%), however, there is no support for their relationships to the other clades or for relationships between the major clades.

Phylogenetic analysis of the *matK* dataset yielded 20,000 most parsimonious trees of 2019 steps (CI = 0.52 excluding uninformative characters, RI = 0.84). The consensus tree from the Bayesian analysis, with branch lengths and support values on the branches as above is shown in fig. 2.2. When the informative indels are mapped on the trees (figs. 2.2-3) the distribution of all are consistent with a single origin, and each provides support for some element in the topology. The three major clades, I-III, can be discerned, and there is weak support (59%, +1, 98%) for a sister relationship between clade II and III. The relationship of *Xanthoceras sorbifolium* to the major clades is uncertain as it is part of a trichotomy that includes clade I and the combined clades II and III, however, in the strict consensus tree from the parsimony analyses (tree not shown) it is placed sister to all remaining ingroup taxa including Aceraceae and Hippocastanaceae. In clade I, Hippocastanaceae plus *Handeliodendron bodinieri*

(Dodonaeoideae) are robustly grouped with Aceraceae (100%, +12, 100%). Clade II is highly robust (100%, +15, 100%) and contains taxa assigned to Dodonaeoideae and *Diplokeleba floribunda* (Cupanieae). Clade III is strongly supported (99%, +6, 100%), and includes all taxa assigned to Sapindoideae, along with *Erythrophysa transvaalensis*, *Koelreuteria paniculata*, and *Ungnadia speciosa* from Dodonaeoideae. The latter taxon is sister to a more robust clade (100%, +12, 100%) comprising all other members of clade III, and this in turn consists of two highly robust lineages, one in which *Erythrophysa transvaalensis* and *Koelreuteria paniculata* are grouped with *Smelophyllum capense,* and *Stadmania oppositifolia* (100%, +24, 100%), and the other comprising all remaining Sapindoideae (100%, +12, 100%). *Paranephelium* and *Schleichera oleosa* are grouped as sister to remaining Sapindoideae, as in fig. 2.1, but with much improved support (100%, +27, 100%). The latter is a large and well-supported (98%, +8, 100%) but weakly resolved clade, within which can be discerned four subclades (A-D) that receive a degree of support and are also identifiable in fig. 2.1. The relationships between these subclades and several other taxa (*Alectryon*, *Elattostachys*, *Haplocoelum foliolosum*, and *Tristiropsis acutangula*) are unresolved (fig. 2.2).

It can be seen that there is considerable congruence between the topologies in figs. 2.1 and 2.2, and the partition homogeneity test indicated no significant conflict in the phylogenetic structure exhibited by the two datasets (P = 0.18). Analysis of the combined dataset yielded 990 most-parsimonious trees of 2431 steps (CI = 0.5 excluding uninformative characters, RI = 0.78). The Bayesian consensus tree with branch lengths (fig. 2.3) resolves *Xanthoceras* as sister to all ingroup representatives, and within the latter three very similar major clades (I-III) can be identified. Clade I is highly robust (99%, +8, 99%), has identical topology to fig. 2.2, and is the moderately supported (75%, +2, 100%) sister to remaining ingroup taxa. Twenty genera from Dodonaeoideae form a strongly supported clade II (100%, +15, 100%, indel *f*). The sister relationship between clades II and III decays at +2 steps, and is weakly supported (75% bootstrap) in the parsimony analysis, but is strongly supported in the Bayesian analysis (100% PP). Once again, species assigned to Koelreuterieae (*Erythrophysa* and *Koelreuteria*) and two monotypic genera of Harpullieae (*Ungnadia* and *Delavaya)* do

not group with the other representatives of Dodonaeoideae, being placed as successive sisters to the rest of clade III with moderate (88%, +4, 100%) and very strong (100%, +10, 100%) support, respectively. There is robust support for the grouping of remaining Sapindoideae, and for its two constituent lineages. Within the larger of these, the equivalents of subclades A-D can be identified, each of which receives good support (fig. 2.3). The placement of the remaining taxa (*Alectryon*, *Tristira tritera*, and *Tristiropsis acutangula*) receives much less support.

Discussion

Circumscription of Sapindaceae

Recent phylogenetic hypotheses based on pollen characters, foliar chemical screening, and morphology have included *Acer* and *Dipteronia* (Aceraceae), and *Aesculus* and *Billia* (Hippocastanaceae) within a broadly defined Sapindaceae (Muller and Leenhouts 1976; Umadevi and Daniel 1991; Judd et al. 1994). Phylogenetic trees based on DNA sequence data for a small number of taxa from Sapindaceae s.l. (Savolainen et al. 2000b; Soltis et al. 2000) have supported this hypothesis, placing the monotypic northern Chinese *Xanthoceras sorbifolium* as sister to all other members of the family (including Aceraceae and Hippocastanaceae). The *rbcL*, *matK* and combined analyses of Harrington and Gadek (2005) using datasets containing far fewer representatives of Dodonaeoideae than contained in these analyses, resulted in differential placement of *Xanthoceras sorbifolium* presumably due to long branch attraction (Felsenstein, 1978; Hendy and Penny, 1989).

In the *rbcL* analysis presented here, rooted using the sister family Malvales and including multiple representatives from each of the other eight families of the Sapindales, *Xanthoceras sorbifolium* receives moderate support as sister to all ingroup representatives (71% bootstrap, +4 decay, 98% PP - fig. 2.1), while there is no support for this relationship from the *matK* analysis (fig. 2.2). The combined analysis provides strong support (97%, +8, 100%) for *Xanthoceras* being the first lineage to diverge within Sapindaceae s.l. (fig. 2.3). It is interesting to note that *Xanthoceras* has a number of morphological features that are rare or unusual in the Sapindaceae s.s., such as

deciduous leaves, an atypical gynoecium with 7-8 ovules per locule, and unique orange hornlike appendages on the disk gland.

The analyses presented here reinforce the close relationship between Aceraceae, Hippocastanaceae, and Sapindaceae s.s. implicit in the broader taxonomic concept of Sapindaceae s.l. adopted by recent workers. It also provides evidence for the inclusion of both within Sapindaceae s.l., but this conclusion relies on the placement of a single taxon (*Xanthoceras*) among the 137 sampled. Even given the near complete sampling of genera in Dodonaeoideae, while there is still the possibility of long branch attraction being a confounding force in this placement, the opposite conclusion – that Aceraceae and Hippocastanaceae are distinct, monophyletic families easily distinguished from nearly all Sapindaceae cannot be ruled out on this basis alone. Nevertheless, support for the basal relationships between the constituent main clades in our analyses is much weaker than the overall support for Sapindaceae s.l. (even when analysed with a broad range of potential sister families). Hence, the alternative to the presently accepted broad family concept would appear to be the recognition of four families (or at least three if Clade I became a single family), one of which would be monotypic, if the new concepts are to be highly robust in the face of further evidence. It is my opinion that a single family recognising the highly robust relationships between all these lineages is preferable, with the major lineages recognised at the subfamily level. If this generally accepted view is to be maintained then it is important to note that Aceraceae Durande (1782) antedates Sapindaceae Jussieu (1789), so unification of the families and retention of the name Sapindaceae would require a 'superconservation' proposal to amend the Taxonomic Code (Turland and Watson, 2004).

Dodonaeoideae, excluding *Xanthoceras*, *Ungnadia*, *Delavaya*, *Erythrophysa* and *Koelreuteria*, is monophyletic in all analyses. A monophyletic Sapindoideae (including the last four genera) appears as derived within Sapindaceae. Our analyses also show that a core set of genera from the tribes of Radlkofer (1890, 1933) and Muller and Leenhouts (1976) are often retrieved in clades, along with taxa that have previously been placed elsewhere.

Radlkofer (1933) and Muller and Leenhouts (1976) considered both these families to be related to the Harpullieae (Dodonaeoideae), whereas Umadevi and Daniel (1991), after a chemosystematic screening of 19 taxa, submerged Hippocastanaceae in Sapindoideae and placed Aceraceae in its own subfamily (table 2.1). Wolfe and Tanai (1987) suggested a close relationship between Aceraceae and Paullinieae (Sapindoideae) because both have multiseriate rays and nonseptate fibres in their wood. Based on the results of a phylogenetic analysis of 27 morphological characters for 19 taxa, Judd et al. (1994) concluded that Aceraceae are related to the more derived tribes of Sapindoideae, but they, like Radlkofer (1933) and Muller and Leenhouts (1976), regarded the Hippocastanaceae as more closely related to Harpullieae. Furthermore, their results indicated that *Aesculus*, *Billia* (Hippocastanaceae), and *Handeliodendron* (Dodonaeoideae) form a monophyletic group derived within Sapindaceae s.l.

In the analyses presented here representatives of all four genera from Aceraceae and Hippocastanaceae, plus the monotypic *Handeliodendron* (Dodonaeoideae/Harpullieae), form a highly robust clade in which Aceraceae (*Dipteronia* and *Acer*) are sister to Hippocastanaceae (*Aesculus* and *Billia*) plus *Handeliodendron* (figs. 2.2 and 2.3). The position of the last taxon is strongly supported within this clade by all support measures and by a synapomorphic indel (figs. 2.2-3: indel *h*), and is congruent with its placement in the non-molecular analyses of Judd et al. (1994) and Forest et al. (2001). The temperate *Handeliodendron bodinieri* is deciduous small trees or shrubs from the Guangxi and Guizhou districts of central China. A potential relationship of *Handeliodendron* with Hippocastanaceae was first alluded to in the description of the genus by Rehder (1935), where he suggested its opposite 5-foliate, palmately-compound leaves resemble those of Hippocastanaceae. His placement of *Handeliodendron* in the heterogeneous Sapindaceae s.s. (Rehder 1935) resulted from a reluctance, prevalent at the time, to enlarge a tightly circumscribed family like Hippocastanaceae (Judd et al. 1994). The analysis of Forest et al. (2001) supported the possession of opposite leaves as a synapomorphy for the *Handeliodendron*/Hippocastanaceae clade; this feature is also found rarely in Sapindaceae in the monotypic *Guindilia*, and some species of *Matayba*. Pollen features

can be used to characterise the group, with *Handeliodendron*/Hippocastanaceae having colporate pollen with distinctive verrucae on the colpus membranes (Adema et al., 1994). *Handeliodendron* also has wood anatomy nearly identical with that of *Aesculus* (Klaassen 1999).

There is no support in these analyses for a relationship between representative species of *Dipteronia*, *Acer* and taxa from Paullinieae/Thouinieae (Sapindoideae) as suggested by Wolfe and Tanai (1987) based on wood anatomy, and Judd et al. (1994) on the basis of similarities in fruit morphology. These results indicate that characters such as multiseriate wood rays, nonseptate fibres, and winged schizocarpic fruits are either incorrectly interpreted or represent parallelisms.

The strong affinities of this group merits recognition at the subfamily level, as Hippocastanoideae Burnett, and the two subclades constitute tribes Acereae and Hippocastaneae.

Dodonaeoideae

Of the representatives assigned to Dodonaeoideae (= Dyssapindaceae), which have two or more apotropous and upright ovules per locule (or rarely one, which is epitropous and hanging), four genera *Ungnadia*, *Delavaya* (Harpullieae) and *Erythrophysa*, *Koelreuteria* (Koelreuterieae) have a closer relationship with other members of the family placed in clade III (figs 2.2-3) and should be excluded from Dodonaeoideae. Within Dodonaeoideae Radlkofer (1895, 1933) considered the tribes Koelreuterieae, Cossinieae and Dodonaeeae to be closely linked based on their possession of a well developed terminal leaflet, and thin convolute cotyledons. Within Koelreuterieae he had assigned three genera (*Koelreuteria*, *Erythrophysa*, and *Stocksia*), while a further newly described genus (*Sinoradlkofer*) was "provisionally" placed in the tribe by Meyer (1977). All genera have been sampled except the monotypic *Stocksia brahuica* from Iran, Afghanistan and Pakistan. This is the only species of Sapindaceae to have thorns which Metcalfe and Chalk (1950) suggest "represent reduced axillary branches". The analyses presented here (figs. 2.1-3) and a similarity in a number of pollen features (Buijsen et al. 2003) clearly indicate that *Sinoradlkofer* belongs in Harpullieae, and that the remaining genera including *Smellophyllum* and *Stadmannia* should be considered as a separate tribe in Sapindoideae (see below).

In the analyses, Dodonaeoideae includes the 20 genera placed in clade II (figs. 2.1-3), one of which was previously assigned to Sapindoideae (*Diplokeleba)*, with all the others previously assigned to the subfamily: *Arfeuillea*, *Averrhoidium*, *Conchopetalum*, *Cossinia*, *Diplopeltis*, *Distichostemon*, *Dodonaea*, *Exothea*, *Filicium*, *Ganophyllum*, *Harpullia*, *Hippobromus*, *Hypelate*, *Llagunoa*, *Loxodiscus*, *Magonia*, *Majidea*, *Sinoradlkofera*, and *Zanha*. Various subsets of these taxa were grouped in each of the analyses, strongly supported by PP, bootstrap and Bremer analysis, and a synapomorphic indel (figs. 2.2-3: indel f). These taxa include representatives of the other four tribes assigned to the subfamily (Cossinieae, Dodonaeeae, Doratoxyleae, and Harpullieae). Analyses of all datasets show that the taxa split into three strongly supported lineages:

i) A combined Cossinieae/Dodonaeeae/Harpullieae comprising genera with dehiscent fruits (*Arfeuillea*, *Conchopetalum*, *Cossinia*, *Dodonaea*, *Diplopeltis*, *Distichostemon*, *Harpullia*, *Llagunoa*, *Loxodiscus*, *Magonia*, *Majidea*, and *Sinoradlkofera* - except for some species of *Diplopeltis* in which the fruit is a schizocarp with indehiscent cocci).

ii) Doratoxyleae with indehiscent fruits (*Exothea*, *Filicium, Ganophyllum, Hippobromus*, *Hypelate* and *Zanha*).

iii) The unexpected combination of *Diplokeleba floribunda* with *Averrhoidium.*

COSSINIEAE/DODONAEEAE/HARPULLIEAE

Harpullieae was considered by Muller and Leenhouts (1976) to be a heterogeneous assemblage, with several genera (*Xanthoceras* and *Ungnadia*) being "difficult to connect" to others in the tribe based on the strikingly different pollen morphologies displayed by *Harpullia, Xanthoceras, Ungnadia,* and *Magonia*. Radlkofer (1933) originally divided Harpullieae into two groups, one containing taxa without a true terminal leaflet (*Conchopetalum*, *Harpullia*, *Magonia*, and *Arfeuillea*), and the other with a small true terminal leaflet (*Xanthoceras*, *Delavaya*, and *Ungnadia*). The analyses indicate the tribe, as previously defined, is polyphyletic. However, if the genera with the terminal leaflet are excluded (plus *Handeliodendron*), the analyses presented here indicate that genera previously assigned to Cossinieae and Dodonaeeae are derived within a monophyletic Harpullieae (also including *Sinoradlkofera*). The combined analyses provide weak support for two distinct groupings of genera within this clade.

The smaller clade (62%, +2, 99% and indel i) contains four genera - *Conchopetalum* (two species endemic to Madagascar), *Majidea* (two species from tropical Africa and Madagascar), monotypic *Magonia* from the savannas of Brazil, Bolivia and Paraguay and monotypic *Arfeuillea* from Laos and Thailand. *Magonia* is the only member of Sapindaceae to have tetrad pollen grains. Muller and Leenhouts (1976) suggested that *Conchopetalum* was the most primitive genus within Harpullieae, with *Majidea* derived from it; however, further data are required to resolve relationships within this clade.

The other larger clade (97% PP) contains a number of unplaced genera and also includes both genera of Cossinieae that were sampled, however, there is no support in any of the analyses for a monophyletic grouping of *Cossinia* (four species – Mauritius, eastern Australia, New Caledonia and Fiji) with *Llagunoa* (three species from tropical South American highlands). The combined dataset indicates support (99%, 97%) for *Cossinia* as sister to Dodonaeeae (minus *Loxodiscus*).

Representatives from three of the four genera assigned to Dodonaeeae (*Dodonaea, Diplopeltis,* and *Distichostemon*) are placed in a monophyletic group within clade II in the *matK* and combined analyses (figs. 2.2-3). The combined analysis places *Loxodiscus* sister to all other Dodonaeeae plus *Cossinia*. In the main, this tribe consists of non-rainforest shrubs, predominantly restricted to Australia except for the monotypic *Loxodiscus* from New Caledonia. It was delimited by Radlkofer (1933) on the imparipinnate or simple leaves, convolute cotyledons, and dry papery capsules found in most members. As mentioned previously, the exceptions are the schizocarps of *Diplopeltis*, and the few species of *Dodonaea* that have paripinnate leaves (West, 1984). A comprehensive analysis of relationships within this tribe is contained in following chapters.

The only other genus of Harpullieae not sampled in this study, the monotypic Chinese endemic *Eurycorymbus cavaerleriei* (Levl.) Rehd. et Hand.-Mazz, is also considered to be as "primitive" as *Conchopetalum* (Buijsen et al. 2003). An unpublished plastid dataset (trn*L-F* – Sven Buerki pers. com.) places *E. cavaerleriei* firmly (100% PP) inside the Cossinieae/Dodonaeeae/Harpullieae clade, but relationships to other genera was unresolved.

DORATOXYLEAE

Doratoxyleae includes species with non-dehiscent, more or less fleshy fruits (Muller and Leenhouts 1976), except for the four species of *Averrhoidium* that have an unusually dehisced fruit in which the seed stays dangling from the placenta (Weckerle and Rutishauser, 2003). Seven of the nine genera assigned to this tribe have been included in this study and constitute a well-supported clade in all analyses that is sister to the Cossinieae/Dodonaeeae/Harpullieae complex. There is little resolution of relationships between the genera, with only some support for sister relationships for *Filicium* and *Hippobromus* (59%, 99%) and *Ganophyllum* and *Zanha* (69%, 98%). The strict consensus tree of a recently published morphological cladistic analysis of 34 characters places the *Doratoxylon* Thouars ex Hook.f. (which is not included in these analyses) as sister to *Zanha* and *Ganophyllum*, while *Euchorium* Ekman &Radlk. received 56% BS as sister to *Hippobromus* (Weckerle and Rutishauser, 2003).

DIPLOKELEBA FLORIBUNDA-AVERRHOIDIUM

This group is strongly supported in all analyses including two synapomorphic indels (b, j, fig. 2.2-3), but there is only weak support for this clade as sister to Doratoxyleae. The monotypic South American genus *Diplokeleba* has had a long and confused history in Sapindaceae, with there even being some thought that it may not belong there at all (van der Ham, 1990). The genus was originally assigned to Melicocceae by Radlkofer (1895), who in turn removed it to Cupanieae where it was considered to belong with a grouping of other South American genera of Cupanieae lomatorrhizae. *Diplokeleba* has largely winged seeds (also found in *Magonia*) which are also unique for Sapindaceae.

The neotropical *Averrhoidium* shares features in common with Doratoxlyeae including the characteristic six ovules in the gynoecium (*Diplokeleba* – 3-carpellate, with single ovule per carpel). However, *Averrhoidium* has the novel feature for Doratoxyleae of dehiscent fruits, and seeds with sarcotesta. Apart from a relatively similar geographical distribution there does not appear to be any macromorphological synapomorphies that currently defines the relationship of these two genera.

Sapindoideae

The analyses provide strong support (100%, +12, 100%, fig. 2.2; 100%, +10, 100%, fig. 2.3) for the delineation of a monophyletic assemblage that includes multiple representatives of all nine tribes of Radlkofer (1890, 1933) and all eight tribes recognised by Muller and Leenhouts (1976), and also *Koelreuteria* and *Erythrophysa*. There is more modest support (88%, +4, 100%) for the inclusion of the monotypic genera *Delavaya* and *Ungnadia* in this group (clade III, fig. 2.3). All save the last four genera are united in having a single ovule per locule. All four genera have traditionally been placed in the Dodonaeoideae because of their having two ovules per locule. Only one ovule in each locule, however, reaches maturity in *Koelreuteria*, indicating that this genus may represent a condition intermediate between the states found in Dodonaeoideae and Sapindoideae (Ronse Decraene et al., 2000).

Ungnadia speciosa, a taxon restricted to northern Mexico and Texas, is placed in clade III as sister to Sapindoideae plus *Koelreuteria* and *Erythrophysa* in both the *matK* and combined analyses (figs. 2.2-3), while *Delavaya* is strongly grouped with it in the *rbcL* and combined analyses. *Delavaya* and *Ungnadia* were grouped on their prolonged basal petal appendages and glabrous stamens in the analysis of Judd et al. (1994), and the wood anatomies of the two genera have been shown to differ from other tropical Harpullieae, but are nearly indistinguishable from Cupanieae (Sapindoideae), because of their scarce paratracheal parenchyma, smaller intervessel pits and the presence of crystals in fibrous elements (Klaassen 1999).

Koelreuteria is from southern China, and was previously placed in the Koelreuterieae, along with *Stocksia* from eastern Iran/Afghanistan, and *Erythrophysa* from Somalia/Madagascar, on the basis of their imparipinnate leaves, convolute cotyledons, zygomorphic flowers and inflated-membranous capsules (Radlkofer 1933). Analysis of the *matK* dataset strongly links *Koelreuteria* and *Erythrophysa* with *Smelophyllum* and *Stadmania*. This is surprising because the last two taxa have been placed in Nephelieae (Sapindoideae; Radlkofer 1933) on the ground that they have a fleshy aril surrounding the seed; a relationship with Koelreuterieae has never been suggested. *Smelophyllum* is a South African monotypic genus, while *Stadmania* contains five species endemic to Madagascar, and another that also extends to mainland

Africa and the Mascarenes. Neither *Smelophyllum* nor *Stadmania* has an "inflated fruit", a character found in Koelreuterieae, *Delavaya,* and *Ungnadia* (Judd et al. 1994); however, they do have some morphological similarities with *Koelreuteria*, viz. trichomes on the anthers, and a resemblance in pollen type, which is three colporate, spheroidal to subprolate, with striate ornamentation (Nowicke, 1976; van der Ham, 1990). I conclude, therefore, that Sapindoideae should be expanded to include Koelreuterieae, *Delavaya* and *Ungnadia*.

The next diverging lineage within Sapindoideae is a clade of southeast Asian taxa comprising *Paranephelium* and the monotypic *Schleichera*, which is sister to the remaining taxa in all analyses. Each genus has been previously assigned to a separate tribe, Cupanieae and Schleichereae, respectively. *Paranephelium* is unusual in Sapindoideae because of its imparipinnate leaves (also found in a few species of *Lepisanthes*) and warty to spiny capsular fruits with seeds without an arillode, whereas *Schleichera* has paripinnate leaves, and fruits which are a dry berry with seeds covered by an arillode. Both genera, however, have a very similar wood anatomy (Cupanieae type IVb – Klaassen 1999) with uniseriate rays, and crystals in both the axial fibrous elements and ray cells.

Discussion of the remaining Sapindoideae will essentially concentrate on the clades that can be identified across all analyses (A-D, figs. 2.1-3).

CLADE A (LEPISANTHEAE–SAPINDEAE

This clade represents the main grouping of Muller and Leenhouts' (1976) group A taxa, that was principally characterised by their paripinnate leaves and the absence of an aril, with each tribe further distinguished on geographical distribution (American-Australasian Melicocceae, African-Asian Lepisantheae, and pantropical Sapindeae). The clade is retrieved in all analyses and includes a core group comprising Asian representatives of Lepisantheae (*Lepisanthes*, *Erioglossum*, and *Hebecoccus*) and African-Australasian Sapindeae (*Sapindus*, *Deinbollia*, *Hornea* and *Atalaya*) that receives some support (figs. 2.2). In addition, there is moderate support from the combined analysis for the inclusion of *Tristira* (Melicocceae) within clade A (fig. 2.3). None of the other three Melicocceae sampled (*Castanospora alphandii*, *Talisia nervosa*, and *Tristiropsis acutangula*) shows a strong relationship to each other. Placement of the

monotypic Australian *Castanospora* with the Australian Cupanieae in clade C on the *matK* data (fig. 2.2) is in line with the close similarity of its syncolporate pollen to that of *Diploglottis* (Klaassen 1999). The relationships of the American Melicocceae (*Talisia*) appear to lie with clade B, *Alectryon* (Nephelieae), and *Tristiropsis* (Melicocceae), although there is no support for this grouping (fig. 2.1), and the last two taxa are strongly grouped within clades C and D in the combined analysis (fig. 2.3). The only other genus of American Melicocceae (*Melicoccus*) although not sampled here, is closely related to *Talisia* based on similarity of fruit morphology (Acevedo-Rodriguez, 2003). The African Lepisantheae sampled (*Pancovia* and *Chytranthus*) differ from the Asian Lepisantheae in their slightly zygomorphic flowers, and entire rather than lobed stigma. They are weakly grouped with *Haplocoelopsis* (African Cupanieae) in an unresolved polytomy with clades A and B (fig. 2.2).

All the sampled Sapindeae (*Atalaya*, *Sapindus, Deinbollia* and *Hornea* a Mauritian monotypic endemic) are included in clade A (fig. 2.2). Three other American genera in this tribe have yet to be sampled (*Thouindium*, *Toulicia*, and *Porocystis*). Several representatives of *Atalaya,* including two endemic species from Australia and two southern African endemic, were included in the *matK* analysis but do not form a monophyletic group. Rather, the Australian species appear closer to species of *Sapindus*, *Lepisanthes*, *Hebecoccus*, and *Erioglossum* than to the African species. A more extensive analysis of this genus is currently underway.

CLADE B (NEPHELIEAE)

Representatives from Nephelieae (*Cubilia, Dimocarpus, Litchi*, *Nephelium*, and *Pometia*), plus a single species of Schleichereae (*Haplocoelum gallaense*), are placed in clade B (figs 2.1-3), and along with *Lepidopetalum* (Australian Cupanieae), *Tristira triptera* (Malesian Melicocceae), and a range of African taxa of the Cupanieae (*Blighia, Haplocoelopsis*), Lepisantheae (*Chytanthus*, *Pancovia*), and Schleichereae (*Haplocoelum foliolosum*, *Lecaniodiscus*) are identified as a potential sister lineage to clade A (figs. 2.2 and 2.3). There is strong support for the inclusion of *Blighia* in clade B on the *rbcL* (fig. 1; 88%, +3, 100%) and combined data (fig. 3; 99%, +8, 100%), but this may be a reflection of the lower taxon densities in these analyses. Muller and Leenhouts (1976) proposed two natural groups within Nephelieae – the *Dimocarpus* and

Pappea groups with indehiscent and dehiscent fruits, respectively. Nine of the 14 genera in Nephelieae have been sampled, but only the five listed above are identified here as core Nephelieae, and all belong to the *Dimocarpus* group. Representatives of the *Pappea* group (*Alectryon*, *Pappea*, *Stadmania*, and *Smelophyllum*) are not resolved as a clade: the last two are strongly grouped with *Koelreuteria* (see above), while the first two are placed near clade C and D (figs. 2.2 and 2.3). The separation of the two species of *Haplocoelum* in the *matK* analysis suggests a need to reassess the generic concept. A study of wood anatomy revealed large infrageneric variation in almost all features measured, to the extent that the boundaries of the genus were obscure (Klaassen 1999).

The Schleichereae was placed with Nephelieae and Cupaneae in group B by Muller and Leenhouts (1976). Only one species of the three genera of Schleichereae sampled was placed within clade B (*Haplocoelum foliolsum*). The Schleichereae comprises mostly African and Madagascan taxa, with the monotypic *Schleichera* being the only genus in Asia. As noted above, the other two African representatives of the tribe were placed within the large polytomy (fig. 2.2) with *Lecaniodiscus* near clades A and B, but the Asian member was robustly placed in the *Paranephelium* clade, reflecting the geographic separation. Further sampling of the five Madagascan genera is needed to clarify relationships within African members of this tribe.

CLADE C (CUPANIEAE)

The vast majority of the taxa attributed to Cupanieae (14 of the 18 genera sampled) are placed in this clade (fig. 2.2), which receives strong support (98%, +7, 100%) in the combined analysis (fig. 2.3). There is no support for the close relationship between these taxa and the *Dimocarpus* group of Nephelieae (placed in clade B) suggested by Muller and Leenhouts (1976). Rather, there is strong support (98%, +5, 100%) for their being closer to Paullinieae/Thouinieae in clade D and *Alectryon* and *Tristiropsis* (fig. 2.3). However, all three African Cupanieae sampled (*Blighia, Haplocoelopsis,* and *Paranephelium*) are placed outside clade C, and there is some evidence to support a link between clade B Nephelieae and *Blighia,* at least (see above).

The Cupanieae is the largest tribe within Sapindaceae s.s., with a pantropical distribution and being distinguished by the presence in many genera of an aril and capsular fruit. It contains approximately 46 genera divided into seven groups on a

number of putatively derived macromorphological characters (Muller and Leenhouts 1976). Apart from the three African genera and one of the Australian representatives (*Lepidopetalum xylocarpum*), all members of the tribe sampled are placed within one clade, and are recognised as core Cupanieae. All but one of these are from Australia and Asia, the exception being *Matayba* sp., from South America. This was included in the *rbcL* analysis, and was grouped with *Toechima tenax* inside clade C with good support (97%, +3, 100%). Poor internal resolution within this clade in the *matK* analysis, and limited taxon representation in the combined analysis, preclude inferences of relationships between members of this clade. Further investigation of relationships within Cupanieae incorporating a far wider sampling regime is currently been undertaken.

CLADE D (PAULINIEAE-THOUINIEAE)

Three of the seven genera attributed to Paullinieae have been sampled, and four of the six in Thouinieae. The strong support (100%, +4, 100%, fig. 2.1) for the monophyly of the former (*Cardiospermum*, *Paullinia*, and *Serjania*) in all analyses is consistent with recognition of tribe Paullinieae. Thouinieae (*Allophylus*, *Bridgesia*, *Diatenopteryx*, and *Guindilia*) are not monophyletic in any analysis (figs. 2.1-3). However, these two tribes together constitute the Nomophyllae (Radlkofer 1933), and group C of Muller and Leenhouts (1976; see table 2.1). The monophyly of this larger taxon set is supported by both the *matK* and combined analyses (figs. 2.2 and 2.3), and is resolved without support in the *rbcL* analysis (fig. 2.1). These results are congruent with those of a cladistic analysis of 21 morphological and pollen characters of genera in the two tribes (Acevedo-Rodriguez, 1993) that showed a paraphyletic Thouinieae that included a monophyletic Paullinieae. Clade D is characterised by zygomorphic flowers, petals with a prominent hood or scale, a reduced unilateral disk, with the compound leaf having a terminal leaflet. The highly speciose Paullinieae are distinguished within this clade by all being tendril climbers, and the presence of stipules.

On the basis of morphological and pollen features, Muller and Leenhouts (1976) suggested that Nomophyllae contain the most derived tribes in Sapindaceae s.l.; members displaying specializations (such as diporate and triporate pollen) not found elsewhere in the family. These authors, and also van der Ham and Tomlik (1994), used

palynological characters to support the argument that Paullinieae originated from a Cupanieae-like ancestor. The molecular data indicate that Thouinieae and Paullinieae are closer to core Cupanieae (clade C, figs. 2.1-3) than any other members of Sapindaceae s.l., and are consistent with their hypothesis. I conclude that tribe Thouineae should be expanded to include all members of Paullinieae.

Conclusion

This study provides a new and more robust inference of phylogenetic relationships within Sapindaceae s.l. The major clades retrieved show a degree of congruence with some of the previous intra-familial groups erected on morphological and anatomical criteria, but there is no support for its division into two subfamilies along the lines of Muller and Leenhouts (1976). Following the classification concepts exposed in Chase et al. (2000), and acknowledging the shared evolutionary history of Sapindaceae, Hippocastanaceae and Aceraceae and the many morphological features that they have in common (Judd et al., 1994), I propose a four subfamily arrangement within Sapindaceae along the following lines:

Sapindoideae should be expanded to include Koelreuterieae, *Delavaya*, and *Ungnadia*;

Hippocastanoideae comprising two tribes: Acereae including *Acer* and *Dipteronia*; Hippocastaneae including *Aesculus, Billia*, and *Handeliodendron*;

Xanthoceroideae including the single genus *Xanthoceras*;

Dodonaeoideae to include the remaining taxa, within which two tribes Dodonaeeae (includes former Harpullieae and Cossinieae) and Doratoxyleae recognisable on fruits dehiscent/indehiscent are supported, with a third new tribe Diplokelebeae (includes *Diplokeleba* and *Averrhoidium*) requiring further morphological and molecular analyses.

Several of the tribes in Sapindoideae (Cupanieae, Lepisantheae, Melicocceae, Nephelieae, Sapindieae, Schleichereae, and Thouinieae) appear polyphyletic in these analyses, and several highly novel associations of genera have been found. There is, however, potential to redefine at least some existing tribes (particularly Cupanieae, Nephelieae, Sapindieae, and Thouinieae) to coincide with major clades identified here. An expansion of both the sequence database and the taxon sampling is needed to

establish the limits of clades and further test their robustness before tribal concepts can be redefined. Morphological synapomorphies for the supported groups remain to be established. This will require some testing of homologies and a detailed re-analysis of a non-molecular database.

Chapter 3: Tempo of evolution in *Acer* and *Aesculus* (Sapindaceae)

The maples, *Acer* L. (Sapindaceae), and buckeyes/horse-chestnuts, *Aesculus* L. (Sapindaceae), are iconic Northern Hemisphere tree genera that have a documented fossil record back to the Late Paleocene, and a postulated history to the Late Cretaceous (Wolfe and Tanai, 1987). They are considered to be Tertiary relict genera whose current disjunct distributions represent responses to climatic oscillations over the last 65 million years (Wen, 1999).

Generally members of *Acer* are temperate to subtropical, broad-leaved trees or shrubs with up to 156 species (de Jong, 2002), and along with *Aesculus* (~15 species), are important components of deciduous and evergreen forests of the Northern Hemisphere. Both genera have similar distributions being found throughout Asia, Europe and North America. A recent phylogenetic study has concurred with previous morphological assessments of a strongly supported sister relationship of *Acer* to the Chinese endemic genus *Dipteronia* Oliv. (2 species) (Harrington et al., 2005). These two genera are strongly supported as sister to a clade that consists solely of *Billia* Peyritsch (two species in the Neotropics), *Handeliodendron* Rehder (monotypic from China) and *Aesculus*, and all together make up Hippocastanoideae, a subfamily of Sapindaceae. The reciprocal monophyly of *Acer* and *Dipteronia* has not been supported in some phylogenetic studies (Pfosser et al., 2002; Tian et al., 2002; Li et al., 2006), but a recent six locus chloroplast dataset including 62 species of *Acer* strongly supports the monophyletic status of that genus (Renner et al., 2007).

The extensive fossil record of leaves and samaras attributed to *Acer* from western North America have been classified into a phylogenetic series that includes 27 sections and 91 extinct species, with over 50 species dating from the Paleocene and Eocene (Wolfe and Tanai, 1987), and there is also a multitude of fossils from Europe and eastern Asia (see Manchester, 1999 for references). The oldest records of Acer-like samaras are from the Late Paleocene (Crane et al., 1990). For *Dipteronia* fossil samaras date from the Late Paleocene to Oligocene from North America (McClain and Manchester, 2001), with no confirmed fossil records from Asia or Europe (Manchester and Tiffney, 2001). Fossil leaves, fruits and seeds attributed to *Aesculus* date from Late

Paleocene in North America (Manchester, 2001), with younger fossils reported from Asia and Europe (see Xiang et al., 1998a for references). There are no fossils reported for *Billia* or *Handeliodendron*.

The currently interpreted paleobotanical record indicates a minimum age for the *Acer*/*Dipteronia* split in the Paleocene ~64 million years ago (Mya) (Wolfe, 1981; Wolfe and Tanai, 1987; McClain and Manchester, 2001) and a similar age has been attributed for the most recent common ancestor (MRCA) of *Billia* and *Aesculus* (Manchester, 2001). The fossil attributed minimum Paleocene age for these genera forms the basis of several scenarios regarding the evolutionary and migratory history of the Hippocastanoideae through space and time across the Northern Hemisphere (Hardin, 1957; Kvaĉek, 1994; Boulter et al., 1996; Xiang et al., 1998a; Forest et al., 2001; Manchester, 2001; Manchester and Tiffney, 2001). Not only are there intercontinental disjunctions between the sister genera, morphological and molecular phylogenetic studies have identified similar disjunctions between sister species in multiple lineages within *Acer* (Wolfe, 1981; Hasebe et al., 1998; Grimm et al., 2006).

An assessment of the currently available but limited molecular evidence presents some conflict with the present interpretation of the fossil record. A three-gene estimation of divergence time for angiosperms (Wikstrom et al., 2001) using a single absolute internal fossil calibration with non-parametric rate smoothing (NPRS) and bootstrap resampling estimation of standard errors produced a maximum likelihood (ML) estimated age for the divergence of *Acer* from *Aesculus* (their node 139) of 23 Mya (SE ± 3) and the MRCA of Sapindales (their node 134) of 80 Mya (SE ± 4). Two other molecular dating studies have included members of Sapindaceae: i) a NPRS of an *rbcL* dataset of the flora from the Cape of South Africa using one fixed (eudicot pollen) and 11 minimum age constraints indicated a stem date for Sapindales ~86 Mya (no estimation of standard deviation given) (Forest et al., 2007); and ii) a penalized likelihood (PL) estimation of divergence times from four combined plastid genes indicated a divergence of Sapindales (represented by a single member of *Acer*) from Brassicales ~92 Mya (Magallón and Sanderson, 2001), which is similar to the age found in the angiosperm study (Wikstrom et al., 2001).

To discern between the potential contrasting reconstructions provided by the fossils and molecules particularly in regard to the Wikstrom et al. (2001) study, the evolutionary history of *Acer* and *Aesculus* is evaluated using a combined fossil based and Bayesian relaxed molecular clock estimations of branch lengths from two plastid genes (*rbcL*, *matK*) that includes 59 Sapindaceae and 31 outgroup sequences, and present evidence that molecular evolution within Sapindaceae potentially conflicts with aspects of the paleobotanical record.

Methods

Sampling and Sequencing

The combined plastid dataset (*rbcL*, *matK*) of Harrington et al. (2005) was expanded by a further 26 ingroup (Sapindaceae) taxa with near complete generic level sampling from the Dodonaeoideae, Hippocastanoideae and Xanthoceroideae and 25% of genera of Sapindoideae. Twenty-one outgroups were drawn from Harrington et al. (2005) or GenBank to include multiple representatives of eudicotyledons (eudicots) and two taxa each from Monocotyledons and Ceratophyllaceae. To determine whether the numbers of samples included in the combined dataset may be influencing the outcomes of analyses, two single gene datasets were also generated that included multiple representatives of *Acer* and *Aesculus*. One of these was an *rbcL* alignment of 1428 base pairs (bp) for 116 taxa which included two outgroups from Malvales and 14 representative genera from seven other families of Sapindales with 32 species of *Acer* and three species of *Aesculus*. The other was an aligned *matK* matrix consisted of 1697 bp with 123 taxa including 12 outgroups from multiple families of Sapindales and two species of *Acer* and eight species of *Aesculus*.

For all additional taxa, extraction of DNA, PCR amplification and sequencing protocols and editing follow those described or referenced in Harrington et al. (2005). Data matrices and trees are deposited in TreeBase (http://www.treebase.org) and are also included on accompanying CD. GenBank accession numbers are listed in appendix 3.1.

Molecular dating and phylogenetic analysis

Bayesian Markov chain Monte Carlo (MCMC) phylogeny and molecular dating estimations of the combined aligned matrix and separate gene analyses were performed using the BEAST package (Drummond and Rambaut, 2007), which is the only currently available method that co-estimates topology and branch lengths (substitution rates) directly from the nucleotide sequences using a Bayesian relaxed clock approach (Drummond et al., 2006).

Nucleotide substitution models for analyses were selected using the corrected Akaike Information Criterion (AICc) as implemented in MrAIC.pl 1.4 (Nylander, 2004), with the likelihood scores under different models estimated using PHYML (Guindon and Gascuel, 2003a). For both the *rbcL* and *matK* genes, the general-time-reversible (GTR) model with gamma distributed rate variation and an assumed proportion of invariable sites were selected. A series of analyses (described below) was performed for the combined and separate data partitions using these parameters. In addition, each gene was further partitioned by codon position with all parameters free to vary independently across codon positions. The uncorrelated relaxed clock model selected allows the rates in each branch to be independently drawn from an assumed log normal distribution with no *a priori* correlation of lineage substitution rates with that of its ancestor (Drummond et al., 2006). The tree branching prior was assumed to follow a traditional Yule speciation process birth rate (constant speciation rate per lineage) (Yule, 1924), and the monophyly of the ingroup was assumed *a priori* (Harrington et al., 2005). The degree of rate autocorrelation and 'clock-likeness' of the data was estimated during the analysis. The BEAST input files are on accompanying CD.

The output was examined using Tracer v1.3 (Rambaut and Drummond, 2004), and summarized (excluding the appropriate burn in) using Logcombiner v1.4.3, and TreeAnnotator v1.4.3 (all part of the BEAST package http://beast.bio.ed.ac.uk/). In evaluating the results of analyses, posterior probability values (PP) ≥95% are interpreted as indicating a degree of support from the molecular data for elements in the topology.

Because the aim of this study is to compare rates of molecular evolution with the paleontological record for Sapindaceae, a series of analyses was performed using different combinations of prior probability distributions as calibration priors based on

the fossil record and also calibrating the phylogeny with the ages on internal nodes being estimated in units of substitutions/site. To determine the relative influence of the joint priors (prior distribution) and joint priors plus the data (posterior distribution) on parameters of interest, all MCMC approximations were also performed without the sequence data (Drummond et al., 2006).

Due to the extremely long run times needed to achieve an effective sample size >200 for all designated parameters, in the initial analyses a pruned dataset was used (with only six outgroups from Sapindales and all ingroup) and dates for *Aesculus* split from *Billia* (64 Mya) and *Acer* from *Dipteronia* (64 Mya) as the only fossil constraints (table 3.1 – analyses A). Both nodes were constrained to a log normal distribution (zero offset 63 Mya standard deviation SD 1.0) providing a hard bound for the fossil date (no potential for younger dates) and a narrow maximum bound for the potential oldest date. Two analyses were run for 10 million generations sampling the topology every 1,000 generations and parameter values every 250 generations. Phylograms with branch lengths in units of substitutions per site were also saved every 10,000 generations.

To determine whether there is any lineage specific variation in evolutionary rates amongst the sampled Sapindaceae, a Bayesian relative rates test was performed which is considered to overcome many of the shortcomings of other forms of relative rate tests (Wilcox et al., 2004). For this test the posterior probability distribution of lengths of all branches from the 1800 phylograms (excluding burnin) returned from the previous analyses are used. Summed branch lengths for all trees were compiled in Cadence v1.0 (Wilcox et al., 2004) from each ingroup terminal to the MRCA of all the descendents of the ingroup i.e. crown Sapindaceae. Rates of evolution were considered to be significantly different between two taxa if their 95% credibility intervals of the posterior probability of their summed branch lengths did not overlap.

Table 3.1 (next page): Prior probability distributions and posterior probability densities from analyses of various combinations of priors mentioned in the text from the Bayesian Markov chain Monte Carlo (MCMC) lognormal relaxed clock phylogeny estimations. For priors with lognormal and normal distributions the mean constraint age in millions of years ago (Mya) is shown with the 95% prior interval (PI), while for the uniform prior the minimum and maximum bounds of the uniform distributions in Mya are shown for the calibration nodes. For the

posterior distributions from the MCMC lognormal relaxed clock phylogeny estimations without and with the nucleotide data shown are the mean and 95% highest posterior density distributions (HPD) for the nodes with prior distributions, and for other nodes estimated during the MCMC approximations (tree prior) discussed in the text. Tree prior refers to node estimations from the MCMC approximations with and without the nucleotide data.

Node	Prior	Mean (95% PI)	MCMC no data Mean (95% HPD)	MCMC posterior Mean (95% HPD)
Analyses A				
Acer/*Dipteronia*	Lognormal	64 (63-68)	65 (63-69)	65 (63-68)
Aesculus/*Billia*	Lognormal	64 (63-68)	64 (63-67)	64 (63-67)
Sapindaceae crown	Tree prior		203 (87-446)	246 (145-360)
Analyses B				
eudicot	Normal	121 (116-126)	121 (118-126)	122 (120-124)
Cornaceae	Lognormal	87 (86-91)	87 (86-90)	87 (86-89)
Myrtaceae	Lognormal	85 (84-89)	85 (84-87)	85 (84-86)
Proteaceae	Lognormal	85 (84-89)	87 (84-89)	85 (84-88)
Sapindaceae crown	Tree prior		76 (60-90)	57 (47-67)
MRCA Acereae/Hippocastaneae	Tree prior		55 (31-73)	38 (24-51)
Acer/*Dipteronia*	Tree prior		24 (21-58)	15 (6-25)
Aesculus/*Billia*	Tree prior		11 (1-20)	12 (4-22)
Analyses C		Min-max		
eudicot	Uniform	118-140	132 (121-140)	138 (134-140)
Cornaceae	Uniform	88-92	90 (88-92)	90 (88-92)
Myrtaceae	Uniform	86-90	88 (86-90)	88 (86-90)
Proteaceae	Uniform	86-90	88 (86-90)	88 (86-90)
Acer/*Dipteronia*	Uniform	64-68	66 (64-67)	66 (64-68)
Aesculus/*Billia*	Uniform	64-70	67 (64-70)	66 (64-69)
Analyses D				
eudicot	Normal	121 (96-145)	134 (110-159)	145 (122-170)
Cornaceae	Normal	87 (62-112)	82 (55-109)	70 (49-90)
Myrtaceae	Normal	84 (59-109)	80 (54-106)	65 (44-84)
Proteaceae	Normal	84 (59-109)	74 (46-101)	84 (61-109)
Acer/*Dipteronia*	Normal	64 (39-88)	57 (33-82)	24 (11-38)
Aesculus/*Billia*	Normal	64 (39-88)	46 (20-72)	21 (9-36)

The next series of analyses used a relaxed clock with the rates for each branch drawn from an underlying uncorrelated lognormal distribution with no nodes constrained to a prior distribution on the dataset with the 21 outgroups mentioned above and also on the two single gene datasets. The mean substitution rate was fixed to one (mean root-to-tip path length equal to one) allowing for a relaxed clock estimation of the relative substitution rates (estimated in substitutions per site) across the phylogeny (i.e. a ratogram). Two independent MCMC chains for each dataset were run for 40 million generations each for the combined dataset and 20 million generations for the individual gene datasets sampling the topology every 10,000 generations and parameter values every 500 generations.

A monophyletic Sapindaceae/Sapindales (Gadek et al., 1996b; Harrington et al., 2005) is part of a large monophyletic group, the eudicotyledons, which has the synapomorphy of tricolpate pollen (Doyle and Hotton, 1991). Within the Sapindales the direct sister families to Sapindaceae remain unresolved (Gadek et al., 1996b; APG, 2003). In the next analyses of the large dataset (table 3.1 – analyses B) four nodal prior distributions based on the fossil record from outside of Sapindales that have been used in other studies and are currently unchallenged are used: i) eudicots stem with a normally distributed prior (mean = 121 Mya, standard deviation, SD = 3.0) which is a generally accepted calibration that has been used in many studies, for some examples see Sanderson (2003), Davies et al. (2004a), Anderson et al. (2005), and Davis et al., (2005); ii) Cornaceae stem with lognormal prior distribution (zero offset 86 Mya, SD = 1.0); iii) Proteaceae stem lognormal prior distribution (zero offset 84 Mya, SD = 1.0), see Anderson et al. (2005) for these fossils descriptions; iv) Myrtaceae stem lognormal prior distribution (zero offset 84 Mya, SD = 1.0) see Sytsma et al. (2004) for fossil description. The application of these prior distributions is a conservative approach to fixing node dates on the Bayesian trees. The use of log normal priors provides a ‘hard’ upper bound (provided by the fossil date) on the relevant stem node and a ‘soft’ lower bound. That is, the descendant lineage must be at least as old as its oldest known fossil representative, while there is a non-zero probability that the stem node is older than the fossil, corresponding to a log-normal distribution. Further justification for this approach is outlined in Yang and Ranala (2006) and Benton and Donoghue (2007).

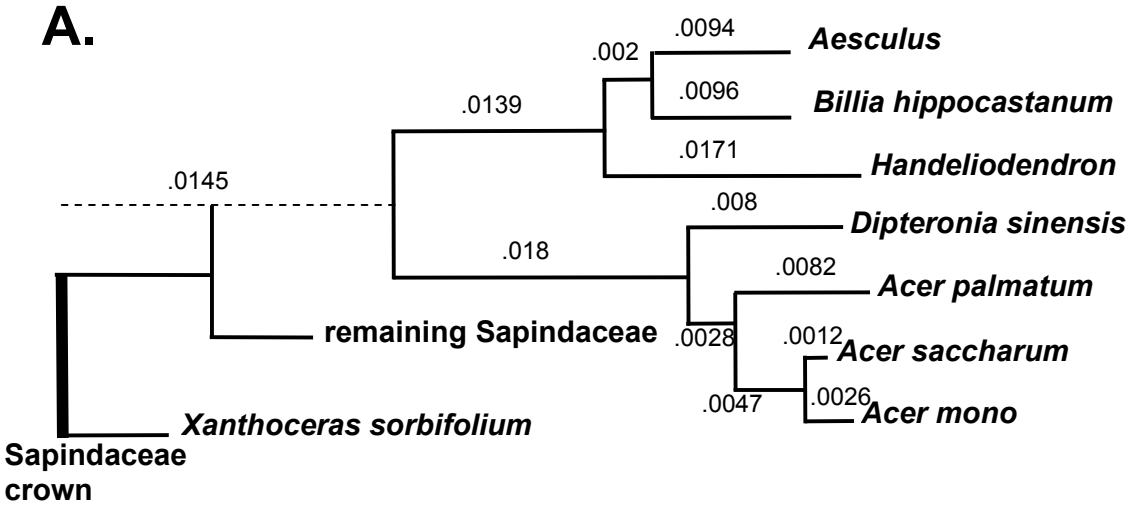
The final two relaxed clock analyses on the large dataset used the four above mentioned fossils and also included the dates for *Aesculus* split from *Billia* (64 Mya) and *Acer* from *Dipteronia* (64 Mya). The approach of Sanders and Lee (2007) was adopted of first assigning hard bounds on narrow uniform distributions to 'fix' these six nodes, thereby generating branch specific substitution rates based on all the calibration nodes being assumed to be correct (table 3.1 – analyses C). The uniform prior distributions with lower and upper bounds in Mya were: eudicots (118-140), Cornaceae (88-92), Myrtaceae (86-90), Proteaceae (86-90), *Acer*/*Dipteronia* (64-68), and *Aesculus*/*Billia* (64-70). For the last analyses relatively wide normal distributions (soft bounds) were assigned to each calibration node which enables them to vary symmetrically (table 3.1 – analyses D). The mean normal prior distributions with their 95% credibility intervals (CI) in Mya were: eudicots 121 (96-145), Cornaceae 87 (62-112), Myrtaceae 84 (59-109), Proteaceae 84 (59-109), *Acer*/*Dipteronia* 64 (39-88), and *Aesculus*/*Billia* 64 (39-88).

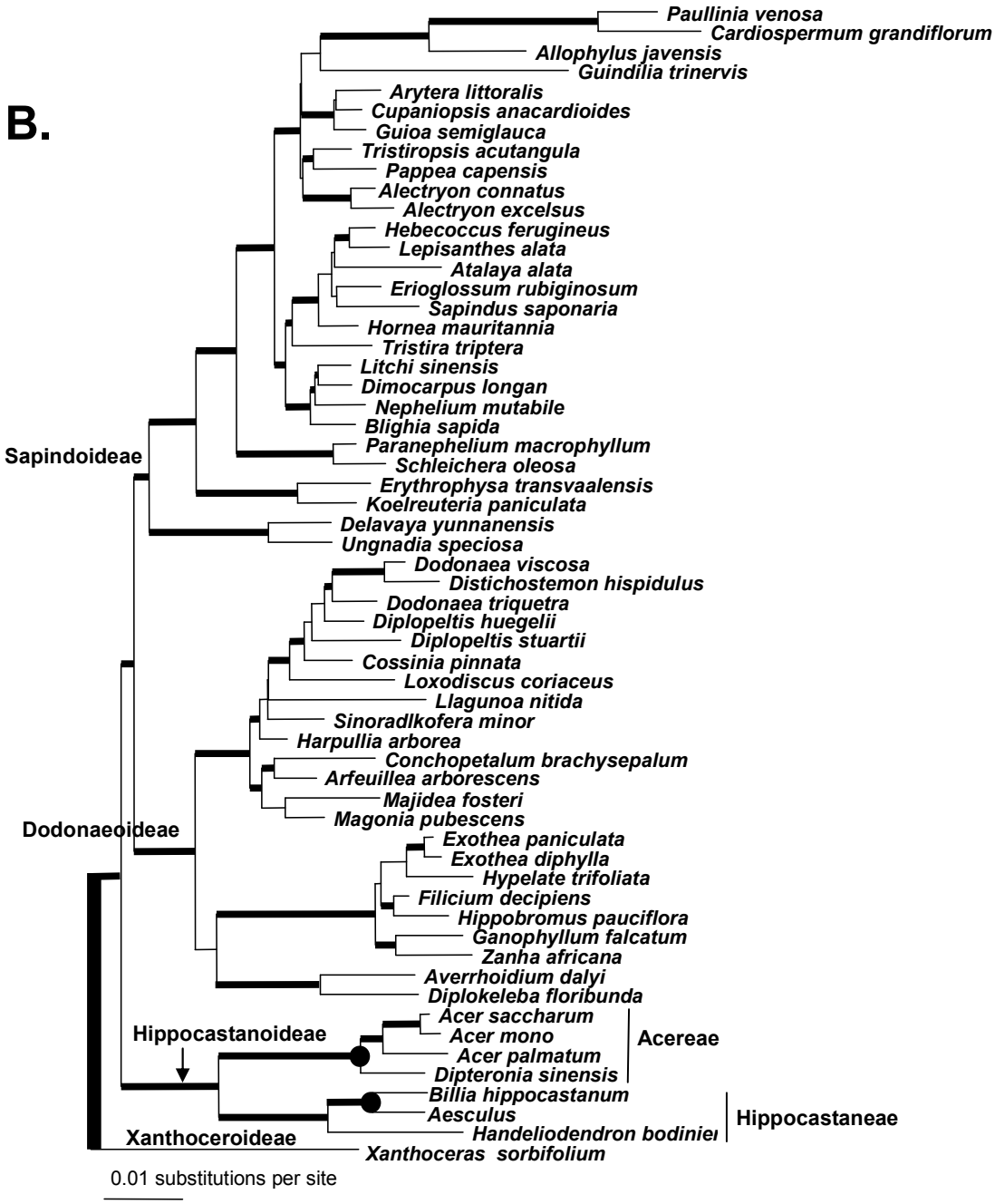
Species diversification rates, assuming an equal rate of random speciation Yule model, were calculated as SDR $=(ln\ n_1 - ln\ n_0)/t$, where n_1 is the number of extant species, n_0 is the initial species diversity, here taken as 1, and t is time in Mya. The 95% upper and lower higher probability distributions of age estimates for the stem of *Acer* from analyses B (table 3.1) were used in calculations of speciation rate.

Results and Discussion

All analyses of the combined and separate gene datasets produced estimations of topology with a high proportion of nodes with PP ≥ 95% and congruent with the phylogenies presented in Harrington et al. (2005), with the four subfamilies (Sapindoideae, Hippocastanoideae, Dodonaeoideae and Xanthoceroideae) each receiving good support. An edited phylogram from the combined analyses of the large dataset (analyses B) is shown in fig. 3.1. The effective sample size (number of independent samples from the marginal posterior or prior distributions) for all designated nodes of interest in all analyses were >200 indicating that the MCMC chains were run for adequate lengths.

Figure 3.1A: Edited consensus phylogram showing Hippocastanoideae (includes Acereae – *Acer* and *Dipteronia*, and Hippocastaneae – *Aesculus*, *Billia*, and *Handeliodendron*) from Bayesian analysis of combined plastid dataset for Sapindaceae with mean branch lengths measured in expected substitutions per site. Dotted line is branch length from Hippocastanoideae crown to Sapindaceae crown. All nodes in this edited phylogram receive 100% posterior probability. **3.1B (next page):** Edited consensus phylogram showing crown Sapindaceae (vertical bar) from Bayesian analysis of combined plastid dataset with mean branch lengths measured in expected substitutions per site. Solid lines lead to clades receiving posterior probabilities ≥95%.The four subfamilies of Sapindaceae are marked along with the two tribes of Hippocastanoideae mentioned in the text. Black circles indicate nodes *Acer*/*Dipteronia* split and *Aesculus*/*Billia* split mentioned in text.





A preliminary examination of the rates of evolution (branch lengths in substitutions per site) in the edited phylogram (fig. 3.1) from the Bayesian analyses (analyses B – table 3.1) of the combined plastid dataset shows that the distance from the Sapindaceae crown group to MRCA of *Acer*/*Dipteronia* has 0.0325 expected substitutions per site. If this rate is extrapolated, say under the simple assumption of a

strict molecular clock (equal rate per unit of time) using the average substitution rate between *Acer* and *Dipteronia* for their terminal branches and assuming a 64 Mya divergence for these lineages, this presents a minimum age for Sapindaceae crown group of around 220 Mya, which is well outside the presumed first appearance (121 Mya) of the distinctive tricolpate pollen, marking the radiation of the eudicots (including Sapindaceae) (Hughes and McDougall, 1990; Doyle, 1992). Reconciling the fossils and molecular clock rate estimates would require either a severe and potentially rapid slow down in the substitution rate in the plastid sequences of all members of the Hippocastanoideae between 64 Mya to Recent or conversely a dramatic quickening in the substitution rate of the remainder of Sapindaceae between 121-64 Mya. A further explanatory scenario is that both of these postulations happened consecutively.

A strict molecular clock estimate is unrealistic since all combined and separate gene relaxed clock estimations indicated a moderate degree of rate heterogeneity amongst branches (95% CI coefficient of variation significantly removed from zero - table 3.2), and there were also significant levels of substitution rate autocorrelation between related branches in some parts of the phylogeny (95% CI covariance significantly removed from zero - table 3.2).

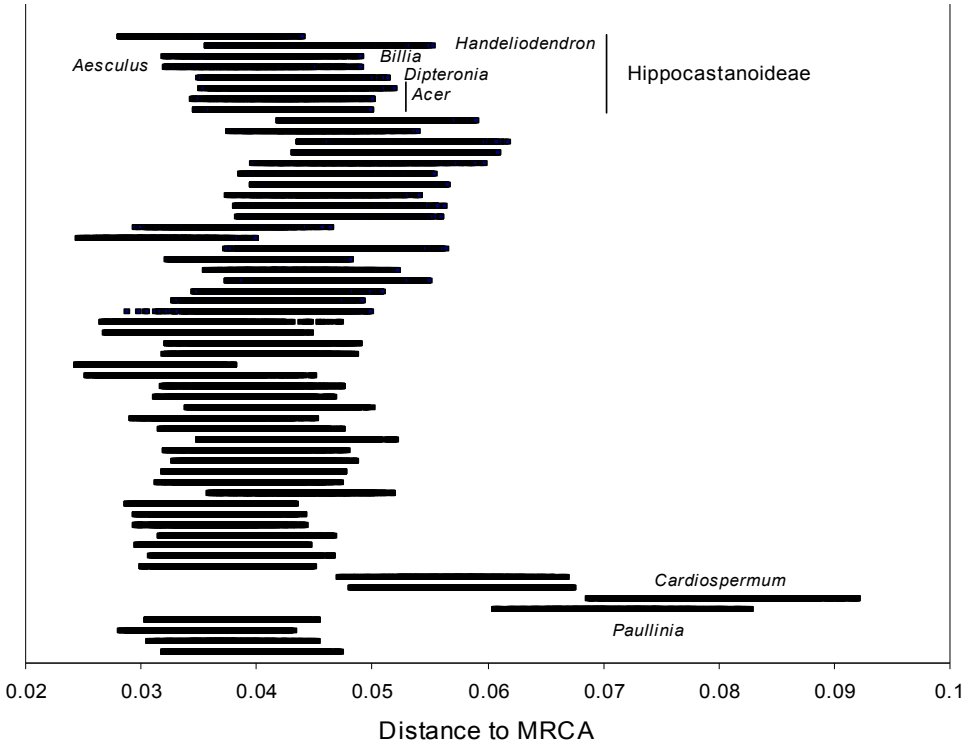
Table 3.2: Coefficient of variation (95% highest posterior density HPD) and covariance (95% HPD) statistics for analyses B of the combined dataset (table 3.1) and for the single gene analyses.

	Combined	*rbcL*	*matK*
Coefficient of variation (95% HPD)	0.72 (0.62-0.82)	0.58 (0.44-0.72)	0.68 (0.54-0.83)
Covariance (95% HPD)	0.18 (0.04-0.32)	0.05 (-0.07-0.19)	0.16 (0.03-0.29)

The Bayesian relative rate test of the combined data indicates that the majority of lineages within Sapindaceae including all genera of Hippocastanoideae have rates of molecular evolution that generally fall within an overlapping range (branch lengths 95% CI distributions from MRCA of ingroup Sapindaceae - fig. 3.2), however, there are two genera with significant rate increases from the remainder of Sapindaceae. These two genera along with two other genera in the same clade that have a generally faster

molecular rate are most likely to be the source of the autocorrelation identified in the relaxed clock analyses. The two genera with rates of evolution significantly faster than all other sampled Sapindaceae are either herbaceous vines (*Cardiospermum*) or woody lianas (*Paullinia*), both with auxiliary tendrils representing modified inflorescences. A Bayesian relative rate analysis of the *rbcL* dataset (result not shown) indicates that the other speciose genus of woody lianas, *Serjania*, also has a significantly faster evolutionary rate. The evolution of the climbing habit key innovation (Gianoli, 2004) in these genera has induced the greatest diversification of any Sapindaceous lineage (lianas - 6 genera, 460 species, remaining Sapindaceae - 134 genera, ca. 1090 species), and this has been shown to potentially correlate with the increased rate of plastid evolution in these genera (Barraclough and Savolainen, 2001).

Figure 3.2: Bayesian relative rates test 95% credibility intervals of the posterior probability distributions of summed branch lengths from the most recent common ancestor (MRCA) of crown group Sapindaceae. Rates of evolution are considered to be significantly different between two taxa if their 95% credibility intervals do not overlap. Only taxa discussed in the text are named.



All analyses performed without data generally returned effective prior distributions not in conflict with the original priors indicating that the MCMC approximations were working, while posterior distributions obtained for all analyses except one (see below) generally departed from the prior distribution indicating informative data (table 3.1). The relaxed clock analyses using the *Acer*/*Dipteronia* split (64 Mya) and *Aesculus*/*Billia* split (64 Mya) with lognormal priors resulted in a mean crown age for Sapindaceae of 246 Mya with the 95% highest posterior density (HPD) ranging from 145-360 Mya (analyses A - table 3.1 - tree and other summary statistics not shown). Clearly, this estimated age of Sapindaceae is strongly at odds with our current understanding of the evolution of this derived family within the eudicots and suggests several potential scenarios similar to the previous analyses. Possible explanations are that the fossil constraints have been incorrectly assigned to the *Acer/Dipteronia* and *Aesculus*/*Billia* splits or that there is a substantial change in substitution rates in either Hippocastanoideae or remaining Sapindaceae. When compared to the analysis run without any nucleotide data the 95% credibility (i.e. posterior) interval was totally encompassed by the 95% prior interval for all nodes (for example see Sapindaceae crown table 3.1) indicating that the joint prior assumptions are strongly influencing the outcome of the posterior distribution (joint priors plus sequence data).

The substitution rates from the ratogram (table 3.3 - tree and other summary statistics not shown) of the analyses of the data without any age constraints were contrasted with node heights from the fossil constrained analyses by converting them to age estimates for the nodes of interest. Using a conservative 122 Mya for the crown group of the eudicots generates a mean date of 14 Mya for the MRCA of *Acer* and *Dipteronia*, and conversely, a 64 Mya for the *Acer*/*Dipteronia* split results in a 704 Mya for crown eudicots (table 3.3). Age estimates for nodes of interest in molecular dating studies have been shown to be influenced by sampling density within a lineage (Linder et al., 2005). However, in this instance because I am only interested in the divergence time estimates for the crown node between two sets of terminal genera, as long terminals with potentially the longest branch to the crown node are included (for example recently derived taxa) the extent of molecular evolution within the genus

should be covered. The species of *Acer* and *Aesculus* included in this study appear to fulfil this requirement (Xiang et al. 1998; Renner et al. 2007). Confirmation for this sampling strategy also comes from the similar age estimates for *Acer/Dipteronia* and *Aesculus*/*Billia* splits obtained from extrapolations of the substitution rate to time for the single gene ratogram analyses with multiple representatives of *Acer* and *Aesculus* (table 3.3).

Table 3.3: Mean substitution rate per site from the Bayesian Markov chain Monte Carlo (MCMC) lognormal relaxed clock phylogeny estimations of the combined and single gene datasets performed without any time constraints. Time extrapolations in millions of years ago (Mya) of the mean substitution rate per site using the bold figure for selected nodes are also shown. Extrapolations from *Acer*/ *Dipteronia* and *Aesculus*/ *Billia* are based on the average of the two mean substitution rates (combined 0.01, *rbcL* 0.008, *matK* 0.015 substitutions per site).

		Eudicot crown	Sapindaceae crown	*Acer*/ *Dipteronia*	*Aesculus*/ *Billia*
Combined	mean subs/site	0.11	0.046	0.013	0.01
Extrapolation	Mya	**122**	51	14	11
Extrapolation	Mya	704	294	**64**	**64**
rbcL	mean subs/site		0.03	0.0096	0.0064
Extrapolation	Mya		240	**64**	**64**
Extrapolation	Mya		**51**	16	11
matK	mean subs/site		0.057	0.013	0.016
Extrapolation	Mya		243	**64**	**64**
Extrapolation	Mya		**51**	12	14

Edited phylogeny and divergence time estimates from the Bayesian analyses (analyses B – table 3.1) using four well-corroborated fossils from outside of Sapindaceae to provide an independent estimate of the age of the lineages of Hippocastanoideae are shown in fig. 3.3. The constrained stem of eudicots dates to 122 Mya (HPD 120-124), the crown Sapindaceae dates to 57 Mya (HPD 47-67) and for the MRCA of Acereae and Hippocastaneae the age is estimated at 38 Mya (HPD 24-51).

Using calibration nodes with age estimates far removed from nodes of interest have been shown to produce underestimations of divergence times (Linder et al., 2005); however, these analyses produced ages for two stem nodes (one sister to Sapindaceae and one internal in Sapindaceae) that are congruent with the fossil record for the associated lineage – *Ailanthus* (*Simaroubaceae*) mean 53 Mya (HPD 41-69), fossil *Ailanthus* ca. 52 Mya (Corbett and Manchester, 2004), and *Alectryon* (Sapindaceae) mean 24 Mya (HPD 18-31), fossil 24-26 (Guerin and Hill, 2006).

The estimated age for the split between *Acer* and *Dipteronia* is in the mid-Miocene about 15 Mya (HPD 6-25), with a slightly younger age for the MRCA of *Aesculus* and *Billia* 12 Mya (HPD 4-22) (fig. 3.3, table 3.1). Potential support for relatively recent age generated by the molecular data for the origin of *Acer* comes from a study on the evolution of the powdery mildew genus *Sawadaea* (Ascomycota: Erysiphaceae), an obligate parasitic fungus of maples whose phylogeny and geographical distribution suggest a close co-evolutionary relationship (Hirose et al., 2005). A molecular clock estimation of divergence time for *Sawadaea* indicate an origin in the Oligocene about 30 Mya with the divergence of the seven major clades within 10 Mya indicating a plausible host (*Acer*) parasite (*Sawadaea*) co-evolutionary relationship. However the current interpretation of the fossil record indicates that rather than being an example of cospeciation, *Sawadaea* instead host shifted onto *Acer* after the divergence of all the main groups and sections of the genus, and then expanded their host ranges congruent with the phylogeny and geographical distribution of *Acer* (Hirose et al., 2005).

The relaxed clock estimates with soft bounds on all normally distributed calibration priors resulted in 95% HPD estimates for the splits between *Acer* and *Dipteronia* and *Aesculus* and *Billia* with a similar range to that in analyses B (table 3.1) and totally outside the range stipulated by the prior distribution for these calibration nodes (analyses D - table 3.1). These results indicate that for relaxed clock estimates of divergence times only narrow hard bound calibration priors for the splits between *Acer* and *Dipteronia* and *Aesculus* and *Billia* can accommodate the minimum dates implied by the fossil record. The use of soft bounds have been shown in simulation studies to be

generally superior to hard bounds in that they better incorporate the data in the analyses; alternatively no amount of data can change hard bounds (Yang and Rannala, 2006).

Figure 3.3: Edited chronogram selected from the highest probability density (HPD) from Bayesian relaxed clock analyses using four fossil constraints (analyses B – table 3.2) showing outgroups and only other taxa mentioned in text. The area covered by the black box indicates the extent of the Miocene. Horizontal bars are the 95% HPD of node heights. The crown of Sapindaceae is marked and the placement of the four fossil constraints represented by bold vertical bars.

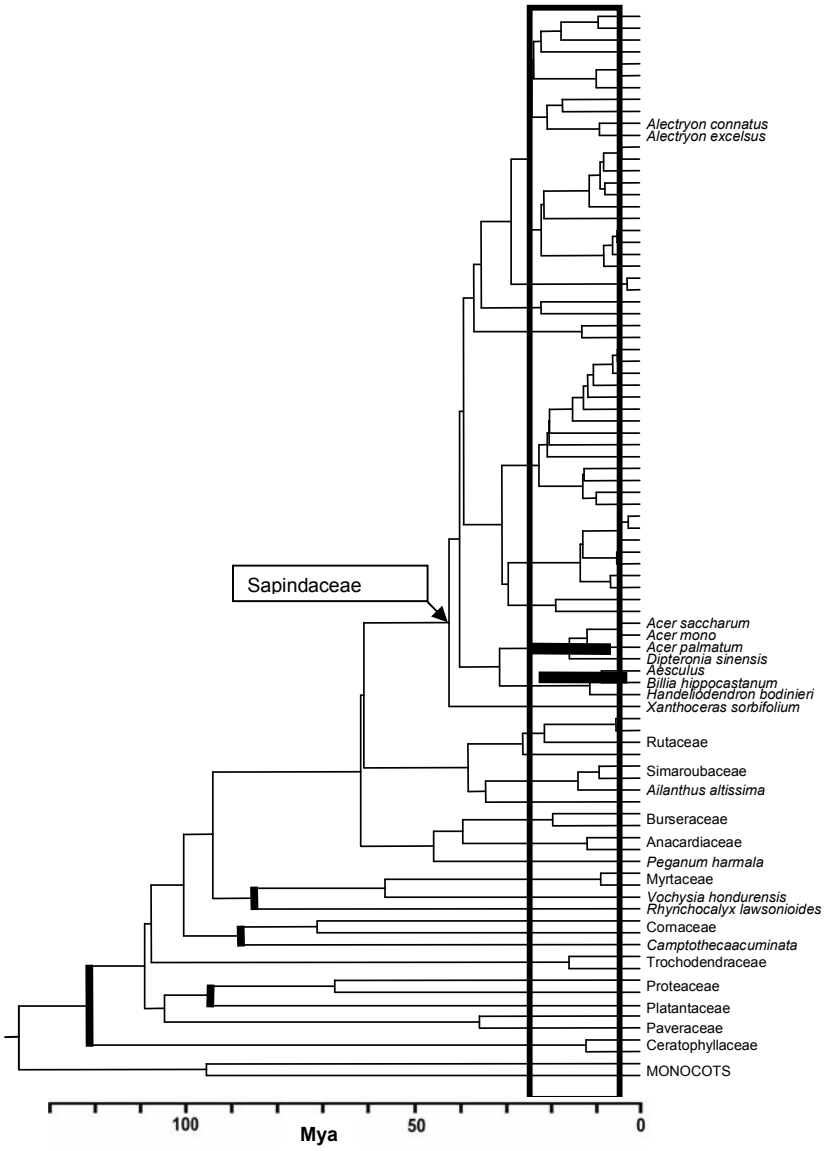
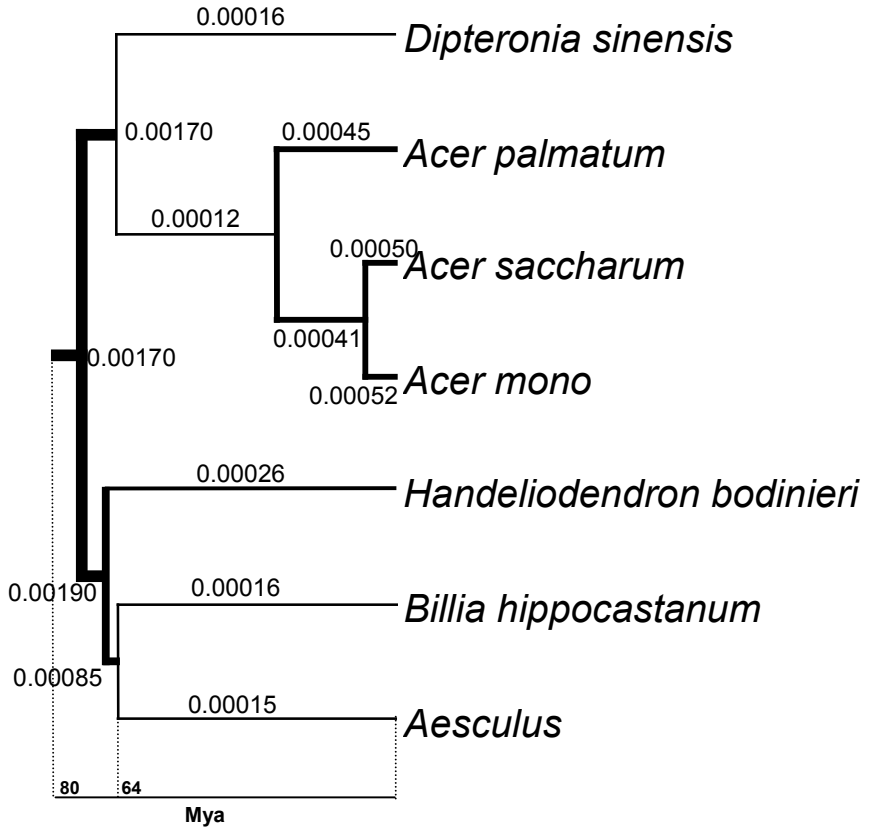


Figure 3.4: Edited chronogram showing substitution rate (per site per million years) for stem and crown Hippocastanoideae selected from the 95% highest probability density from Bayesian relaxed clock analyses C (table 3.2). The thicker the line the faster the evolutionary rate.



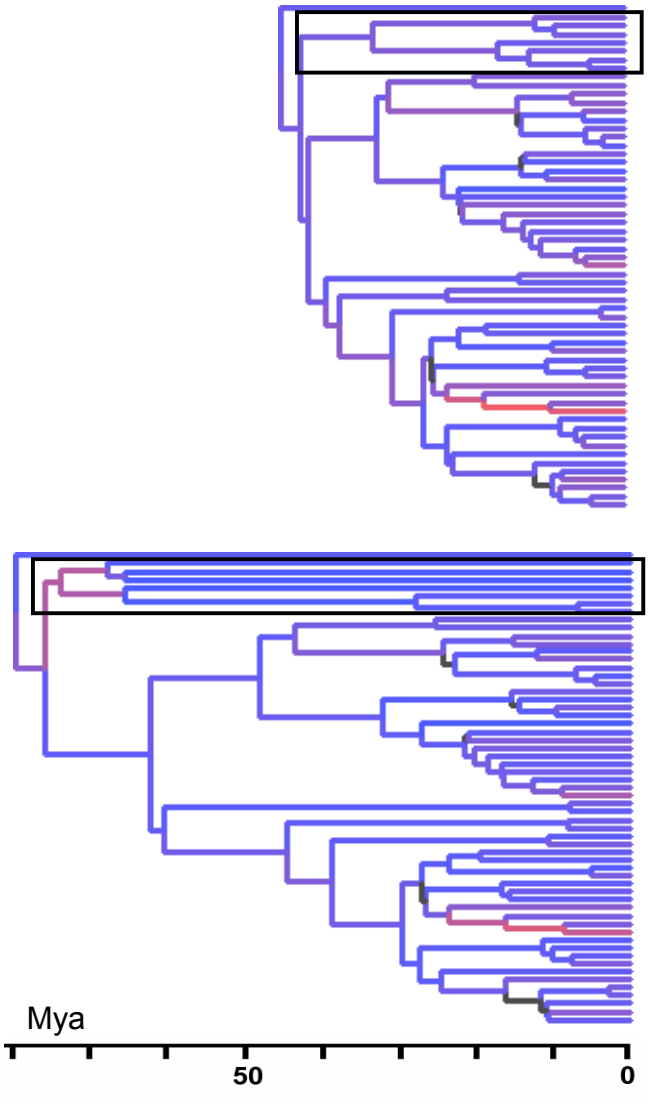
While the Bayesian relative rate test indicated no significant difference in the overall substitution rates in Hippocastanoideae from most of the remaining Sapindaceae sampled (fig. 3.2), it does not preclude concerted rate shifts (accelerations and/or decelerations) along some of the branches that contribute to the overall rate. The partial chronogram from the reconstruction of the ancestral rates of molecular evolution for the six fossil calibrations under hard bound uniform distributions is shown in fig. 3.4 (Sapindaceae chronogram fig 3.5). Incorporating the Late Paleocene split between *Acer* and *Dipteronia* as a constraint in relaxed clock estimates of divergence times requires at least a ten fold change in the mean substitution rate between the branches leading to the split and the branches after the split. The rate of molecular evolution on the stem of Hippocastaneae and Hippocastanoideae (fig. 3.4) is similarly faster than the rate on branches to the terminals. It is difficult to interpret how this apparent long period of

mutation rate stasis can be biologically plausible. Generally it is considered that the underlying causes of accelerations and decelerations in evolutionary rates have depended on a combination of biological and ecological factors (eg. generation time, metabolic rate, efficacy of DNA repair). While the current interpretation of the extensive fossil record does suggest a possible increase in extinctions of *Acer* species at the end of the Oligocene (Wolfe and Tanai, 1987; Boulter et al., 1996) there are no other indications that there has been any dramatic change in life-history traits within *Acer* and *Aesculus* nor has it been suspected for the progenitors of *Acer* and *Aesculus*. Alternatively there is no evidence that there has been a change in the evolutionary rates in all the other Sapindaceae.

A more realistic interpretation of the evolutionary rates within Hippocastanoideae is provided by analyses B (table 3.1, fig. 3.5), which corroborates the results of the relative rates test (fig. 3.2) and shows that there is virtually no difference in the stem and crown rates of Hippocastanoideae, and similarly little difference from the range of values in the remaining Sapindaceae except for the higher rates for the lianas. It is interesting to note that there is a relatively high mean evolutionary rate on the stem of Sapindaceae (0.0016 substitutions per site per million years), however, this is probably an artefact of undersampling of Sapindales outgroups, non-resolution of relationships between families of Sapindales, and extinction.

While the fossil evidence has previously suggested great antiquity for these generally deciduous temperate genera, this study shows that there is molecular evidence that the large genus *Acer* diverged from its sister genus *Dipteronia* in the Miocene, as did *Aesculus* from *Billia*. This was around the time of a Miocene cooling period that continued until the Pleistocene ice ages. There are geographical disjunctions between the genera of Hippocastanoideae and between multiple species pairs within *Acer* and *Aesculus*. If a consequence of vicariance, the disjunctions between North America, Europe and Asia can only have arisen via the Bering land connection which was the last Tertiary land connection between Eurasia and North America. Floristic exchange could have continued across this corridor until ~5.5 Mya. Episodic dispersal events also cannot be discounted as there are two species of *Acer* on the recent volcanic island of Ullung (1.8 Mya), which is 137 km from peninsula Korea (Pfosser et al., 2002).

Figure 3.5: Top - Edited chronogram showing substitution rate (per site per million years) for crown Sapindaceae selected from the 95% highest probability density (HPD) from Bayesian relaxed clock analyses B (table 3.2). Below - Edited chronogram showing substitution rate (per site per million years) for crown Sapindaceae selected from the 95% highest probability density (HPD) from Bayesian relaxed clock analyses D (table 3.2). In both cladograms a black box surrounds Hippocastanoideae and the branching order for the unnamed taxa is generally the same. Branch colour indicates substitution rate (per site per million years) with the transition being from the blue slower rate to the orange faster rate (approximately 10x slowest rate). Black branches are posterior probabilities <50%.



The relatively recent date for the emergence of *Acer* from MRCA resolved from the molecular data potentially has a profound impact on the interpretation of the evolutionary history of the genus. For example, if we compare a simple Yule birth rate estimation of diversification (no extinction) using the dates obtained under the Bayesian relaxed clock analyses (analyses B, table 3.1), the estimated rate of species accumulation per million years for *Acer* ranges from 0.18 (28 Mya) – 0.84 (7 Mya), which is still higher than the angiosperm family median (0.12) at the lower rate (Eriksson and Bremer, 1992), while the higher rate is comparable to that of the insular radiation of the Hawaiian silverswords (Baldwin and Sanderson, 1998). There have been multiple phylogenetic studies using many different molecular markers aimed at interpreting the intersectional relationships within *Acer*, with both chloroplast and nuclear, both singly and combined, having by and large been unsuccessful in determining deep level relationships (Ackerly and Donoghue, 1998; Suh et al., 2000; Pfosser et al., 2002; Tian et al., 2002; Grimm et al., 2006; Li et al., 2006; Renner et al., 2007). While a number of monophyletic sections within the genera have been molecularly supported, generally relationships between the many clades remain unresolved, with very short or zero branch lengths and a broom-handle phylogeny suggestive of a history of extinction and of recent, rapid radiation (Richardson et al., 2001; Crisp et al., 2004). A similar result was obtained in the only molecular phylogenetic analyses of *Aesculus* (Xiang et al., 1998).

There have been many examples where age estimates generated from molecular dating studies have deviated from ages suggested by the paleontological record. The molecular estimates for divergence times within Hippocastanoideae presented in this paper call into question the current assignment of the pre-Miocene fossils attributable to *Acer*, *Aesculus* and *Dipteronia*. It may be possible that the extinct progenitors of the Hippocastanoideae lineage may have had a similar convergence in morphological trends that has been repeated in the extant group, and as such the fossils may represent extinctions on the stem lineages leading to the various splits in the tribes and genera. The oldest fossils could even belong to extinct convergent lineages on the stem of the Sapindaceae.

It is difficult to reconcile the two competing evolutionary scenarios for the Hippocastanoideae that have been presented in this study. If we accept the molecular evidence as providing the more biologically realistic interpretation of evolutionary rates within Sapindaceae, then we have to totally reassess vast quantities of well-preserved micro and macrofossils that have been attributed to *Acer*, *Aesculus*, and *Dipteronia* from the Paleocene to the Miocene. Alternatively maintenance of the current fossil record attributed to *Acer*, *Aesculus*, and *Dipteronia* requires further investigation into the biological probability and reasons for the abrupt and dramatic change in evolutionary rates in one subfamily of Sapindaceae.

Chapter 4: Comparative study of the evolution of nuclear ribosomal spacers incorporating secondary structure analyses within Dodonaeoideae, Hippocastanoideae and Xanthoceroideae (Sapindaceae)

Phylogenetic relationships within Sapindaceae have been investigated using a combined plastid dataset of *rbcL* and *matK* genes (Harrington et al., 2005, and Chapter 2). The plastid gene tree is well resolved in all of the deepest branches; however, there is a lack of resolution within the Dodonaeoideae. Since there is near complete sampling for this subfamily and because of its significance to understanding sister relationships to *Dodonaea* (the subject of phylogenetic analyses in chapter 5), further data are required to resolve relationships unequivocally. Biological knowledge of the evolution of Dodonaeoideae has been difficult to ascertain using morphological and pollen characters, and it has generally been regarded as a heterogeneous assemblage of genera (Radlkofer, 1890; Muller and Leenhouts, 1976; Buijsen et al., 2003). In an attempt to improve the robustness of phylogenetic hypotheses within Dodonaeoideae (Sapindaceae) and to assess congruence with the plastid analyses, DNA sequences from spacer regions that separate ribosomal genes (rDNA) were generated.

The use of molecular data from the internal transcribed spacers (ITS1 and ITS2 = ITS) of the nuclear ribosomal repeat for phylogenetic inference at infrageneric levels, or between closely related genera is widely popular in plant systematics (see Hershkovitz et al., 1999; Alvarez and Wendel, 2003), while the use of a partial fragment 5'external transcribed spacer region (ETSf) at the same taxonomic level is usually dependent on the ease and ability of primer design (Hershkovitz et al., 1999). However usage of these spacer regions in phylogeny reconstruction is not without concern, particularly with regards to several evolutionary peculiarities that are apparent when sequencing and analysing these regions. These include non-complete concerted evolution of the ribosomal repeats resulting in intrasequence polymorphisms, preferential sequencing of paralogous non-functional copies (pseudogenes), non-independence of many nucleotide sites due to the properties of secondary structure

constraints and the potential for compensatory base changes (CBC), and also noticeable homoplasy, alignment and rooting difficulties due to the faster evolutionary rate of change of nucleotides within spacer sequences (Hershkovitz et al., 1999; Alvarez and Wendel, 2003; Bailey et al., 2003). In this study these problems were addressed by adopting specific protocols developed by other workers to identify possible pseudogenes, and to override many of the other concerns alignment and analyses methods were used that incorporate secondary structure (further discussion on all these issues below). Also a general framework is developed for analysing small RNA molecules that incorporates testing for functional constraints.

Both ITS and ETSf are part of the cistron that encodes the 18S, 5.8S and 26S single strand rRNAs that occur in tandem arrays located at one or more chromosomal loci. Due to lower selective constraints compared to the coding regions, the primary sequence of these spacers can be highly variable and possibly unalignable between more distantly related species due to length and intrasequence nucleotide heterogeneity between sampled taxa.

rDNAs encode RNA genes, which are single stranded but develop secondary structure where the molecule folds onto itself to form generally short regions of Watson Crick base-pairings (G:C and A:U) and the intermediate non-cannonical pair (G:U) in stems, and single stranded loops (see table 4.1 for glossary). Throughout the rest of this chapter the RNA equivalent U (uracil) for the T (thymine) in the DNA is used. Stems are generally conserved over evolutionary time and the pairings are maintained by compensatory mutation (CBC – compensatory base change). Knowledge of the secondary structure of a sequence can provide information in terms of optimal base pairing that can aid alignment (Coleman and Mai, 1997; Denduangboripant and Cronk, 2001). Manual alignment hypotheses that are aided and constrained by secondary structure conventions have been shown to favor phylogenies more congruent with other sources of data than other alignments (see Kjer, 2004 for multiple references). Pseudogenes will have unstable secondary structures so they may be identified by comparison of their secondary structures with those of stable functional copies (Buckler et al., 1997b; Mayol and Rossello, 2001).

Conventional phylogenetic analysis methods use models that generally assume that nucleotide sites in a sequence are evolving independently and are spatially distinct. For RNA encoding regions, however, secondary structure dictates that this is not a valid assumption, so that paired-site models developed to deal specifically with stem structures in RNA encoding sequences are more appropriate (Jow et al., 2002; Hudelot et al., 2003). It has also been shown that phylogenetic reconstructions that employ independent assumptions for non-independent data can over-estimate support (in terms of bootstrap) for internal branches (Jow et al., 2002; Galtier, 2004; Smith et al., 2004).

Table 4.1: Glossary of RNA secondary structure terms (adapted from Gillespie, 2004)

Term	Definition
Helix (stem)	A double helix composed of a succession of complementary hydrogen-bonded nucleotides between paired strands. Pairing generally involves the Watson-Crick A:U, G:C pairs and the noncanonical G:U pair
Single strand loop	Unpaired nucleotides separating helices
Terminal loop	Succession of unpaired nucleotides at the end of a stem
Lateral bulge	Succession of unpaired nucleotides on one strand of a helix
Internal bulge	Group of nucleotides from two parallel strands unable to form canonical pairs
Mismatch pair	Any pairing in a secondary structure model that does not involve A:U, G:C or G:U pairs
Compensatory base change (CBC)	Subsequent mutation on one strand of a helix to maintain canonical pairing following initial mutation of a complementary base

Virtually all of the studies utilising ITS and/or ETSf primary sequences have ignored or barely considered the secondary structure of these molecules. It has been shown that for a range of rRNA sequences the use of models of sequence evolution that allow stem and loop regions to evolve according to separate models significantly improve likelihood-based estimates of phylogeny compared to independent models (Muse, 1995; Schoniger and von Haeseler, 1999; Savill et al., 2001; Telford et al., 2005), and recently this approach has also been shown to be appropriate for use with ITS rDNA sequences (Biffin et al., 2007). There have been no studies that I am aware of

that have incorporated secondary structure and paired-sites models for analyses of ETSf sequences.

Chemical determination of secondary structure is rarely attempted; rather molecular biologists use a comparative biology approach across a sequence alignment to define a putative secondary structure, with the support for the core structure coming from covariation analyses. A helix is considered 'proven' when it contains at least one CBC, or contains conserved structural motifs or other specific structural elements (e.g. tetraloops - for further examples and references see Gutell et al, 2002). The comparative method assumes that there is a generally maintained secondary structure for a group of sequences, and that the evolutionary processes of selection and mutation do not alter the structure and function of the molecule (Gutell et al., 2002).

This study incorporates an investigation of the secondary structures of the nuclear ribosomal spacers into phylogenetic analyses of relationships amongst three of the four major lineages in the Sapindaceae, the Xanthoceroideae, Hippocastanoideae and Dodonaeoideae, for which there is near complete sampling. The aims of this study are: 1) to generate secondary structure predictions for ITS and ETSf for Sapindaceae subfamilies Xanthoceroideae, Hippocastanoideae and Dodonaeoideae, 2) to use these predictions to assist alignment of sequences, 3) to use the alignment and the associated structural partitioning mask to select an appropriate model for phylogenetic analysis, and 4) to compare the phylogeny drawn from the analysis with that from the plastid data.

Methods

Sampling, DNA extraction, amplification and sequencing

Twenty-five ITS sequences were produced, with a further six sequences (including two outgroups from Burseraceae) added from GenBank, and 33 ETSf sequences plus two outgroups from GenBank, for a total combined matrix of 40 taxa (see appendix 4.1).

Total genomic DNA was isolated from leaf tissue using a CTAB protocol (Doyle and Doyle, 1990), and further cleaned using Jetquick (Genomed). The internal transcribed spacer region of nrDNA (ITS) was amplified using either ABI101/ABI102

(Sun et al., 1994) or ITS4/ITS5 (White et al., 1990) primer pairs, with ETS18S/9bp (Wright et al., 2001) for ETSf. Amplification and sequencing reactions were as outlined in Harrington and Gadek (2004) except for the addition of 10% dimethyl sulfoxide (DMSO) to all amplification reactions, which has been shown to lessen the likelihood of sequencing non-functional repeats (Buckler et al., 1997b). Double stranded PCR products were purified with the UltraClean™PCR Clean-up Kit (*MO BIO* Laboratories Inc., Solana Beach, California) and sequenced in the forward and reverse direction using dyenamic ET dye terminator kit (Megabase) chemistry (Amersham Biosciences). Cleaned PCR products were sequenced on a Megabase 1000 (Amersham Biosciences) at the Genetic Analysis Facility of James Cook University. Forward and reverse sequences were edited with ChromasPro Version 1.32 (Technelysium Pty Ltd).

Secondary structure prediction and sequence alignment

A similar approach to that used by other workers was followed, using a procedure that integrates sequence alignment and secondary structure prediction (Mai and Coleman, 1997; Subbotin et al., 2006; Biffin et al., 2007). This process combines a comparative approach by identification of evolutionarily conserved stem pairings and CBC across an alignment, with a thermodynamic approach that uses an energy model to predict a secondary structure with the lowest or near lowest free energy.

Both spacer sequences (minus 5.8S gene) were originally aligned by eye in BioEdit version 7.0.1 (Hall, 1999). Boundaries of spacers with the RNA genes were determined by comparison to published sequences following Harrington and Gadek (2004). The next step was to generate an estimation of potential conserved secondary structures across the spacers for all sequences. To accomplish this, three manually aligned datasets (ETSf, ITS1 and ITS2) were presented to the KNetFold server (http://knetfold.abcc.ncifcrf.gov/). The *k*-nearest neighbour algorithm used by KNetFold has been shown to be statistically more accurate at predicting structures than other programs (Bindewald and Shapiro, 2006). Given an alignment KNetFold combines thermodynamic and compensatory information to produce a common structure prediction at homologous positions across the alignment. The common structures are then used to readjust the initial alignment which is then resubmitted to KNetFold.

Secondary structure predictions for individual sequences were produced using the energy minimisation principle in RNAstructure version 4.1.1 (Mathews et al., 2004). Multiple individual optimal and suboptimal predictions were evaluated for conserved helices and where necessary generalised constraints as determined in KNetFold were placed on the folding algorithm.

The final determination of a common secondary structure mask across the sampled taxa (including outgroups) was achieved by combining the free energy structural predictions for individual taxa with a comparative sequence analysis. Because small changes in secondary structure may have occurred over evolutionary time between taxa, a consensus secondary structure is used for the alignment. A 50% majority rule measure is used, which has generally been adopted by recent researchers (Gardner and Giegerich, 2004; Gowri-Shankar and Rattray, 2006; Biffin et al., 2007; Subbotin et al., 2007) for determining whether nucleotide sites are involved in stem pairing. If the pairing does not occur in >50% of species in the alignment it is not maintained, and is designated as independently evolving.

Positional covariations (CBC) across the alignment were determined with cbcAnalyzer (Wolf et al., 2005). All secondary structures were drawn with *PseudoViewer* software (Han and Yanga, 2003).

Phylogenetic analysis - model selection

Selection of the appropriate nucleotide substitution model that 'fits' the data is the critical choice for phylogeny estimation (see Posada and Buckley, 2004 for multiple references). For RNA encoding data the choice is from a greater range of substitution models because of the secondary structure information. For instance, individual nucleotides contained within unpaired sites (single strand loops and lateral, internal or terminal loops – see table 4.1) evolve independently from other nucleotides and structures and are consequently under no structural constraint, so can be modeled using traditional DNA nucleotide substitution models (Whelan et al., 2001). For RNA stems, however, the assumption that each residue evolves independently does not hold, because substitution events that maintain or restore hydrogen bonding between paired sites are generally favored (selected for) over substitutions that disturb hydrogen bonding.

While some recent workers have chosen a numerical recoding scheme for analysing stem pairs (Smith et al., 2004; Subbotin et al., 2006), there is a range of Markov models of RNA sequence evolution incorporating correlated changes between stem pairing nucleotides in a RNA helix that have been developed (table 4.2). The models vary in complexity (fig. 4.1) and differ in their treatment of mismatch pairs (MM), symmetry or asymmetry between GC and CG, between AU and UA, between UG and GU, and the treatment of transitional states between CBC (e.g. A:U↔G:C) treated as two instantaneous changes (double substitution) or as a two staged transition via a stable intermediate G:U. For a full evaluation of these models see Savill et al. (2001). A similar likelihood estimation and permutation test (Felsenstein, 2003) approach to those used by Telford et al., (2005) and Biffin et al., (2007) was used to determine whether it is appropriate to use RNA specific models or standard four-state nucleotide substitution models for combined analyses of ITS and ETSf sequences.

Table 4.2: General description of RNA specific models implemented in software package PHASE version 2.0b 2005, http://www.bioinf.man.ac.uk/resources/phase.

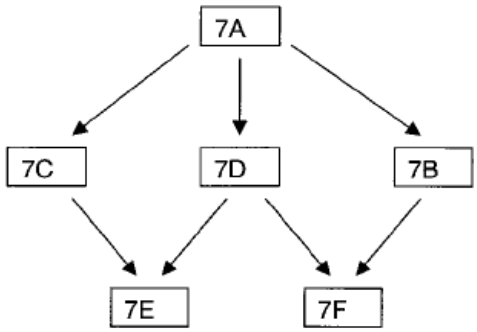
6 state models	Recognises only the six matching pairs AU GU GC UA UG GC ignoring the other 10 mismatch pairs
7 state models	Recognises the six matching pairs and is a simple treatment of mismatch pairs that groups all 10 mismatch pairs into a simple state
16 state models	Recognises the six matching pairs and includes all 10 possible mismatch states

For this study, given the small amount of data (40 sequences and length 940 basepairs - bp) and the relatively high amount of mismatch pairs (~ 5-7% of empirical stem pairings), and considering that the most complex and time consuming models (16-state models) potentially statistically overparamaterise the data (for review see Sullivan and Joyce, 2005), it was deemed most appropriate to compare and choose from the 7 state models.

To determine the best model or combinations of four-state nucleotide substitution models and RNA 7 substitution models, likelihoods were estimated using the Optimizer module of the PHASE software package (PHASE version 2.0b 2005 -

available from www.bioinf.man.ac.uk/resources/phase). The Optimizer program computes maximum likelihood (ML) estimates of designated evolutionary parameters (substitution model parameters and branch lengths) for a given topology and sequence alignment. The ML estimates were compared using the corrected Akaike Information Criterion (AIC_c), which provides a penalty for overparameterisation and provides a correction for the small sample size. The use of the AIC for guiding model selection has been shown to be superior to likelihood ratio tests because they allow for assessment of model selection uncertainty and model averaging (Posada and Buckley, 2004). The model with the lowest AIC_c value is generally considered the best. The relative plausibility of the other models is assessed using the theoretical approach of Burnham and Anderson (2003) by evaluating the ΔAIC_c of each model from the best model and adopting the following standards in terms of support for the model from the data - $\Delta \leq 2$ indicates substantial support, $4 \leq \Delta \leq 7$ considerably less support, $\Delta > 10$ no support.

Figure 4.1: Relationships between the RNA7 classes of models. The most complex model with the largest number of free parameters is 7A and the models at the head of each arrow are less complex derivatives of the model at the tail (adapted from Savill et al. 2001)



The topology used to test all models was that selected by AIC_c in MrAIC.pl 1.4 (Nylander, 2004), which tested for the best-fit model from the 56 available DNA nucleotide substitution models, with the trees and likelihood scores under different models estimated using PHYML (Guindon and Gascuel, 2003b). For both ETSf and ITS, the general-time-reversible (GTR) model, with variable rates (I) and gamma distribution (Γ) that allows for rate variation across sites was selected; this is equivalent

to the REV + I + Γ model implemented in PHASE. The same model (REV + I + Γ) was also selected for separate analyses of ITS and ETSf loops only datasets (data not shown). To determine the best-fit model and the utility of independent (four-state nucleotide substitution models) verses nonindependent substitution (RNA paired-site structural models) across the dataset and data partitions, a hierarchical series of AIKc comparisons were made. The data partitions were designated as ITS and ETSf, and while recognizing that ITS1 and ITS2 are functionally independent (Musters et al., 1990) and have possible different evolutionary histories, the more complex partition (ITS1 + ITS2 + ETSf) was not considered since it would over-parameterise the already small data partitions.

The different models of sequence evolution and the order in which they were run are listed in table 4.3. While it is also possible to run a multitude of other combinations of RNA models, e.g. ITS RNA7A with ETS RNA7B, initial tests using the AIC_c criterion of all RNA7 models on individual stem partitions for ITS and ETS indicated that RNA7C was the best model for each partition, so it was decided to run the same RNA7 model for each test. Two permutations tests were also applied to determine whether randomly repartitioned datasets of i) combined ITS plus ETSf loops partition and stems partition, and ii) ITS loops, ITS stems plus ETSf loops and ETSf stems (4 partitions) performs as efficiently, in terms of AICc, as the non-partitioned data. The null hypothesis of the test was that partitioning into stems and loops does not significantly improve the likelihood of the data. Maximum-likelihood and AICc scores were determined for 100 randomly repartitioned datasets, achieved by moving one column of the dataset while maintaining the original position of the secondary structure mask.

Phylogenetic analyses

To determine phylogenetic relationships amongst the 40 taxa (including two outgroups) the combined dataset with four structural partitions (ITS loops + ITS stems + ETSf loops + ETSf stems) was analysed with a Bayesian statistical method using Markov-chain Monte Carlo (MCMC) techniques in the *mcmcphase* program of the PHASE package. All loops were analysed under the REV model, while stems were analysed with the 'best-fit' RNA7 model (see results). For all partitions a discrete gamma

distribution of rates = yes, number of gamma categories = 6, invariant sites = yes was used, with all parameter values estimated directly from the dataset. The analyses were started with a random unrooted MCMC tree and run for 1,100,000 iterations with the first 100,000 iterations discarded as burnin, and with the posterior sampled every 150 cycles. The PHASE program *mcmcsummarize* was used to generate all statistics (including best tree and consensus tree) from each run. The likelihood scores and all the sampled parameters of the substitution model for all trees were examined to confirm that they had maintained stability after the burnin. Two other independent analyses with different starting points were performed to confirm convergence, and the trees generated were all combined (20,001 trees) and a 50% majority rule tree was generated using PAUP*4.08b (Swofford, 2002).

Table 4.3: Hierarchical selection process to determine the most appropriate model/s to analyse the combined ITS and ETSf datasets using a series of AIC_c comparisons in the order in which they were compared. REV = REV + I + 6Γ (includes invariant sites and six gamma rate categories).

Partition	Models for partitions
Combined	REV
ITS + ETSf	REV + REV
Combined loops + combined stems	REV + REV
Permuted data combined loops + stems (x100)	REV + REV
ITS loops + ITS stems + ETSf	REV + REV+ REV
ITS + ETSf loops + ETSf stems	REV + REV+ REV
ITS loops + ITS stems + ETSf loops + ETSf stems	REV + REV+ REV + REV
Permuted loops + stems + loops + stems (x100)	REV + REV+ REV + REV
ITS loops + ITS stems + ETSf loops + ETSf stems	REV + RNA7A + REV + RNA7A
ITS loops + ITS stems + ETSf loops + ETSf stems	REV + RNA7B + REV + RNA7B
ITS loops + ITS stems + ETSf loops + ETSf stems	REV + RNA7C + REV + RNA7C
ITS loops + ITS stems + ETSf loops + ETSf stems	REV + RNA7D + REV + RNA7D
ITS loops + ITS stems + ETSf loops + ETSf stems	REV + RNA7A + REV + RNA7A

Results and discussion

Alignment and secondary structure

The conserved length (164 bp) and very low levels of sequence variation in the 5.8S gene of ITS, along with no substitutions in the highly conserved positions of the spacers amongst all samples is a good indication that functional ITS sequences have been generated (Buckler et al., 1997b; Hershkovitz et al., 1999).

Figure 4.2a: ITS1 RNA transcript secondary structure for *Majidea fosterii*. Boxed region contains the conserved structural motif of Liu and Schardl (1994). Free energy (ΔG = kcal/mol) for each structure is shown.

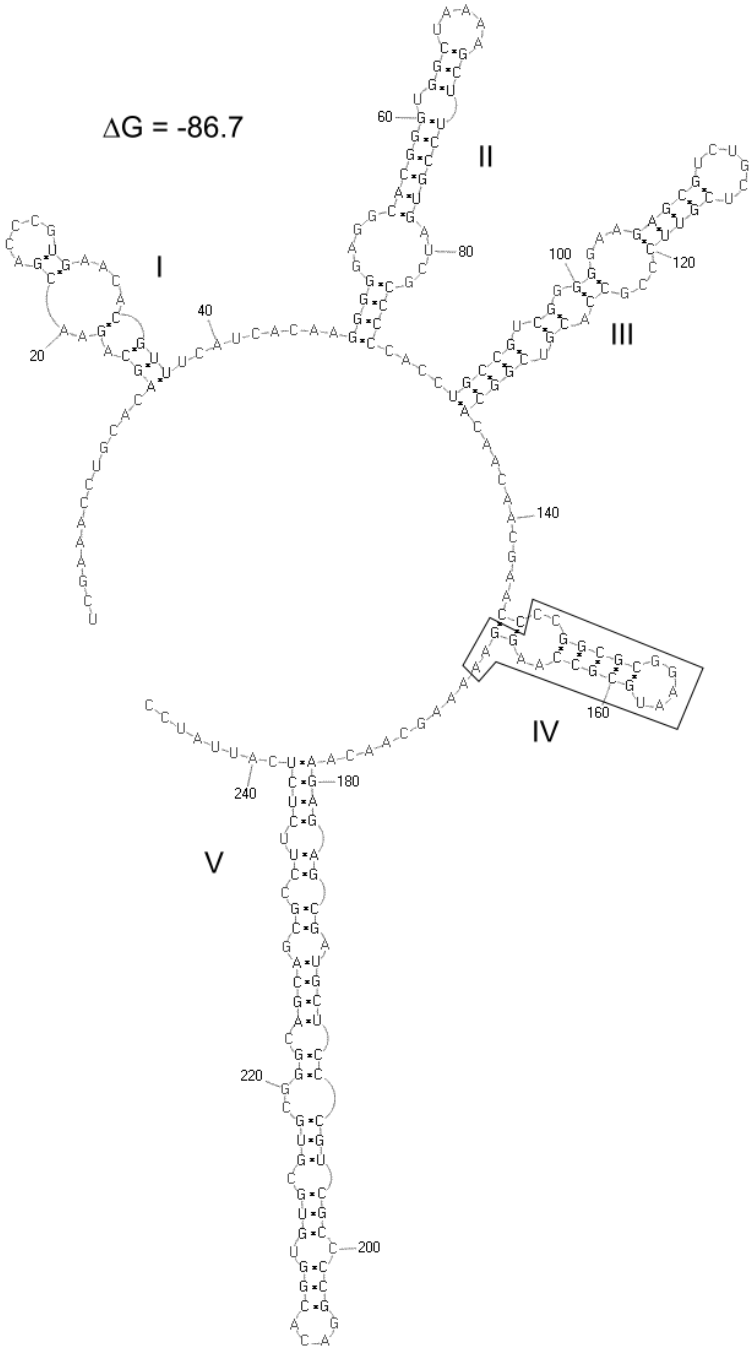


Figure 4.2b: ITS 2 RNA transcript secondary structure for *Majidea fosterii*. Boxed regions contain the key conserved structural motifs of Schultz et al. 2005.

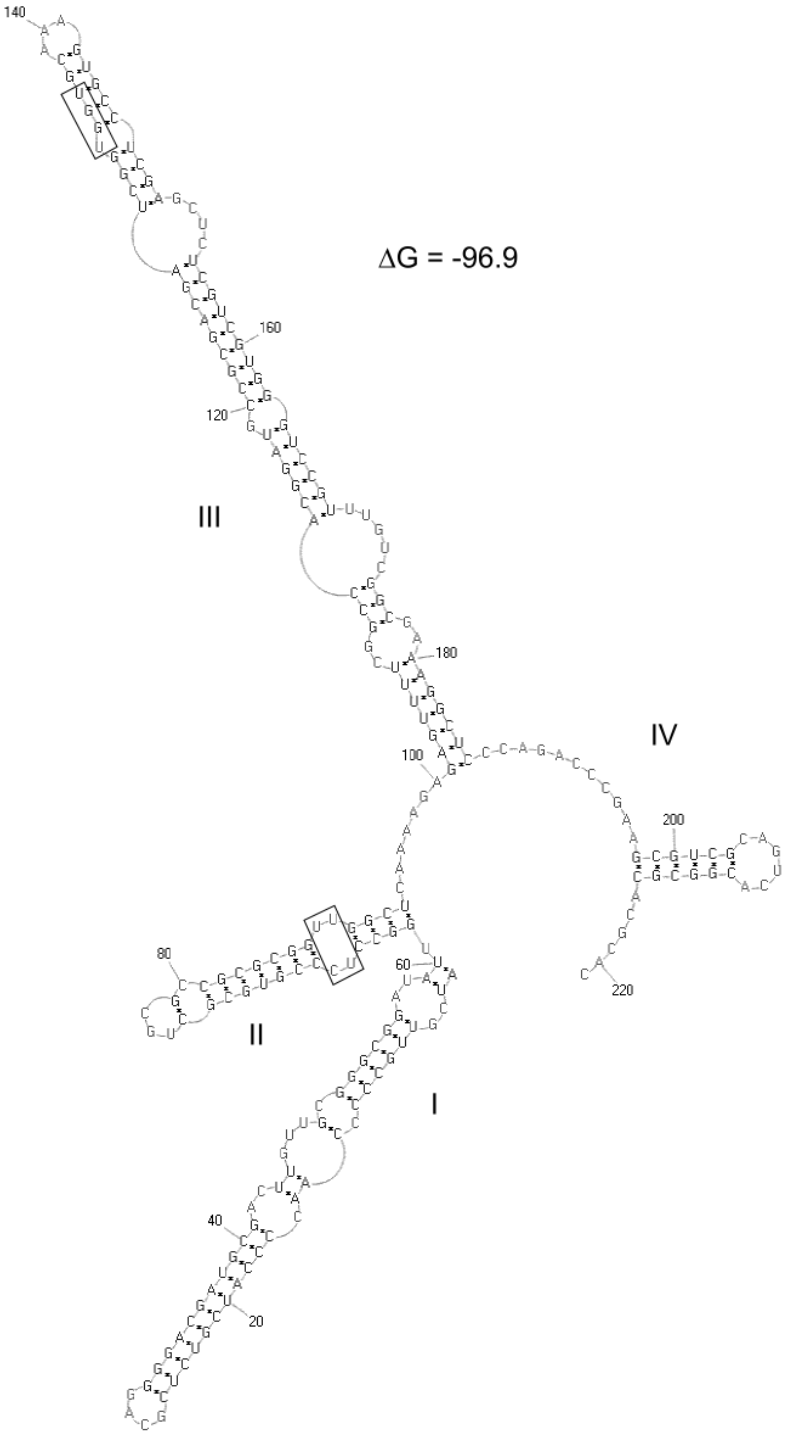
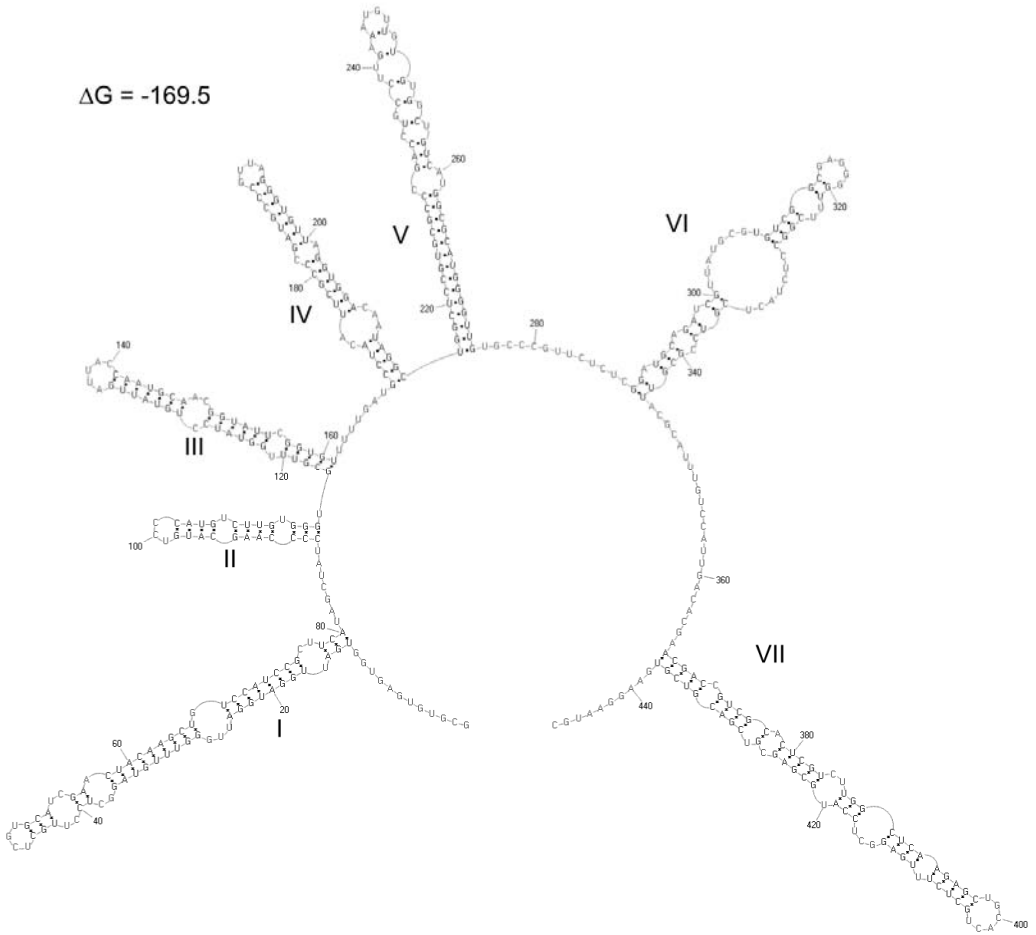


Figure 4.2c: Tentative secondary structure for the ETSf RNA transcript for *Majidea fosterii*.



The concatenated alignment with secondary structure mask is contained on the associated CD. Figure 4.2 a-c shows putative secondary structure models for a representative species for the three regions and highlights key conserved structural motifs. Within ITS there are 121 stem pairs (55 ITS1 and 66 ITS2) over nine helices. The structure of ITS1 is generally in accord with those described in other studies (Gottschling et al., 2001; Mayol and Rossello, 2001; Goertzen et al., 2003; Albach and Chase, 2004; Biffin et al., 2007). Helix IV of ITS1 (fig. 4.2a) is uniform in length and structure across the alignment with only a few nucleotide changes in the terminal loop, and contains the conserved structural motif of Liu and Schardl (1994). ITS 2 (fig. 4.2b) is generally more conserved than ITS1, making secondary structure prediction more consistent. ITS 2 has all the distinct characteristics of core structure described for the Eukaryotes by Schultz (2005): i) standard four helices (or four-fingered hand); ii) helix III as the longest; (iii) a pyrimidine mismatch loophole in helix II; and iv) helix III containing the conserved angiosperm UGGU motif 5' to the apex (Mai and Coleman, 1997). Helix I also displays length and sequence variability and the C- to G-rich transition from 5' to 3' described for angiosperms (Hershkovitz and Zimmer, 1996). Several of the helices of both spacers contain CBC with the most being between outgroup and ingroup (appendix 4.2a).

Approximately 445 bp were sequenced from the conservative region of the 5' ETSf which also contains the hyper-variable segment upstream of the 18S rRNA coding sequence (Volkov et al., 2003). The sequences start with a motif that has been shown to be highly conserved across a range of plant families, and is possibly part of a series of motifs that are involved in pre-rRNA metabolism as a signal for primary processing of rRNA (Polanco and Perez de La Vega, 1994). The generally conserved length of the region from this motif to the 18S gene has been suggested to correspond to a key function for this region in the processing of the rRNA gene transcript (Bena et al., 1998). Length conservation has also been shown to strengthen the prediction of secondary structure (Gardner and Giegerich, 2004).

The tentative secondary structure model for ETSf (fig. 4.2c) contains 136 stem pairs spread over seven helices of varying lengths and stabilities. While to my knowledge there have been no other ETS secondary structure predictions for flowering plants, the putative multi-stemmed structure is generally similar to the 10 stemmed structure proposed for the 700 bp 5'ETS of the yeast *Saccharomyces*

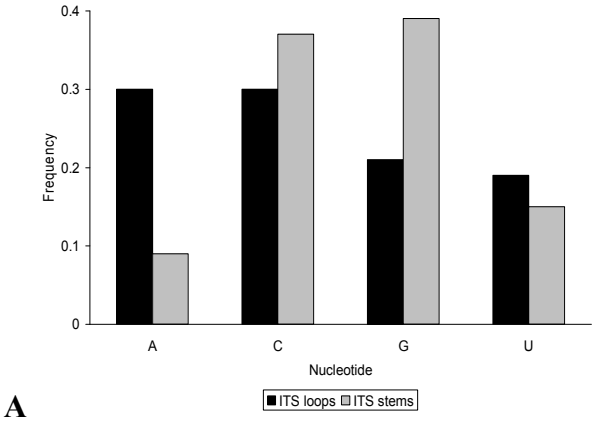
cerevisiae (Lee-Chuan and Lee, 1992). All of the stems are supported by CBC (appendix 4.2b).

While precautions were taken not to preferentially sequence paralogous copies (pseudogenes) in PCR reactions by addition of DMSO, the ETSf sequences for both species of *Llagunoa* have features of pseudogenes when compared to the other sequences (see Buckler et el., 1997). These features include a increased rate of molecular evolution in terms of numbers of CBC (appendix 4.2b), and decreased GC content due to a possible higher number of methylation-related mutations by deamination (C→T and G→A). While *Llagunoa* does have the lowest GC content, the proportions of the different nucleotides does vary widely across the alignment (table 4.4). Pseudogenes are also potentially identifiable by reduced secondary structure stability of the transcript in terms of their higher free energy (ΔG = kcal/mol) (Buckler et el., 1997). However while free energies also vary across the alignment (ΔG-198 to ΔG-137), *Llagunoa* is still within the range and at the lowest end (ΔG-141). The fact that ETSf from both *Llagunoa* could be amplified may be evidence that they are functional copies, since they maintain the conserved priming site CAUGGGCGUGUGAGUGGUGAU (also see phylogenetic analysis section below).

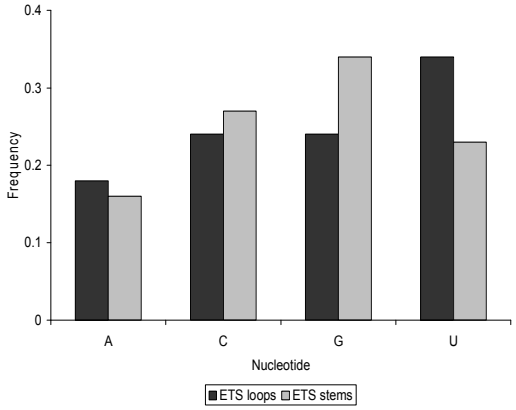
Spacer molecular evolution comparison

The ITS stems and ETSf stems had almost identical mean evolutionary rates (ETSf/ITS = 1.01), while ITS loops had a higher mean substitution rate than ETSf loops (ETSf/ITS = 0.7), demonstrating that there are different evolutionary rates in different parts of the dataset. There is distinct nucleotide compositional bias between the stems and loops of both of the spacers, but the bias is different in each spacer (fig. 4.3 a-c). There is also a strong bias towards overestimation of mismatch pairs (ITS empirical 0.07, RNA7C estimated mean 0.15, ETSf empirical 0.05, RNA7C estimated mean 0.14) due to the combination of a base-pair model with a variable rates model (Jow et al., 2002). In terms of MM pairs, both spacers have similar numbers of stem sites under the weakest evolutionary constraints for conserved structure.

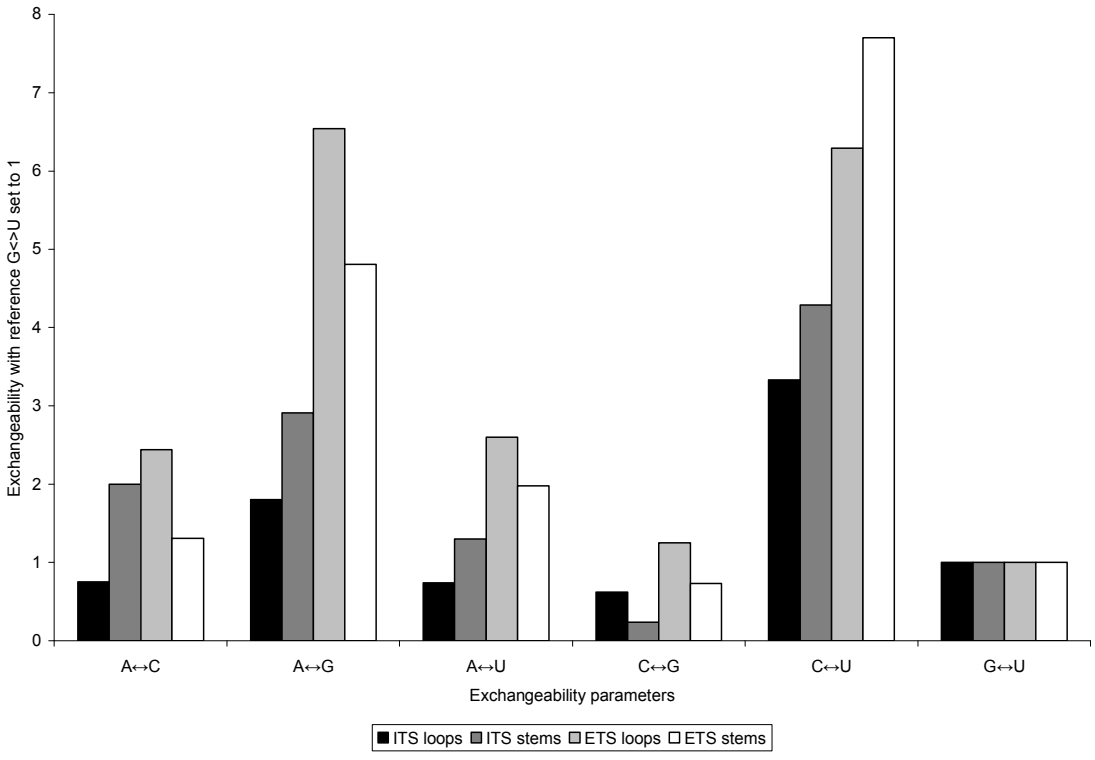
Figure 4.3: (A-B) Nucleotide frequencies comparisons estimated using separate GTR + I + Γ models for stem and loop partitions across both spacers, and (C) rate ratio (exchangeability parameters) for each category of nucleotide change estimated from the reference parameter rate (G↔U set to 1).



A



B



C

For the ITS data this study identifies a similar nucleotide mutational processes to that outlined by Biffin et al. (2007) for Myrtaceae ITS. The slower evolving and more stable helical regions are rich in the thermodynamically stronger (lower free energies) cytosine and guanine, while in loops there is evidence of a preference for adenine (3x more than in stems) which may be the result of their involvement in particular stabilised structural motifs (Gutell et al., 2000). A common

feature of stems and an indication of selection operating to maintain stability (base pairing) is a greater transition/transversion ratio, ie. higher A↔G and C↔U rate ratios (fig. 4.3c), with the single mutation still maintaining stable secondary structures through the stable intermediate GU/UG pairings, for example G:C (C↔U) G:U (G↔A) AU.

Table 4.4: Empirical nucleotide frequencies across the three spacer regions with other selected examples for ETSf within Sapindaceae.

	A	C	G	U	C/G
ITS1	22	33	28	16	61
ITS2	15	31	32	21	63
ETSf	16	25	29	30	54
Dodonaea	16	24	27	32	51
Diplokeleba	13	31	32	24	63
Hippobromus	15	27	30	27	57
Llagunoa	20	19	26	34	45

Similar mutational dynamics are found within ETSf, such as higher G/C content in stems than in the rest of the structure, and high A↔G and C↔U rate ratios. However across the whole molecule there is a higher proportion of uracil in ETSf than in the ITS (28% vs 17%), with the adenine bias in loops replaced by a uracil bias, and there is also a far higher proportion of the less thermostable A:U/U:A pairs and transitional G:U pairs than in ITS (table 4.5). The basis for the selective preference for these weaker bonding stem pairs remains to be elucidated. Because an adenine bias in RNA loops is the generally accepted norm, as they can be associated with several structural motifs (Gutell et al., 2000), and since in terms of adenine content the ITS is like other RNA molecules (Telford et al., 2005; Biffin et al., 2007), a further permutation test was performed on the ETSf loop and stem data only to test whether the bias toward uracil was possibly structurally significant. The null hypothesis was that in the absence of selection for specific nucleotide content in loop and stems, a randomly partitioned dataset should perform equally in terms of AIC_c to the loop and stem data. Using the REV + REV model, the stem and loop partition (AIC_c = 9996) strongly outperformed the 100 random partitions (AIC_c = 10018 ± 7.6). As with the stem pairs, the biological basis for the selective preference for uracil in ETSf loops is unknown.

Table 4.5: Base pair frequency, mutability of base pairs (net rate at which a base pair changes to other states from the best-fit rate matrix which is rescaled so that its average substitution rate is 1.0 i.e. <1 = evolve more slowly than average) and substitution rate parameters (r_d – double transitions, r_v – double transversions, r_f – single transition forward, r_b– single transition reverse) inferred using the RNA7C model.

	fGC	**fCG**	**fAU**	**fUA**	**fGU**	**fUG**	**fMM**
Base pair frequencies							
ITS	0.30	0.27	0.10	0.06	0.06	0.07	0.14
ETS	0.19	0.23	0.16	0.11	0.10	0.07	0.14
Mutabitity of base pair							
ITS	0.68	0.52	0.53	0.79	2.28	1.65	2.07
ETS	0.82	0.45	0.58	0.77	1.48	2.53	1.65
Substitution rates	**r_d**		**r_v**		**r_f**		**r_b**
ITS	0.04		0.00		0.35		0.84
ETS	0.03		0.003		0.38		0.73

The pivotal study of Biffin et al. (2007) indicated that, unlike other RNA molecules, both model selection and substitution rates indicate that there is evidence for toleration of the 'intermediate' base pairs (GU/UG) in the mutational dynamics in stems in ITS of Myrtaceae. The results presented here for Sapindaceae ITS concur with those of Biffin et al. 2007, and suggest a similar absence of selection against the intermediate base pairs for ETSf. Both spacers have nil or negligible rates of double transitions (r_d) and double transversions (r_v), and similar overall 'forward' (AU→GU, UA→UG, GC→GU = r_f) and 'backward' (GU→GC, GU→AG, UG→CG, UG→UA = r_b) single transition rates; they differ, however, in that there is bias in the mutation rates of the symmetry pairs (AU/UA, GC/CG). Within ETSf the 'forward' transition from AU to the less stable intermediate GU is nearly twice that of ITS, while the substitution rate is higher but virtually reversed for the transition from UA→UG. The intermediate GU has nearly twice the 'backward' transition rate to the stable GC pair in ITS than in ETSf, while intermediate pairs are fractionally more likely to change to stable AU/UA in ETSf, maintaining the higher rate of AU/UA pairings and overall uracil content than in ITS. The greater GU/UG content in ETSf may have biological significance due to the increased functional variety of

their distinctive structural, chemical and thermodynamic conformational properties which these pairings confer (Gautheret et al., 1995; Varani and McClain, 2000).

Mutation in stems of both spacers follows a very similar pattern (table 4.5). A mutation in a MM pair or in the intermediates (GU/UG) is up to four times more likely to occur than in the other more structurally conserved pairs. The only general difference between the spacers is that the highest mutability in ITS is found in GU pairs (table 4.5 - 2.28) while for ETSf it is in the asymmetric UG pair (2.53).

The overall differences in the patterns of nucleotide organisation, content and subsequent secondary structure association within the spacers are possibly due to the differing roles and sequential timing of cleavage events of various parts of the spacers and other interrelated processes that occur in precursor rRNA processing pathways.

Model selection

For each individual spacer (data not shown) and a combined data matrix partitioned into separate loops and stems (ITS loops + ITS stems + ETSf loops + ETSf stems) returned substantially better AIC_c scores than non-partitioned or randomly partitioned data (table 4.6). These results are indicative that models which consider secondary structure interactions more accurately represent the mutational processes that operate at different loci within a RNA encoding sequence.

The 'best fit' evolutionary model for analyses of the combined data matrix as determined by the AIC_c for the separate stem partitions for the each spacer regions was the covariation RNA7C model. This model is a simpler version of the RNA7A model and does not allow for double substitutions, treating all nucleotide changes as single substitutions. Selection of a RNA model over standard 4-state models is evidence that over half of the nucleotides in both of the spacers that are involved in stem pairing are not evolving independently, and dinucleotide evolutionary models for stems perform 'better' (lower AIC_c) than those that only allow independent evolution of nucleotides.

Table 4.6: Comparison of AICc scores for various partitioning schemes beginning with the simplest. While not all of the model scores can be directly compared because of the differing partitioning schemes the best RNA7 model is highlighted in bold. In all instances REV = REV + I + Γ. AICc = AIC + 2k(k +1)/n – k where AIC = -*L* + 2k - 1, with k the number of estimable parameters, *L* the likelihood estimate and n the number of nucleotides positions (949).

Partition	***Models***	***k***	**L**	***AIC$_c$***
Combined	REV	11	-10488	20998
ITS + ETSf	REV + REV	22	-10444	20933
Combined loops + stems	REV + REV	22	-10413	20871
Permuted data combined loops + stems (x100)	REV + REV	22	-10475±4.8	20995
ITS loops + ITS stems + ETSf	REV + REV+ REV	33	-10374	20816
ITS + ETSf loops + ETSf stems	REV + REV+ REV	33	-10442	20920
ITS loops + ITS stems + ETSf loops + ETSf stems	REV + REV+ REV + REV	44	-10357	20806
Permuted data loops + stems + loops + stems (x100)	REV + REV+ REV + REV	44	-10432±5.2	20956
Combined loops + combined stems	REV + RNA7C	28	-9378	18813
ITS loops + ITS stems + ETSf loops + ETSf stems	REV + RNA7A + REV + RNA7A	78	-9316	18802
ITS loops + ITS stems + ETSf loops + ETSf stems	REV + RNA7B + REV + RNA7B	72	-9318	18792
ITS loops + ITS stems + ETSf loops + ETSf stems	**REV + RNA7C + REV + RNA7C**	**56**	**-9322**	**18763**
ITS loops + ITS stems + ETSf loops + ETSf stems	REV + RNA7D + REV + RNA7D	44	-9338	18768
ITS loops + ITS stems + ETSf loops + ETSf stems	REV + RNA7E + REV + RNA7E	40	-9341	18766

Phylogenetic analyses

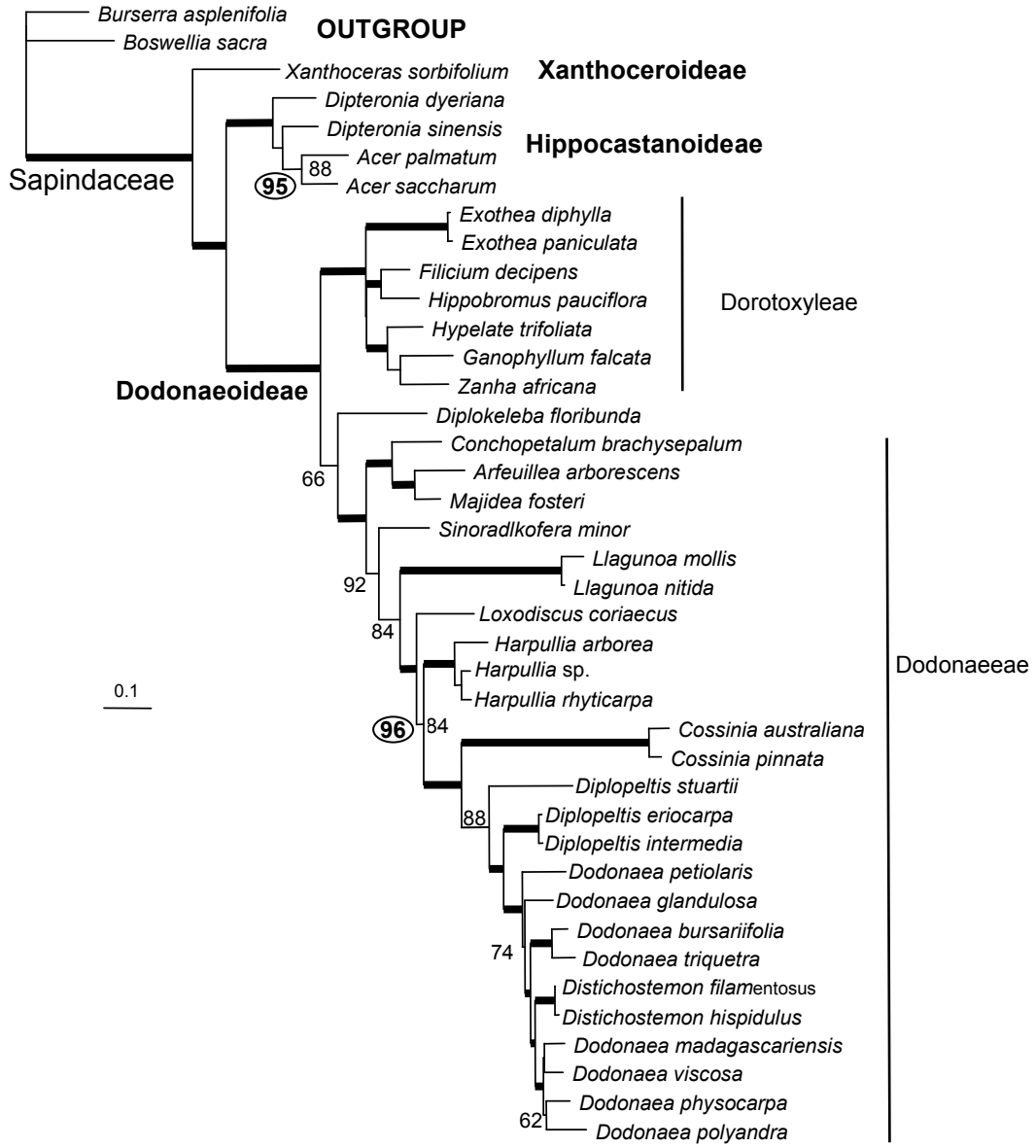
The Bayesian estimation of phylogeny using the REV + RNA7C + REV + RNA7C model across the loop and stem alignment for the combined spacer data is well resolved and the branching order is congruent with a plastid phylogeny using combined *rbcL* and *matK* genes (Harrington et al., 2005; Chapter 2) except for the placement of *Diplokeleba floribunda* (fig. 4.4). The three subfamilies are strongly

supported, with the node determining relationships between these clades found in 96% of the sampled topologies.

The major goal of molecular phylogenetic analyses is to generate hypotheses of evolutionary relationships as near to the 'correct' phylogeny as possible. When compared to the phylogenies produced using traditional 4-state models for the spacer partitions i.e. REV + REV there is virtually no difference in branching order and support for nodes receiving 98-100% posterior probabilities (PP). However, the simpler model generally overestimates confidence for all of the weakly supported branches (< 95% PP). For example in the 4-state model analysis *Diplokeleba* receives 89% PP for its sister relationship to Dodonaeeae, whereas this relationship is only found in 66% of the RNA7C model topologies (fig. 4.4). Alternatively, the plastid topology had weak support (71% bootstrap – chapter 1) for a sister relationship between *Diplokeleba* and Doratoxyleae. There are also two nodes (values shown in circles in fig. 4.4) in the 4-state model tree that receive ≥ 95% PP ('good' support) that do not receive equivalent support in the RNA model phylogeny.

The RNA7C analysis fails to resolve with any confidence the direct sister relationships of the two species of *Llagunoa*, the sequences for which were identified as having some features of pseudogenes; however, they are confidently placed within the well supported grouping of members of Dodonaeeae, as also identified in the plastid analyses.

Figure 4.4 (next page): Phylogeny for 40 taxa of Xanthoceroideae, Hippocastanoideae and Dodonaeoideae (Sapindaceae) of combined ITS and ETS secondary structure alignment estimated using Bayesian analyses with the model REV + I + Γ for the separate loop partitions and RNA7C for each of the separate stem pair partitions. Branches with posterior probabilities ≥ 95 % are shown in bold, with other values > 50% shown next to the branch they support. Numbers in circles are posterior probabilities ≥ 95 % from the four-state model analysis for the two partitions (REV + REV) that don't receive equivalent support in this analysis. Branch lengths are the mean posterior estimates of the evolutionary distances between the two complementary clades they define i.e. the longer the branch, the bigger the evolutionary distance between the incident nodes.



The use of oversimplified models of nucleotide substitution is known to produce spurious levels of support in terms of high posterior probabilities and can confound estimates of phylogeny (Buckley, 2002; Huelsenbeck and Rannala, 2004). In this study we can have confidence in well supported branches that are the result of using a combination of models that more accurately approximate sequence evolution in each data partition.

The methods used in this study do not allow for 'better' resolution to the tree in terms of an increased number of nodes with PP ≥95%, due to the decreased numbers of independently evolving nucleotides attributable to the paired-site model; however, they do provide a more accurate and robustly arguable hypothesis of

phylogenetic relationships and estimation of base composition and relative nucleotide substitution rates (branch lengths) for branches within the tree.

Conclusion

This study provides a framework for the incorporation of alignment and analyses procedures to avoid the potential pitfalls associated with the molecular structural design and evolution of ribosomal spacers. Secondary structure predictions have been developed for nuclear ribosomal spacer regions that hypothesize that over half of the nucleotides are potentially involved in stem-pairing. The incorporation of separate non-independent models of nucleotide evolution for these sites allows estimation of base composition and substitution rates that aid an understanding of the evolutionary processes and selection forces operating on these ribosomal spacers.

A molecular phylogeny for Dodonaeoideae, Hippocastanoideae and Xanthoceroideae (Sapindaceae) has been produced that has incorporated separate models that account for the variation in nucleotide evolution in the independent (loops) and covarying (stems) regions of the ribosomal spacers. By using models that incorporate potential functional constraints on nucleotide evolution in stem regions the generation of potentially falsely high confidence levels can be avoided for branches in the phylogeny, and also the comparative estimates of branch lengths generated more appropriately represent the evolutionary history between the sequences.

Chapter 5: Understanding the origins of the Australian flora: plant adaptation and evolution in increasingly arid environments - phylogeny and evolutionary history of Hopbushes (*Dodonaea* – Sapindaceae) and Pepperflowers (*Diplopeltis* – Sapindaceae) based on nuclear ITS and partial ETS sequences

Sapindaceae are principally a tropical family, within Australia the majority of species are restricted to Queensland rainforests; however, within the largest by far Australian genus *Dodonaea* Miller (Hopbushes), the majority of species are found in temperate and arid communities, and in the tropics none is found within rainforests. It is the only genus of Sapindaceae in Australia whose species are found in south-western Australia and Tasmania.

Dodonaea comprises ca. 64 species (West, 1984; 1987), of which 61 species are endemic to Australia. *Dodonaea polyandra* also occurs in the Western District of Papua New Guinea, while *D. viscosa* extends from Australia throughout Southeast Asia/India, through the islands of the Indian and Pacific Oceans to the tropics and temperate zones of South America and Africa. At various times a number of other extra-Australian species have been recognised (Sherff, 1945; 1947; Brizicky, 1963; Leenhouts, 1983), all of which appear to be closely related to *D. viscosa*. *D. madagascariensis* is endemic to the uplands of central Madagascar, and is one of the few amazing examples of disjunct distributions within a genus from Australia to Madagascar.

Within Australia, *Dodonaea* grow in a wide range of habitats across the continent from the arid zones of central and Western Australia, across temperate southern Australia to the savannas and woodlands of the tropics. They are prostrate or erect multistemmed shrubs or trees of woodland, forest or shrubland communities, and are absent only from closed canopy rainforest, dense heath and alpine communities. The small petalless flowers have pollen dynamics and structure obviously adapted for wind pollination (Delphino, 1890; Reddi et al., 1980; Keighery, 1982; West, 1982a; West, 1984). While anemophily is exceedingly rare amongst the predominantly tropical Sapindaceae, the structure of the pollen of the closely related *Distichostemon* and *Diplopeltis* is also possibly related to wind

pollination (Muller and Leenhouts, 1976). This small group of genera (*Dodonaea*, *Diplopeltis* and *Distichostemon*) is a further paradigm of wind pollination being phylogenetically derived from a substantially larger predominantly animal pollinated group of genera within Sapindaceae (Soltis et al., 2005). While virtually all other Sapindaceae flowers have extrastaminal nectiferous disks, they are highly reduced and intrastamminal as part of a gynophore in *Dodonaea*, with no nectar produced (West, 1980).

The first infrageneric categories within *Dodonaea* were erected by Miquel (1844), who recognised two sub-sections based on leaf morphology: Section *Eu-Dodonaea* (simple leaves) and Section *Remberta* (pinnate leaves). The Sectional classification was broadened by Bentham (1863) with the inclusion of 39 species in five series – the pinnate-leaved species (Series *Pinnatae*) and four simple-leaved series further divided on capsule appendage morphology (Series *Cyclopterae*, *Platypterae*, *Cornutae* and *Apterae*).

There have been two other extensive reviews of the genus. The first by the great German taxonomist Radlkofer (1933) was part of a monograph of the whole of Sapindaceae. He placed *Dodonaea* in the tribe Dodonaeeae, within Dyssapindaceae (=Subfamily Dodonaeoideae) together with *Loxodiscus*, *Diplopeltis* and *Distichostemon*. *Dodonaea* and *Distichostemon* are considered closely related (Muller and Leenhouts, 1976), with both having regular flowers that lack petals and an insignificant disk which is intrastaminal, but are distinguished from one another predominantly on stamen number: *Loxodiscus* and *Diplopeltis* have zygomorphic flowers, are petaloid and have a nectariferous extrastaminal disk.

Within Sapindaceae Radlkofer (1933) believed that evolution proceeded via a series of reductions in morphological structures, so he regarded Dodonaeoideae as derived from Sapindoideae, and Dodonaeeae was numbered the 12th derived tribe placed between Cossinieae and Doratoxyleae (Chapter 2 – table 2.1). In an examination of macromorphology and pollen micromorphology of Sapindaceae, Muller and Leenhouts (1976) supported Radlkofer's concept of two subfamilies but they proposed a direction of evolution that is the reverse of his scheme. They suggested that the Dodonaeoideae displayed relictual features in their pollen and morphology, and that this along with their predominantly austral distribution, was evidence that this subfamily was an "assemblage of relicts", whereas Sapindoideae were more derived. In this system Dodonaeeae were suggested to have "antedated"

all other tribes in the family, with *Dodonaea* "the most primitive genus" in this tribe (Muller and Leenhouts, 1976). The results of recent Bayesian phylogenetic analyses of molecular data for Sapindaceae which have included near complete sampling of Dodonaeoideae (Harrington et al., 2005; Harrington et al. 2009; Harrington et al., in review), place *Cossinia* sister to *Dodonaea*, *Distichostemon*, and *Diplopeltis* within Dodonaeoideae, and with the next sister relationship being *Loxodiscus*.

Generally *Dodonaea* has been perceived to have had a relatively long evolutionary history in Australia, with pollen attributed to *Dodonaea* (aff. *D. triquetra*) dating from the Late Eocene about 40 million years ago (Mya) (Martin, 1994; Martin, 1997). However Bayesian relaxed clock molecular dating analyses (Harrington et al., in review) point towards a Late Miocene origin for *Dodonaea* – stem 10.4 (Mya) 95% higher probability distribution (HPD) 6-15.4 Mya – which suggests a recent radiation for the genus in Australia associated with the increased continental aridity from the Late Miocene to the Recent.

Radlkofer (1933) recognised 54 species (52 Australian) of *Dodonaea* which he classified into three series, *Cyclopterae*, *Platypterae* and *Aphanopterae*, and six subseries. Within each of the series, the species were grouped into two subseries largely on seed and fruit characters. It appears that the most important classifiers were the presence or absence of an aril, the type of hyaline layer covering the seed, and the presence or absence and types of glandular structures of the leaves. As a result simple- and compound-leaved species were grouped into the same subseries in some cases.

The most recent revision of the genus undertaken by West (1984) recognised six species groups using a combination of characters and postulated evolutionary trends in breeding system, morphology and extant distributions. Group 1 contained species with the greatest number of primitive character states and group 6 the most derived states. For example possession of an aril was regarded as an advanced character. With regard to their distributions, West (1982, 1984) recognised in general that several groups showing advanced character states included species with restricted distributions, while the widespread species belonged to those groups with mostly primitive character states. The infrageneric groups with a majority of primitive character states also contained species in which adaptations to aridity were pronounced. To account for the current distribution patterns and phylogenetic relationships, West (1984) considered that there were two possible hypotheses to

explain the origin of the arid zone taxa. However West (1984) was unable to determine whether those species that are now endemic to arid zones had evolved from previously more widespread species of the wetter, pre-arid environment of ancient Australia or relatively recent invaders from surrounding temperate and tropical areas

All *Dodonaea* for which a chromosome number are known (17 species and numerous regional subspecies and extra-Australian species of *D. viscosa*) have been recorded as n=14, 2n=28 (Love, 1975; 1976; 1984; West, 1984; Gill et al., 1990; Oginuma et al., 1997), which suggests that allopolyploidy has not prompted speciation in *Dodonaea*. The taxonomic treatment of West (1984) recognised 63 species of *Dodonaea* one of which was a putative hybrid (*D. tepperi*). Her review of herbarium collections and her own field observations suggested that there could be a number of other possible hybrid taxa or hybrid zones between a number of sympatric taxa.

Many of the species of *Dodonaea* are important components of Aboriginal pharmacopoeia, being used to treat a variety of ailments and also to poison fish (Ghisalberti, 1998). Because of their tolerance to dry environments, *Dodonaea* are also becoming increasingly important in horticultural plantings.

The main goals of this study are to use molecular data to test the various classifications and hypotheses of interspecific relationships based on morphology within *Dodonaea*. Once a well-supported estimate of phylogeny has been established, it will be possible to gain an understanding of the temporal and spatial history of *Dodonaea*, an important element in the Australian flora.

To choose an appropriate molecular marker a number of chloroplast DNA (*trnL* intron*, trnL-trnF* spacer, *matK*, *psbA-trnH*, *ndhF*, and *rpl16*) and nuclear DNA regions (internal transcribed spacers ITS, partial external transcribed spacer ETSf and partial rpb2) were analysed for variability and utility for resolving relationships amongst a sample taxon set. All of the plastid and rpb2 datasets showed little to no variability amongst a sample taxon set, while both the ITS and ETSf appeared to provide some potential for resolving relationships within *Dodonaea*.

Materials and Methods

Taxon sampling and outgroup selection

All of the currently recognised species of *Dodonaea* were sampled for this study. Samples were either field collected, taken from the Australian National Botanic Gardens (ANBG) living collection, grown from vouchered seed collections deposited at ANBG, or extracted from herbarium specimens. A list of species sampled, voucher information and Genbank accession numbers for sequences are given in appendix 5.1.

Preliminary analysis indicated that all six species of *Distichostemon* were ingroup taxa as they nested within *Dodonaea*. Seven Australian species of *Harpullia*, all species of *Diplopeltis*, two species of *Cossinia* and *Loxodiscus coriaceous* were included as outgroups.

DNA extraction, amplification and sequencing

Total genomic DNA was isolated from leaf tissue using a CTAB protocol (Doyle and Doyle, 1990), and further cleaned using Jetquick (Genomed). For *D. polyandra* multiple samples of DNA extracted from fresh or silica dried leaf material did not produce amplified product via the polymerase chain reaction (PCR). However successful product was obtained from CTAB extractions of seedling grown root material, and from a biopsy of the vascular cambial zone from under the bark of branches.

The ITS region was amplified using either ABI101/ABI102 (Sun et al., 1994) or ITS4/ITS5 (White et al., 1990) primer pairs, ETSf was amplified using ETS18S/9bp primer pairs (Wright et al., 2001). Amplification and sequencing reactions were as outlined in Harrington and Gadek (2004), except for the addition 1 μL of dimethyl sulfoxide (DMSO) in all amplification reactions to lessen the likelihood of sequencing non-functional repeats (Buckler et al., 1997a). For some older herbarium samples a 'touchdown' protocol as described in Becerra (2003) was employed. Double stranded PCR products were purified with the UltraClean™PCR Clean-up Kit (*MO BIO* Laboratories Inc., Solana Beach, California) and sequenced in the forward and reverse direction using dyenamic ET dye terminator kit (Megabase) chemistry (Amersham Biosciences). Cleaned PCR products were sequenced on a Megabase 1000 (Amersham Biosciences) at the Genetic Analysis Facility of James Cook University.

Alignment and secondary structure prediction

Forward and reverse sequences were edited with Sequence Navigator 1.0.1. (PE Biosystems Inc., Foster City, California) and in the first instance aligned manually in BioEdit version 7.0.1 (Hall, 1999). Boundaries of ITS1 and ITS2 were determined by comparison to published sequences following Harrington and Gadek (2004). The ribosomal 5.8S gene (164 basepairs - bp) was invariant in length and nucleotide composition across both ingroup and outgroup sequences, so was excluded from the alignment.

For secondary structure prediction the putative models for ITS1, ITS2 and ETSf for Dodonaeoideae delineated in Harrington et al. (2009) were used. For each region the consensus secondary structure mask was added to the alignment, and in a few instances this mask and the initial alignment were adjusted to account for novel secondary structure features, particularly in helix II of ITS1, where there is a up to 29 bp deletion within some members of the outgroups and all of the ingroup (see results). The secondary structure masks were also adjusted to lessen the number of mismatch pairs across the alignment. Potential positional covariations (compensatory base changes - CBC) across the alignment were determined with cbcAnalyzer (Wolf et al., 2005). The alignment and trees generated from this study can be accessed at TreeBASE under the study number.

Model selection and phylogeny estimation

A Bayesian perspective was used to develop and interpret hypotheses of phylogenetics relationships amongst the sampled taxa. A similar likelihood estimation approach to that used by Telford et al. (2005) and Biffin et al. (2007) was implemented to determine whether it is appropriate to use RNA specific models compared to standard 4-state nucleotide substitution models for combined analyses of ITS and ETSf sequences. The corrected Akaike Information Criterion (AICc) was used to evaluate combinations of RNA specific paired site models (excluding the most complex and time consuming RNA16 set of models, and the RNA6 models that exclude mismatch pairs) and the standard 4-state DNA models as implemented in the Optimizer module of the PHASE software package (PHASE version 2.0b 2005, http://www.bioinf.man.ac.uk/resources/phase). For a detailed description of all models see Savill et al. (2001). The RNA7 range of models tested recognise the six matching stem pairs (A↔U, G↔U, G↔C) and is a simple treatment of mismatch

pairs that groups all 10 possible mismatch pairs into one state. The topology used to test all model combinations was that selected by best AICc score in MrAIC.pl 1.4 (Nylander, 2004) for the combined alignment, where the trees and likelihood scores under the 56 different models tested are estimated using PHYML (Guindon and Gascuel, 2003a).

Because there were 19 informative indels (table 5.1) that could be coded and added to the dataset for phylogenetic analyses, the data were first analysed without coded indels to generate model statistics for the ITS and ETSf loop and stem alignments. Both the single data partitions and the combined dataset were analysed with four structural partitions (ITS loops + ITS stems + ETSf loops + ETSf stems) with a Bayesian statistical method using Markov-chain Monte Carlo (MCMC) techniques to sample from the posterior probability density in the *mcmcphase* program of the PHASE package. All loops were analysed under the REV model (equivalent to GTR), while stems were analysed with the 'best-fit' RNA7 model (see results – table 5.2). For all partitions a discrete gamma distribution of rates = yes, number of gamma categories = 6, invariant sites = yes were used, with all parameter values estimated directly from the dataset in all analyses. Proposal priorities for each model were set to the number of free parameters estimated for each model (REV = 11, RNA7C = 17).

The analyses were started with a random unrooted MCMC tree and run for 1,100,000 iterations with the first 100,000 iterations discarded as burnin, and with the posterior sampled every 500 cycles. The PHASE program *mcmcsummarize* was used to generate all statistics (including mean model parameters and consensus tree) from each run. The likelihood scores and all the sampled parameters of the substitution model for all trees were examined to confirm that they had maintained stability after the burnin. Two other independent analyses with different starting points were performed to confirm convergence, and the trees generated were all combined (6,000 trees) and a 95% majority rule tree was generated using PAUP*4.08b (Swofford, 2002).

To determine phylogenetic relationships amongst *Dodonaea* (plus samples from four outgroup genera), 19 coded indels (coded A/U) were added to the loops partition in the spacer region in which they occur. Three independent analyses of the separate and combined data partitions were carried out as above. In evaluating the

results of all analyses, posterior probability values (PP) ≥95% are interpreted as indicating a degree of support from the molecular data for elements in the topology.

Relative evolutionary rates analysis

Bayesian Markov chain Monte Carlo (MCMC) phylogeny and molecular evolutionary rates estimations of the combined aligned matrix without indels were performed using the BEAST package (Drummond and Rambaut, 2007). For each ITS and ETSf stem and loop partition, the general-time-reversible (GTR) model with gamma distributed rate variation and an assumed proportion of invariable sites were applied. All parameters were free to vary independently across partitions. A relaxed clock with an uncorrelated lognormal distribution with no fossil constraints was implemented. The mean substitution rate was fixed to one to allow for an estimation of the relative substitution rates (estimated in substitutions per site) across the phylogeny (i.e. a ratogram). The tree branching prior was assumed to follow a traditional Yule speciation process birth rate of a constant speciation rate per lineage. Two independent MCMC chains for the combined dataset were run for ten million generations sampling the topology every 10,000 generations and parameter values every 500 generations. The output was examined using Tracer v1.4, and summarised (excluding the appropriate burn in) using Logcombiner v1.4.6, and TreeAnnotator v1.4.6 was used to generate a maximum clade credibility chronogram scaled to mean node heights (all part of the BEAST package http://beast.bio.ed.ac.uk/). Divergence times were determined by scaling the relative node heights into time by setting the divergence of *Dodonaea* from *Diplopeltis* to 10.4 Mya (95% HPD 6-15.4), corresponding to the estimated date for this node from Harrington et al. (in review).

Relative rate test

To determine whether there is any lineage specific variation in evolutionary rates amongst the sampled *Dodonaea*, a Bayesian relative rates test was performed which is considered to overcome many of the shortcomings of other forms of relative rate tests (Wilcox et al., 2004). For this test the posterior probability distribution of lengths of all branches was used from the 1800 phylograms (excluding burnin) returned from the previous analyses. Summed branch lengths for all trees were compiled in Cadence v1.0 (Wilcox et al., 2004) from each ingroup terminal to the MRCA of all the descendents of the ingroup, i.e. crown *Dodonaea*. Rates of evolution were considered to be significantly different between two taxa if their 95%

confidence interval of the posterior probability of their summed branch lengths did not overlap.

Diversification rate

Both the ultrametric trees and phylograms from the Bayesian relaxed clock with an uncorrelated lognormal distribution analyses were used to analyse diversification rates within *Dodonaea*. Since there is near complete sampling of all currently described species of *Dodonaea*, no correction measures or the use of simulated datasets were required.

The tempo of extant lineage accumulation within *Dodonaea* over time (speciation minus extinction) was visualised with a lineage through time (LTT) plot (using Genie, ver. 3.0; Pybus and Rambaut, 2002). The logarithm of the number of extant lineages against node height were plotted for the maximum clade credibility chronogram scaled to mean node heights and with all outgroups removed. A constant birth-death speciation model will return a straight line (Barraclough and Vogler, 2002), and deviation from a constant rate was tested for using the γ-statistic, which describes the distribution of nodes within a phylogeny under the assumption that if the clade diversified with constant rate it follows a normal distribution with a mean of zero and a standard deviation of one. The null hypothesis of a constant rate is rejected if $\gamma<-1.645$ (speciation rate decrease towards present) and $\gamma>+1.645$ (speciation rate increase towards present) (Pybus and Harvey, 2000).

To test whether the rate of lineage accumulation was gradual or the result of punctuated episodes such as having periods of evolutionary stasis followed by bursts of speciation, the Delta test was used (Webster et al., 2003). The frequency of such evolutionary processes has been shown to be relatively common amongst plant lineages (Pagel et al., 2006). Bias in the estimation was controlled for by testing for a node density effect which has been shown to confound estimates of punctuational evolution (Venditti et al., 2006). For the analyses, 1000 phylograms from the Bayesian relaxed clock analyses were submitted to the website http://www.evolution.reading.ac.uk/pe/index.html.

To identify possible ancestral branches within *Dodonaea* and their sister clade *Diplopeltis* that had significantly higher than expected rates of lineage accumulation (diversification rate shifts), the relative cladogenesis statistic (P_k) was measured as implemented in the program End-Epi ver. 1.0.1 (Purvis et al., 1995; Rambaut et al., 1997) on the Bayesian maximum clade credibility chronogram scaled

to mean node heights. The test calculates the probability (P_k) under a constant-rates birth death model that a lineage at time t will have k tips compared with the total number of tips by time 0 (the present). To identify diversification rate shifts from the whole tree analysis that may be artefacts of the trickle-down effect (Moore et al., 2004), a series of subtree analyses were carried out, beginning with recent nodes and progressing along all nodes to the root of the tree. Significantly supported nodes ($P < 0.01$) not supported in the subtree analyses were considered as artefacts of the trickle-down effect.

The rate of diversification as a speciation-extinction process ($\hat{r}_e$) was estimated for the seventy species of *Dodonaea* and their sister clade *Diplopeltis* with the same stem age and 95% HPD from previous analyses and the method-of-moments estimator for crown age equation (7) and stem age equation (6) of Magallón and Sanderson (2001) (see equations below). Extinction was modelled at both the zero level (r_0) and also at the relatively very high extinction rate ($r_{0.9}$) following the methodology and justifications of Magallón and Sanderson (2001).

Equations (6) and (7) of Magallón and Sanderson (2001) for estimation of diversification rate for stem and crown age respectively. Where ε is relative extinction rate, *t* is time interval, and *n* is diversity.

$$\hat{r}_\epsilon = \frac{1}{t}\log[n(1-\epsilon)+\epsilon] \qquad (6)$$

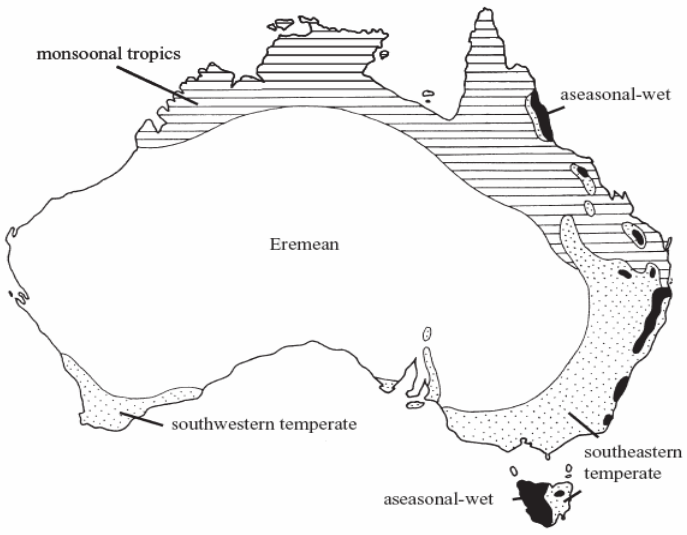
$$\hat{r}_\epsilon = \frac{1}{t}\left\{\log\left[\frac{1}{2}n(1-\epsilon^2)+2\epsilon + \frac{1}{2}(1-\epsilon)\sqrt{n(n\epsilon^2-8\epsilon+2n\epsilon+n)}\right]-\log 2\right\} \qquad (7)$$

Biogeographical analyses

Species distributions were taken from Australia's Virtual Herbarium (http://www.chah.gov.au/avh) and West (1984; 1985). Current distributions of each species were coded into six discrete categories based on the map of Australian biomes from Crisp et al. (2004) (fig. 5.1). These areas are 1) arid - Eremean, 2)

southeastern temperate, 3) southwestern temperate, 4) monsoonal tropics, 5) aseasonal-wet, and 6) extra-Australia. Species distributions per biome were mapped onto the Bayesian posterior probability (PP) 95% majority rule consensus topology of the combined dataset from the PHASE analyses (fig. 5.3). Lineage accumulations of the sister genera *Dodonaea* and *Diplopeltis* for each biome over one million year time intervals were also plotted.

Figure 5.1: Australian biomes from Crisp et al. 2004. Southeastern temperate = sclerophyll (eucalypt) forest, woodland and heath, seasonally dry. Southwestern temperate = sclerophyll (eucalypt) forest, woodland and heath, Mediterranean climate. Monsoonal tropics = savannah, mostly sclerophyll (eucalypt and acacia) seasonally dry. Eremean = arid shrubland, low woodland and grassland. Aseasonal-wet = year-round high rainfall, tropical to temperate or subalpine, closed-canopy rainforest (volcanic soils) to heath (oligotrohic soils).

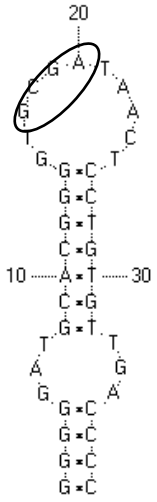


Results

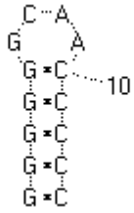
The 85 taxa included in this study comprise 15 outgroup (from *Cossinia*, *Diplopeltis Harpullia* and *Loxodiscus*) and 70 ingroup taxa (*Dodonaea* and *Distichostemon*). The aligned combined dataset plus coded indels is 957 bp and includes 253 bp of ITS1 (66 stem pairs), 224bp of ITS2 (75 stem pairs) and partial 462 bp of ETSf (142 stem pairs). There are compensatory base changes (CBC) in both the ITS and ETSf datasets, with the greatest number between outgroups, or ingroup and outgroups eg. in ITS – six CBC between *Harpullia cupanioides* and *Diplopeltis stuartii*, in ETSf – seven CBC between the sister species, *D. falcata* and *D. peduncularis*, and *Harpullia arborea*.

There are a number of insertion and deletion events in the sequences of some taxa that alter their secondary structures and are phylogenetically informative. The first of these is a significant evolutionary event, since it entails a major change in the secondary structure of ITS1 (fig. 5.2). All species of *Dodonaea*, *Distichostemon* and four of the five species of *Diplopeltis* (*D. eriocarpa*, *D. intermedia*, *D. huegelii* and *D. petiolaris*) have a deletion of up to 29 bp from Helix II of ITS1. This deletion involves the removal of an internal bulge and a second set of stems pairings that are found in all species of *Cossinia*, *Harpullia*, *Loxodiscus* and *Diplopeltis stuartii*. While the biological significance of this loss is unknown, the simplest explanation is that the original stem extension is functionally superfluous. The terminal loop in all of the taxa with the stem and bulge deletion is an example of a tetraloop (in this case generally GCAA or GCGA), which is more thermodynamically stable than other configurations of unpaired bases in single strand structures due to interactions between the bases (Higgs, 2000). All of the 19 indels are described in table 5.1.

Figure 5.2: Representative secondary structures for helix II of ITS1 RNA transcript for: A) all species of *Cossinia*, *Harpullia*, *Loxodiscus* and *Diplopeltis stuartii* with possible tetraloop circled, and B) all species of *Dodonaea*, *Distichostemon* and four of the five species of *Diplopeltis* (*D. eriocarpa*, *D. intermedia*, *D. huegelii* and *D. petiolaris*) with loop (GCAA) an example of a relatively more stable tetraloop. Free energy (ΔG = kcal/mol) for each structure is shown.



A *Harpullia rhyticarpa* ΔG =-7.1



B *Dodonaea viscosa* ΔG =-5.2

Table 5.1: Secondary structure description of all indels for the combined data matrix and the taxa within which they occur, also including alignment position and sequence region. bp = basepair.

Indel	Alignment position	Region	Secondary structure description	Taxa
1	52-79	ITS1	Removal of internal bulge and multiple stem pairings in Helix II	All species of *Dodonaea*, *Distichostemon* and four of the five species of *Diplopeltis* (*D. eriocarpa*, *D. intermedia*, *D. huegelii* and *D. petiolaris*)
2	87	ITS1	Adenine deletion in loop between helices II and III	All species except all samples of *Harpullia*
3	96-98	ITS1	3 bp insertion which may be the result of a replication slippage event that causes a lateral bulge in helix III	*D. biloba* and *D. procumbens*
4	113-119	ITS1	2 bp insertions of various nucleotide characters in terminal loop of helix III plus gain of additional stem pairing adjacent to loop	As in 1
5	197	ITS1	Adenine or guanine insertion as lateral bulge in helix V	All species except all samples of *Cossinia*, *Harpullia* and *Loxodiscus*
6	231	ITS1	Adenine insertion as lateral bulge in helix V	All species of *Distichostemon*
7	272	ITS2	1 bp insertion of all nucleotide characters as lateral bulge of helix I	All species except all samples of *Harpullia*
8	276	ITS2	Uracil or cytosine insertion that generates an additional stem pairing (U:G or C:G) in helix I	All species except all samples of *Harpullia* plus *Loxodiscus*
9	289	ITS2	Deletion of single nucleotide lateral bulge in helix 1	*D. adenophora, D. baueri, D microzyga, D. pachyneura, D. platyptera, D. ridigia* and *D stenophylla*
10	295	ITS2	Guanine deletion from stem pair leaving lateral bulge in helix 1	*D. adenophora, D microzyga* and *D. platyptera*
11	333	ITS2	Deletion of nucleotide from terminal loop stem II	All samples except all species of *Harpullia* and *Loxodiscus*
12	461	ITS2	Deletion of nucleotide from terminal loop stem IV	*D. triangularis* and *D. triquetra*
13	597-8	ETSf	Guanine/guanine or guanine/adenine insertion in lateral bulge helix II	*Diplopeltis* (*D. eriocarpa*, *D. intermedia*, *D. huegelii* and *D. petiolaris*)
14	681	ETSf	Uracil/adenine insertion in terminal loop helix IV	*D. caespitosa, D. ceratocarpa, D. divaricate, D. ericoides, D. hexandra, D. humifusa* and D. *pinifolia*
15	847-51	ETSf	Various nucleotide deletions in loop between helices 6 and 7	*Diplopeltis* (*D. eriocarpa*, *D. intermedia*, *D. huegelii* and *D. petiolaris*)
16	849	ETSf	Guanine insertion in loop between helices 6 and 7	Both species of *Cossinia*
17	889-899	ETSf	Guanine insertion as lateral bulge in helix 7 and loss of two loop	*Dodonaea* (multiple species – see fig. 5.3)

			nucleotides	
18	889-899	ETSf	Guanine insertion as lateral bulge in helix 7 and loss of three loop nucleotides	*Diplopeltis* (*D. eriocarpa*, *D. intermedia*, *D. huegelii* and *D. petiolaris*)
19	909	ETSf	Cytosine deletion from a 3 bp lateral bulge in helix 7	*D. boroniifolia, D. filiformis, D. multijuga, D. pinnata* and *D.rupicola.*

Model selection

The 'best fit' evolutionary model for analyses of both the separate ITS and ETS partitions as determined by the best AIC_c score was REV + I + Γ, which is equivalent to the general time reversal model with variable rates (I) and gamma distribution (Γ). The combined data matrix partitioned into separate loops and stems (ITS loops + ITS stems + ETSf loops + ETSf stems) returned a substantially better AIC_c score than non-partitioned (table 5.2). These results are indicative that models which consider secondary structure interactions more accurately represent the mutational processes that operate at different loci within a RNA encoding sequence. The 'best fit' evolutionary model for combined analyses was REV + RNA7C + REV + RNA7C (table 5.2). The covariation RNA7C model is a simpler version of the RNA7A model, does not allow for double substitutions and treats all nucleotide changes as single substitutions. It has 10 rate parameters and seven frequencies.

Table 5.2: Comparison of AICc scores for various partitioning schemes beginning with the simplest. The best RNA7 model is highlighted in bold. In all instances REV = REV + I + Γ which is equivalent to the general time reversal model with variable rates (I) and gamma distribution (Γ). AICc = AIC + 2k(k +1)/n – k where AIC = -*L* + 2k - 1, with k the number of estimable parameters, *L* the likelihood estimate and n the number of nucleotides positions (combined = 939).

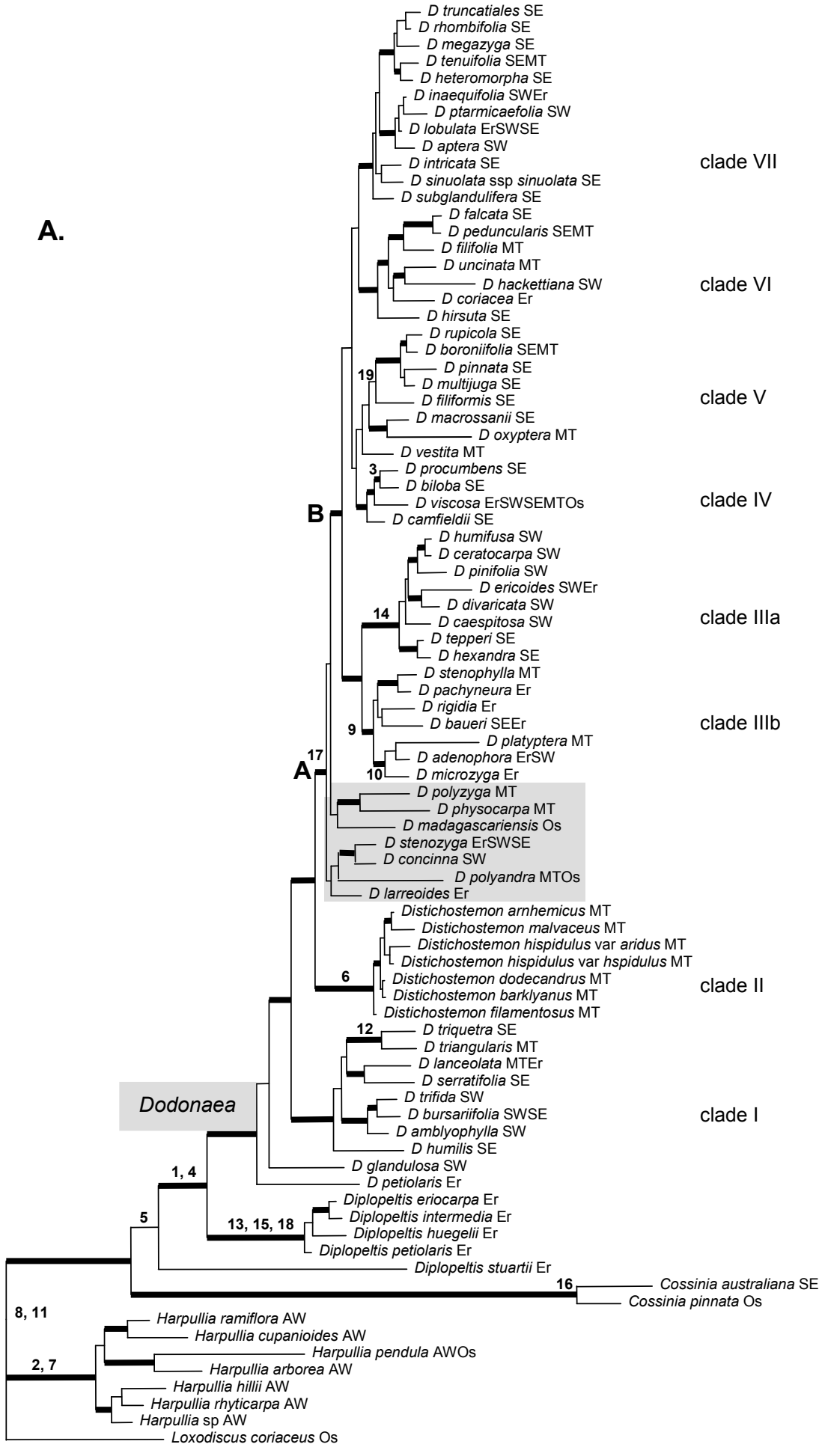
Partition	Models	k	L	AIC_c
ITS + ETSf	REV + REV	22	-9729	19503
ITS loops + ITS stems + ETSf loops + ETSf stems	REV + REV+ REV + REV	44	-9660	19414
ITS loops + ITS stems + ETSf loops + ETSf stems	REV + RNA7A + REV + RNA7A	78	-8597	17354
ITS loops + ITS stems + ETSf loops + ETSf stems	REV + RNA7B + REV + RNA7B	72	-8599	17364
ITS loops + ITS stems + ETSf loops + ETSf stems	**REV + RNA7C + REV + RNA7C**	**56**	**-8603**	**17325**
ITS loops + ITS stems + ETSf loops + ETSf stems	REV + RNA7D + REV + RNA7D	44	-8641	17372

Bayesian MCMC estimation of phylogeny

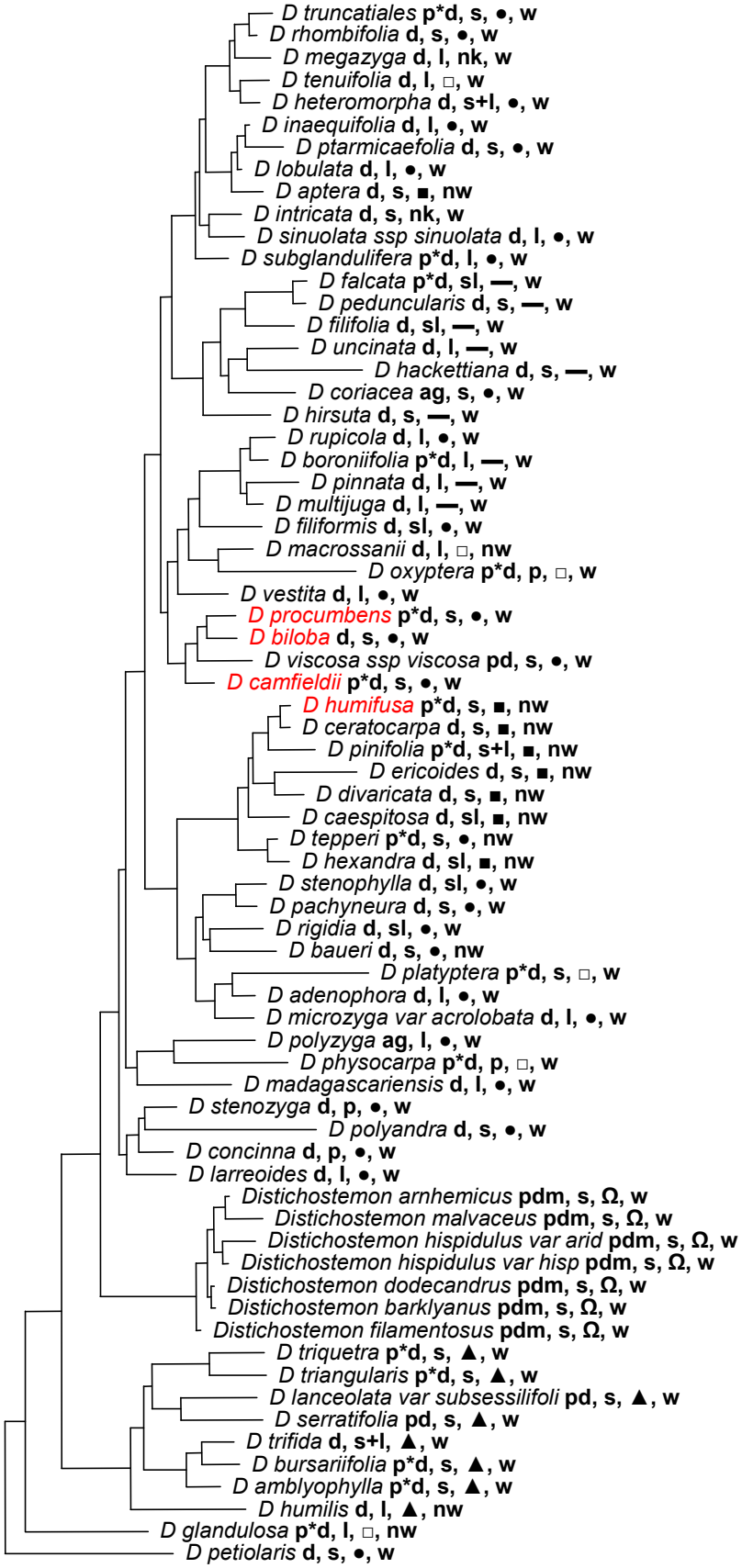
Phylogeny estimation is being driven by the ITS dataset, since the ITS stems had a higher relative mean evolutionary rate than ETSf stems (ETSf/ITS = 0.87), whereas ITS loops had a higher relative mean substitution rate than ETSf loops (ETSf/ITS = 0.73). Mean posterior estimates of the sampled model parameters (including nucleotide frequencies and rate ratios) for combined analyses of ITS loops + ETSf loops + ITS stems + ETSf stems are in appendix 5.2.

The Bayesian posterior probability (PP) 95% majority rule consensus topology of the combined dataset is shown in fig. 5.3 with branch lengths proportional to the mean posterior estimates of the evolutionary distances between the two complementary clades they define. Phylogeny estimations for the two separate data partitions (trees not shown) did not show any substantial conflicting nodes receiving ≥ 95% PP.

Figure 5.3 (next page): **A.** Phylogeny for *Dodonaea* and *Diplopeltis* derived from combined ITS and ETS secondary structure alignment estimated using Bayesian MCMC analyses with the model REV + I + Γ for the separate loop partitions and RNA7C for each of the separate stem pair partitions. Branches with posterior probabilities ≥ 95 % are shown in bold. Branch lengths are the mean posterior estimates of the evolutionary distances between the two complementary clades they define i.e. the longer the branch, the bigger the evolutionary distance between the incident nodes. Numbers on branches refer to indels characterised in table 5.1. Letters A and B refer to nodes discussed in the text. Species distributions relative to Australian biomes described in fig. 5.1 are coded after species name: Er = Eremean, SW = southwestern temperate, SE = southeastern temperate, MT = monsoonal tropics, AW = aseasonal wet, Os = overseas from Australia. **(following page) B**. Same phylogeny for *Dodonaea* only, with breeding system (pd polygamo-dioecious, p*d rarely polygamo-dioecious, d dioecious, ag andromonoecious or gynomonoecious, m monoecious), leaf type (**i** imparipinnate, **l** linear, **p** parapinnate, **s** simple), seed type (see fig. 5.10 for examples, ● black lenticular – f g i, ▲ hyaline membrane lifting at margin only – a b c, □ hyaline membrane lifting over whole seed – no example, ■ infundibular aril, ▬ lenticular compressed towards margin, **Ω** funiculus with raised annular rim around the hilum - h, **nk** not known) and capsule appendage (**w** winged, **nw** not winged or otherwise angled or horned) noted for each species. Species in red are prostrate with branches rooting at nodes.


A.
D truncatiales SE
D rhombifolia SE
D megazyga SE
D tenuifolia SEMT
D heteromorpha SE
D inaequifolia SWEr
D ptarmicaefolia SW
D lobulata ErSWSE
D aptera SW
D intricata SE
D sinuolata ssp sinuolata SE
D subglandulifera SE
clade VII
D falcata SE
D peduncularis SEMT
D filifolia MT
D uncinata MT
D hackettiana SW
D coriacea Er
D hirsuta SE
clade VI
D rupicola SE
D boroniifolia SEMT
19
D pinnata SE
D multijuga SE
D filiformis SE
D macrossanii SE
D oxyptera MT
D vestita MT
clade V
3
D procumbens SE
D biloba SE
D viscosa ErSWSEMTOs
D camfieldii SE
clade IV
B
D humifusa SW
D ceratocarpa SW
D pinifolia SW
D ericoides SWEr
D divaricata SW
14
D caespitosa SW
D tepperi SE
D hexandra SE
clade IIIa
D stenophylla MT
D pachyneura Er
D rigidia Er
9
D baueri SEEr
D platyptera MT
D adenophora ErSW
clade IIIb
17
A
10
D microzyga Er
D polyzyga MT
D physocarpa MT
D madagascariensis Os
D stenozyga ErSWSE
D concinna SW
D polyandra MTOs
D larreoides Er
Distichostemon arnhemicus MT
Distichostemon malvaceus MT
Distichostemon hispidulus var aridus MT
Distichostemon hispidulus var hspidulus MT
6
Distichostemon dodecandrus MT
Distichostemon barklyanus MT
Distichostemon filamentosus MT
clade II
12
D triquetra SE
D triangularis MT
D lanceolata MTEr
D serratifolia SE
Dodonaea
D trifida SW
D bursariifolia SWSE
D amblyophylla SW
D humilis SE
clade I
D glandulosa SW
1, 4
D petiolaris Er
Diplopeltis eriocarpa Er
Diplopeltis intermedia Er
5
13, 15, 18
Diplopeltis huegelii Er
Diplopeltis petiolaris Er
Diplopeltis stuartii Er
16
Cossinia australiana SE
Cossinia pinnata Os
8, 11
Harpullia ramiflora AW
Harpullia cupanioides AW
Harpullia pendula AWOs
2, 7
Harpullia arborea AW
Harpullia hillii AW
Harpullia rhyticarpa AW
Harpullia sp AW
Loxodiscus coriaceus Os

B.


D truncatiales p*d, s, ●, w
D rhombifolia d, s, ●, w
D megazyga d, l, nk, w
D tenuifolia d, l, □, w
D heteromorpha d, s+l, ●, w
D inaequifolia d, l, ●, w
D ptarmicaefolia d, s, ●, w
D lobulata d, l, ●, w
D aptera d, s, ■, nw
D intricata d, s, nk, w
D sinuolata ssp sinuolata d, l, ●, w
D subglandulifera p*d, l, ●, w
D falcata p*d, sl, —, w
D peduncularis d, s, —, w
D filifolia d, sl, —, w
D uncinata d, l, —, w
D hackettiana d, s, —, w
D coriacea ag, s, ●, w
D hirsuta d, s, —, w
D rupicola d, l, ●, w
D boroniifolia p*d, l, —, w
D pinnata d, l, —, w
D multijuga d, l, —, w
D filiformis d, sl, ●, w
D macrossanii d, l, □, nw
D oxyptera p*d, p, □, w
D vestita d, l, ●, w
D procumbens p*d, s, ●, w
D biloba d, s, ●, w
D viscosa ssp viscosa pd, s, ●, w
D camfieldii p*d, s, ●, w
D humifusa p*d, s, ■, nw
D ceratocarpa d, s, ■, nw
D pinifolia p*d, s+l, ■, nw
D ericoides d, s, ■, nw
D divaricata d, s, ■, nw
D caespitosa d, sl, ■, nw
D tepperi p*d, s, ●, nw
D hexandra d, sl, ■, nw
D stenophylla d, sl, ●, w
D pachyneura d, s, ●, w
D rigidia d, sl, ●, w
D baueri d, s, ●, nw
D platyptera p*d, s, □, w
D adenophora d, l, ●, w
D microzyga var acrolobata d, l, ●, w
D polyzyga ag, l, ●, w
D physocarpa p*d, p, □, w
D madagascariensis d, l, ●, w
D stenozyga d, p, ●, w
D polyandra d, s, ●, w
D concinna d, p, ●, w
D larreoides d, l, ●, w
Distichostemon arnhemicus pdm, s, Ω, w
Distichostemon malvaceus pdm, s, Ω, w
Distichostemon hispidulus var arid pdm, s, Ω, w
Distichostemon hispidulus var hisp pdm, s, Ω, w
Distichostemon dodecandrus pdm, s, Ω, w
Distichostemon barklyanus pdm, s, Ω, w
Distichostemon filamentosus pdm, s, Ω, w
D triquetra p*d, s, ▲, w
D triangularis p*d, s, ▲, w
D lanceolata var subsessilifoli pd, s, ▲, w
D serratifolia pd, s, ▲, w
D trifida d, s+l, ▲, w
D bursariifolia p*d, s, ▲, w
D amblyophylla p*d, s, ▲, w
D humilis d, l, ▲, nw
D glandulosa p*d, l, □, nw
D petiolaris d, s, ●, w

The two species of *Cossinia* are strongly supported (100% PP) sister to *Diplopeltis* and *Dodonaea*. However *Diplopeltis* is paraphyletic, with *Diplopeltis stuartii* not grouping with the strongly supported clade (100% PP) comprising all of the other species *Diplopeltis*. A visual inspection of the tree shape within *Dodonaea* reveals a series of splits between a smaller sister clade and the remaining species. The monophyly of *Dodonaea* (including all *Distichostemon*) is well supported (PP 100%). The relationships between *D. petiolaris* and *D. glandulosa*, and all the other species in the remaining strongly-supported clade (100% PP) are only weakly resolved. Within this large clade, eight species (clade I - 100% PP) are sister to remaining *Dodonaea*. *Distichostemon* (clade II - 100% PP) is the next to diverge. Within the remaining 53 species of *Dodonaea* (node A - 100% PP) the phylogeny in many places is poorly supported (≤ 95% PP). There are, however, a number of internal nodes in this clade receiving good support (≥ 95% PP).

Evolutionary rates analyses

The Bayesian uncorrelated log normal estimation of phylogeny and substitution rates indicated a moderate degree of rate heterogeneity amongst branches with the 95% CI coefficient of variation significantly removed from zero (mean 0.6, HPD 0.46-0.74), justifying the use of a relaxed molecular clock analyses. There was no substitution rate autocorrelation between related branches in the phylogeny since the 95% CI of the covariance statistic was not significantly removed from zero (mean 0.03, HPD -0.1-0.17). The mean substitution rate per site across the ratogram was 0.118 substitutions per site (95% HPD 0.083-0.16). The effective sample sizes (number of independent samples from the marginal posterior or prior distributions) for all designated nodes of interest in all analyses were >200, indicating that the MCMC chains were run for adequate lengths.

The ratogram is shown in fig. 5.4 (appendix 5.3 – 95% HPD interval). The ratogram with branch lengths extrapolated from the mean substitution rate per site to the dimension of time based on the stem node of *Dodonaea* (mean 10.4 Mya, 95% HPD 6 -15.4 MYA) from the analyses of Harrington et al. in review is shown in fig. 5.5. Extrapolation of the substitution rates (mean 0.118, 95% HPD 0.083-0.16) to time using the same mean date (10.4 Mya) is shown in table 5.3.

Rate of extant lineage accumulation

The LTT plot is shown in fig. 5.6. There are a number of regions of the plot that show small movements away from a straight line. The γ-statistic for *Dodonaea* plus

four species of *Diplopeltis* was -2.17 (SD 0.5), and for *Dodonaea* only was -2.7 (SD 0.6), both values indicating significant deviation from a homogenous speciation process, with the negative value signifying that there are more nodes towards the root of *Dodonaea* and consequently there has been a significant slowing of the speciation process towards recent time.

The Node-density Artefact Analyser server returned that there is no significant evidence for a relationship between nodes and pathlengths ($\beta = 0$, no punctuated evolution) or the node density artefact. The relative cladogenesis statistic for the whole tree analysis indicated significant diversification rate shifts ($P < 0.01$) that generally correlate with nodes that contribute to the unequal ladderised phylogeny. Further subtree analyses indicated that actual significant diversification rate shift was on the node within *Dodonaea* that included all the species in clades I-VI (fig. 5.5).

Figure 5.4: (next page) Ratogram from the Bayesian Markov chain Monte Carlo phylogeny and molecular evolutionary rates estimations of the combined aligned matrix without indels performed using the BEAST package (Drummond and Rambaut, 2007) with the mean substitution rate across the phylogeny equal to one ie. mean root to tip node height equal to one. The 95% HPD interval for each node receiving >50% PP is shown in appendix 5.3. For each ITS and ETSf stem and loop partition the general-time-reversible (GTR) model with gamma distributed rate variation and an assumed proportion of invariable sites were applied. Black lines indicate branches receiving <50% posterior probability. Branches are coloured by rate from orange-red fastest substitution rate to darkest blue slowest substitution rate. For example the fastest rate *Harpullia penda* 2.6 substitutions per site, with the slowest terminal *D. rhombifolia* 0.56 substitutions per site.

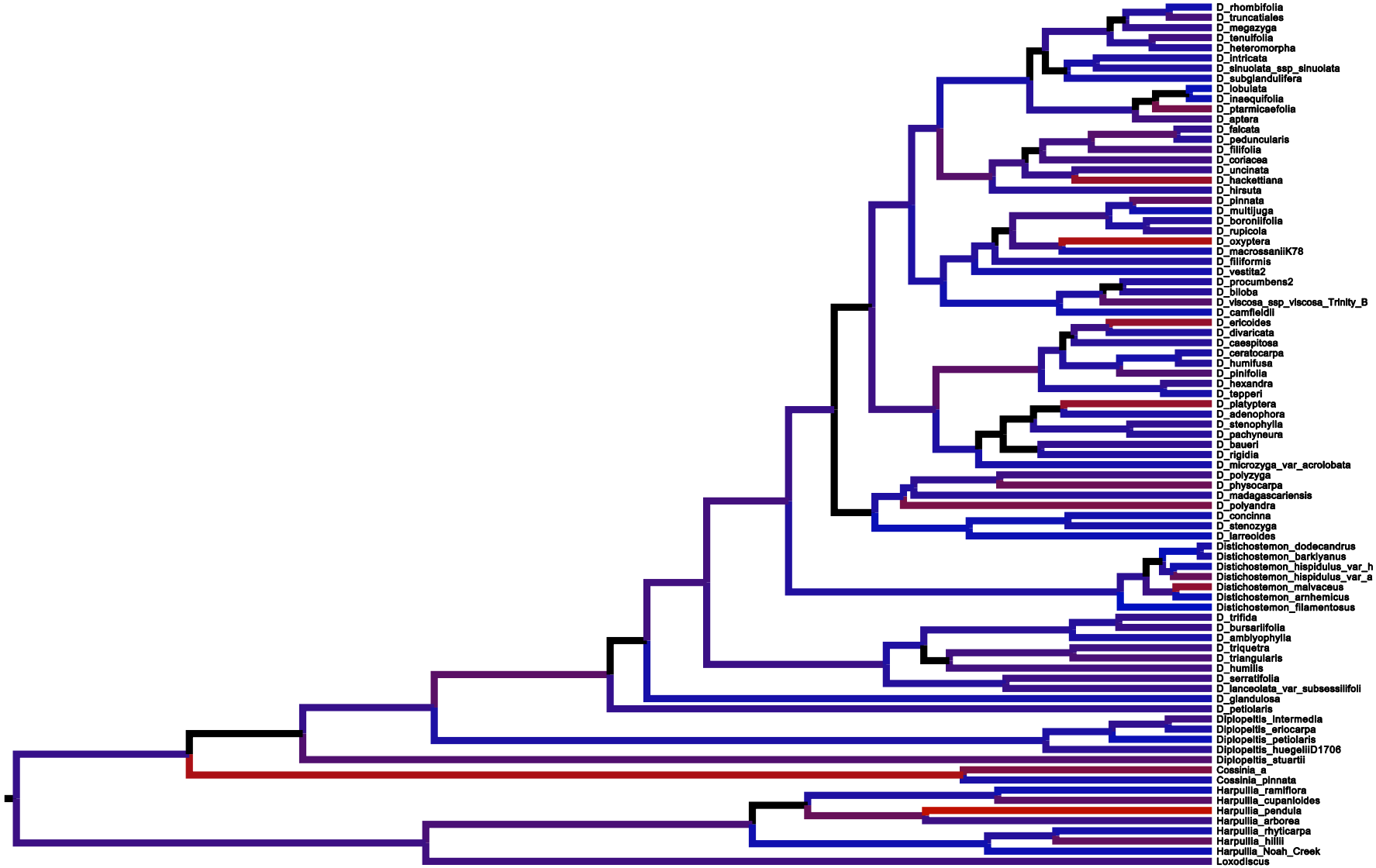
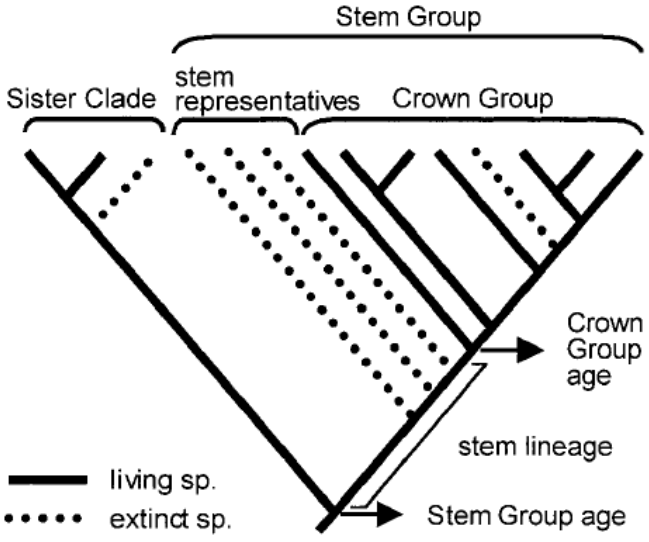

D_rhombifolia
D_truncatiales
D_megazyga
D_tenuifolia
D_heteromorpha
D_intricata
D_sinuolata_ssp_sinuolata
D_subglandulifera
D_lobulata
D_inaequifolia
D_ptarmicaefolia
D_aptera
D_falcata
D_peduncularis
D_filifolia
D_coriacea
D_uncinata
D_hackettiana
D_hirsuta
D_pinnata
D_multijuga
D_boronifolia
D_rupicola
D_oxyptera
D_macrossaniiK78
D_filiformis
D_vestita2
D_procumbens2
D_biloba
D_viscosa_ssp_viscosa_Trinity_B
D_camfieldii
D_ericoides
D_divaricata
D_caespitosa
D_ceratocarpa
D_humifusa
D_pinifolia
D_hexandra
D_tepperi
D_platyptera
D_adenophora
D_stenophylla
D_pachyneura
D_baueri
D_rigidia
D_microzyga_var_acrolobata
D_polyzyga
D_physocarpa
D_madagascariensis
D_polyandra
D_concinna
D_stenozyga
D_larreoides
Distichostemon_dodecandrus
Distichostemon_barklyanus
Distichostemon_hispidulus_var_h
Distichostemon_hispidulus_var_a
Distichostemon_malvaceus
Distichostemon_arnhemicus
Distichostemon_filamentosus
D_trifida
D_bursariifolia
D_amblyophylla
D_triquetra
D_triangularis
D_humilis
D_serratifolia
D_lanceolata_var_subsessilifoli
D_glandulosa
D_petiolaris
Diplopeltis_intermedia
Diplopeltis_eriocarpa
Diplopeltis_petiolaris
Diplopeltis_huegeliiD1706
Diplopeltis_stuartii
Cossinia_a
Cossinia_pinnata
Harpullia_ramiflora
Harpullia_cupanioides
Harpullia_pendula
Harpullia_arborea
Harpullia_rhyticarpa
Harpullia_hillii
Harpullia_Noah_Creek
Loxodiscus

Figure 5.5: (next page) Chronogram from the Bayesian Markov chain Monte Carlo phylogeny and molecular evolutionary rates estimations of the combined aligned matrix without indels performed using the BEAST package (Drummond and Rambaut, 2007). For each ITS and ETSf stem and loop partition the general-time-reversible (GTR) model with gamma distributed rate variation and an assumed proportion of invariable sites were applied. Mean branch lengths (substitutions per site) have been scaled into units of time using the mean age for the split between *Dodonaea* and *Diplopeltis* of 10.4 Mya from the analyses of Harrington et al. (in review). The Pliocene epoch has been shaded gray. Black circle designate where there has been a significant ($P < 0.01$) diversification rate shift identified by the relative cladogenesis statistic (P_k) (Rambaut et al., 1997). Solid vertical lines refer to clades at either end of the line that have disjunct distributions on either side of the Nullarbor Plain and are discussed in the text. In all instances the clade from the southeastern temperate biome is at the top of the line, and clade from southwestern temperate biome at the bottom of the line.

From Magallon and Sanderson (2001). Definition of stem group and crown group age. The age of the stem group is the time of divergence of the clade from its sister taxon. The age of the crown group is the time of the deepest bifurcation within the crown group. Solid lines represent living species; dotted lines represent extinct species.

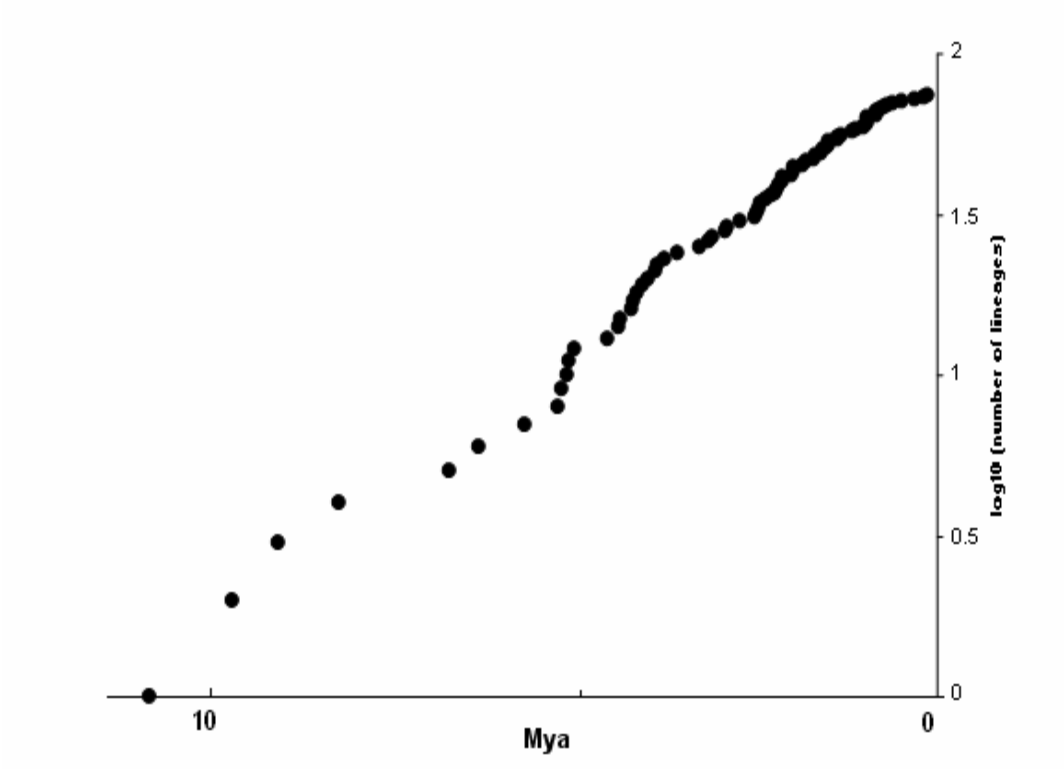



D rhombifolia
D truncatiales
D megazyga
D tenuifolia
D heteromorpha
D intricata
D sinuolata
D subglandulifera
D lobulata
D inaequifolia
D ptarmicaefolia
D aptera
D falcata
D peduncularis
D filifolia
D coriacea
D uncinata
D hackettiana
D hirsuta
D pinnata
D multijuga
D boroniifolia
D rupicola
D oxyptera
D macrossanii
D filiformis
D vestita2
D procumbens
D biloba
D viscosa
D camfieldii
D ericoides
D divaricata
D caespitosa
D ceratocarpa
D humifusa
D pinifolia
D hexandra
D tepperi
D platyptera
D adenophora
D stenophylla
D pachyneura
D baueri
D rigidia
D microzygaa
D polyzyga
D physocarpa
D madagascariensis
D polyandra
D concinna
D stenozyga
D larreoides
Dist dodecandrus
Dist barklyanus
Dist hispidulus var h
Dist hispidulus var a
Dist malvaceus
Dist arnhemicus
Dist filamentosus
D triquetra
D triangularis
D humilis
D trifida
D bursariifolia
D amblyophylla
D serratifolia
D lanceolatai
D glandulosa
D petiolaris
Diplopeltis intermedia
Diplopeltis eriocarpa
Diplopeltis petiolaris
Diplopeltis huegelii
Diplopeltis stuartii
Cossinia austral
Cossinia pinnata
Harpullia ramiflora
Harpullia cupanioides
Harpullia pendula
Harpullia arborea
Harpullia rhyticarpa
Harpullia hillii
Harpullia sp.k
Loxodiscus
10
Mya

Table 5.3: Extrapolation of the substitution rates (substitutions per site) from the Bayesian uncorrelated log normal estimation of phylogeny and substitution rates for the combined dataset to units per time (substitutions per site per million years) using the mean date for the stem node of *Dodonaea* (mean 10.4 Mya, 95% HPD 6 -15.4 Mya) returned from the analyses of Harrington et al (in review).

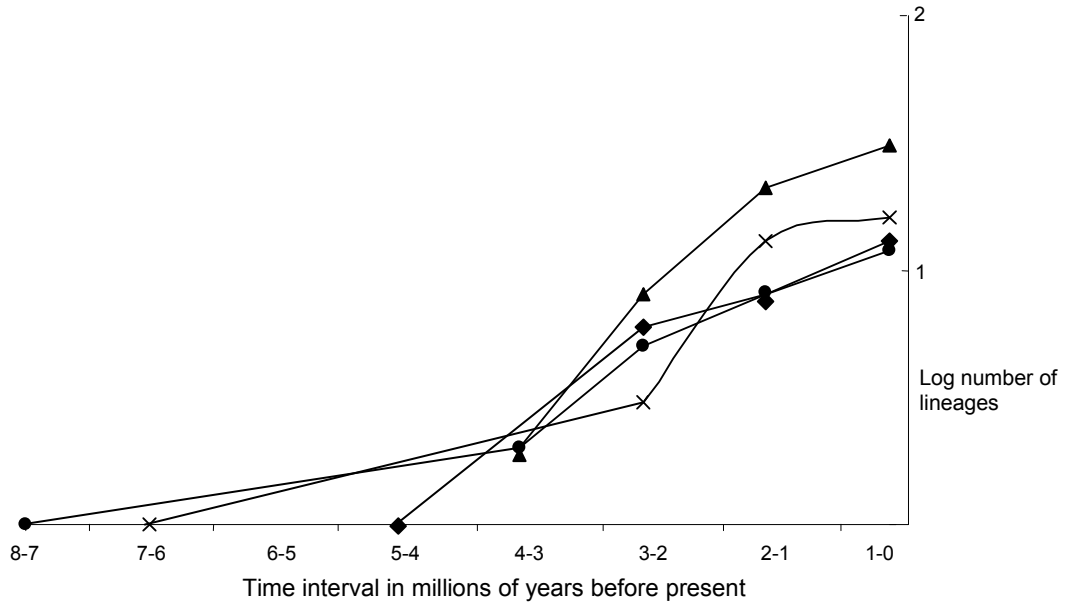
	mean	95% HPD lower	95% HPD upper
Node height for *Dodonaea* split from *Diplopeltis* (sub/site)	0.088	0.07	0.11
Extrapolation from mean (10.4 Mya)	10.4	8.0	13.0
Extrapolation substitutions per site per million years (x 10^{-9}).	8.5	6.7	10

Figure 5.6: Lineage through time plot for *Dodonaea*. Log number of extant lineages against node height scaled to 10.36 Mya for the split of *Dodonaea* and *Diplopeltis* plotted for the maximum clade credibility chronogram scaled to mean node heights and with all outgroups removed.



A lineage accumulation plot based on extant species current distributions in Australian biomes (fig. 5.1) for combined *Dodonaea* and *Diplopeltis* over one million year time intervals is shown in fig. 5.7.

Figure 5.7: Lineage accumulation over time of extant *Dodonaea* and *Diplopeltis* in Australian biomes over one million year time intervals. ● = eremean, x = southwestern temperate, ▲ = southeastern temperate, ♦ = monsoon tropics.

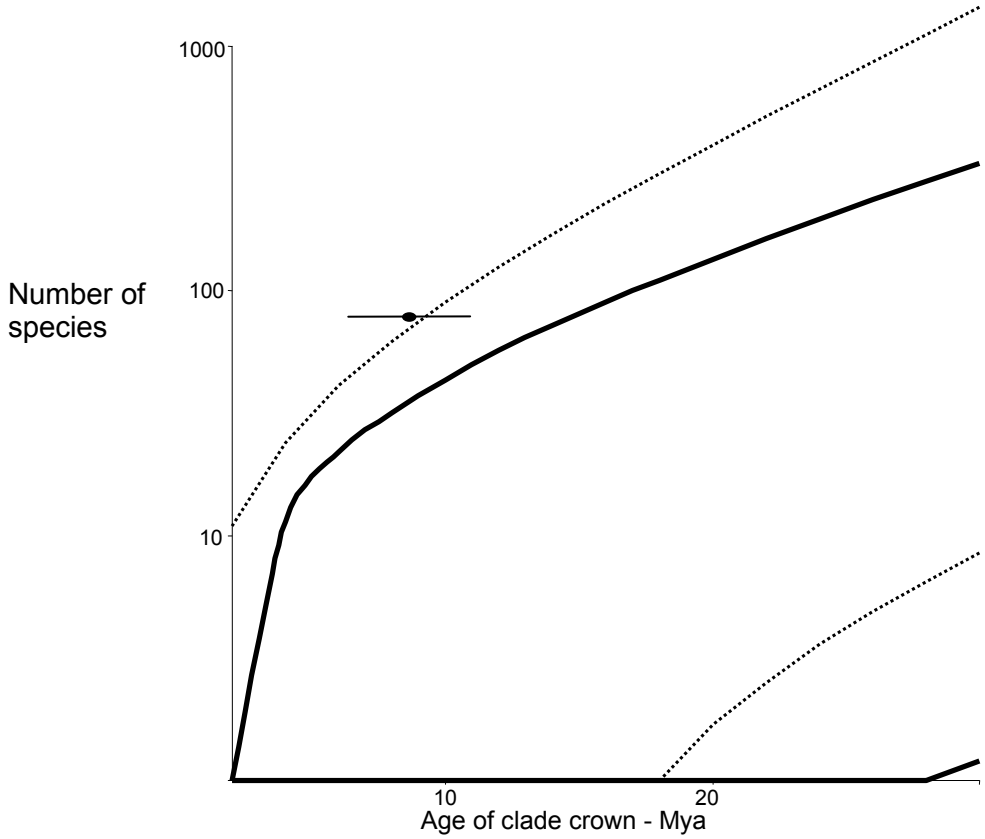


Diversification rate estimates at both high and low extinction rate estimates for stem and crown group age for *Diplopeltis* and *Dodonaea* are given in table 5.4. Diversification rate of crown *Dodonaea* relative to rate of diversification of angiosperms as a whole (Magallon and Sanderson, 2001) is shown in fig. 5.8.

Table 5.4: Net rate of diversification per million years for *Dodonaea* and *Diplopeltis* in the absence of extinction (ε = 0.0) and at relatively high extinction rate (ε = 0.9) for stem and crown groups. Stem and crown ages are derived from an extrapolation of relative substitution rates mean and 95% HPD from the Bayesian uncorrelated log normal estimation of phylogeny and substitution rates for the combined dataset to units per time (substitutions per site per million years) using the mean date for the stem node of *Dodonaea* (10.4 Mya) returned from the analyses of Harrington et al (in review)

	ε = 0.0 mean	95% HPD upper	95% HPD lower	ε = 0.9 mean	95% HPD upper	95% HPD lower
Dodonaea/*Diplopeltis* stem	0.41	0.54	0.33	0.22	0.26	0.16
Dodonaea only stem	0.35	0.28	0.45	0.21	0.17	0.28
Diplopeltis only stem	0.1	0.08	0.13	0.08	0.06	0.1
Dodonaea crown	0.49	0.57	0.37	0.28	0.35	0.22
Diplopeltis crown	0.5	1.57	0.28	0.38	1.2	0.2

Figure 5.8: Confidence intervals of expected species diversity according to age of crown group. The 95% confidence interval of expected species diversity through time of a clade that diversifies with a rate equal to that of the mean of angiosperms under a high relative extinction rate (ε = 0.9; $r_{0.9}$ = 0.0767) is shown as solid lines. The rate of diversification corresponding to that of the earliest radiation within the angiosperms under a high relative extinction rate (ε = 0.9; $r_{0.9}$ = 0.123) according to a starting crown group age is shown as dashed lines. *Dodonaea* is mapped according to crown group age (mean with 95% HPD) and standing species diversity (70). A clade that falls above the upper limit of the highest confidence interval (i.e., ε = 0.9; $r_{0.9}$ = 0.123) is considered extremely species rich (Magallon and Sanderson, 2001).



Discussion

Delimitation of Diplopeltis

The endemic Australian genus *Diplopetis* (pepperflowers) was erected by Endlicher (1837) based on the type species *Diplopeltis huegelii*. Mueller (1863b) added a further two species, *Diplopeltis petiolaris*, *Diplopeltis stuartii*, and a variety (which later became *Diplopeltis eriocarpa*) to the genus. Mueller (1863a) further suggested that *Diplopeltis stuartii* might be considered a separate genus based on the

distinctive zygomorphic flowers and the dehiscent fruit not found in other members of the genus, but instead he created a new monotypic section (Diplopholis) within *Diplopeltis* to accommodate this species. The last species (*Diplopeltis intermedia*) was added by George and Erdtman (1969) when they revised the genus after a palynological and morphological re-examination. Erdtman, who undertook the pollen analysis, was intrigued by the two distinct pollen types within the genus: the small prolate grains of *Diplopeltis stuartii* that were elliptical in equatorial view, as opposed to the larger perprolate, rounded grains of the other species that were rhomboidal in equatorial. He commented "so different are the pollen grains in the two sections that had they been found as dispersed fossil spores, they would probably have been referred to different sporomorph genera" (p. 93, George and Erdtman, 1969).

Perprolate - Describing the shape of a pollen grain or spore in which the ratio between the polar axis and the equatorial diameter is more than 2.
Prolate - Describing the shape of a pollen grain or spore in which the polar axis is larger than the equatorial diameter.
From - Erdtman, G. 1943. An Introduction to Pollen Analysis. Waltham, Massachusetts.

In this molecular investigation that incorporates all species of *Diplopeltis*, *Diplopeltis* is paraphyletic (fig. 5.3). *Diplopeltis stuartii* is not closer to other species of the genus than it is to species outside the genus. There are also a number of evolutionary elements in the molecular data that support *Diplopeltis stuartii* as distinct from the other members of the genus. *Diplopeltis stuartii* is the only member of the genus not to have lost the stem and loop extension (figs. 5.2 and 5.9) from ITS1 helix II, which is also maintained in all other outgroup taxa. There is a further deletion (indel 2 – table 5.1) and an insertion (indel 4) from all other species of *Diplopeltis* and *Dodonaea* that support the monophyly of all these taxa. There are also three further mutations to sequences of all other species of *Diplopeltis* that support their monophyletic status (indel 13, 15 and 18) and do not occur in *Diplopeltis stuartii*. There is also a CBC in the stem of helix V ITS1 (alignment positions 202/224) that is unique to *Diplopeltis stuartii* (all *Harpullia*, *Cossinia*, and clade I have C-G, the remaining *Dodonaea* and *Diplopeltis* have the intermediate T-G, while *Diplopeltis stuartii* has T-A). To preserve the monophyly of *Diplopeltis* sensu stricto, *Diplopeltis stuartii* needs to be transferred to a new genus.

The four species of *Diplopeltis* (*D. eriocarpa*, *D. huegelii*, *D. intermedia* and *D. petiolaris*) are monoecious low shrubs that generally have a continuous distribution within an 80 kilometres margin along the west coast of Australia between latitudes 20° and 33° south. They have five sepals and four (sometimes five in *D. eriocarpa*) equal-sized pink or white petals, and produce fruit of three cocci that fall to the ground entire. By contrast the herbaceous perennial *Diplopeltis stuartii* has a disjunct distribution from the other four species, being found in the desert regions from the Hamersley Range eastwards almost to the Northern Territory-Queensland border. It has distinctive zygomorphic flowers (four petals, unequal but paired), and produces 3-celled capsules that dehisce loculicidally. Hence, there is a clear distinction among the species of *Diplopeltis* sens. str. in both distribution and reproductive morphology.

Figure 5.9: Nucleotide alignment positions 47-80 from helix II ITS1 for selected taxa. All other species of *Dodonaea* have a similar alignment to *D. viscosa* ssp. *viscosa* Trinity Beach.

D viscosa ssp viscosa Trinity B	GGGGG----------GCAA--------------CCCCC
Distichostemon arnhemicus	GGGGG----------GCAA--------------CCCCC
Distichostemon barklyanus	GGGGG----------GCAA--------------CCCCC
Distichostemon dodecandrus	GGGGG----------GCAA--------------CCCCC
Distichostemon filamentosus	GGGGG----------GCAA--------------CCCCC
Distichostemon hispidulus var h	GGGGG----------GCAA--------------CCCCC
Distichostemon hispidulus var a	GGGGG----------GCAA--------------CCCCC
Distichostemon malvaceus	GGGGG----------GCAA--------------CCCCC
Diplopeltis intermedia	GGGGG----------GCAA--------------CCCCC
Diplopeltis eriocarpa	GGGGG----------GCAA--------------CCCCC
Diplopeltis huegeliiD1706	GGGGG----------GCAA--------------CCCCC
Diplopeltis petiolaris	GGGGG----------GCAA--------------CCCCC
Diplopeltis stuartii	GGGGGAAGCATGAG--GCGAAAGCCCTTGTGACAACCCC
Harpullia arborea	GGGGGATGCACGGGGTGCGATAACTCCTGTGTCAACCCC
Harpullia rhyticarpa	GGGGGATGCACGGGGTGCGATAACTCCTGTGTTGACCCC
Harpullia cupanioides	GGGGGATGCACGGGGTGCGATAACTCCTGTGTCAACCCC
Harpullia Noah Creek	GGGGGATGCACGGGGTGCGATAGCTCCTGTGTCGACCCC
Harpullia hillii	GGGGGATGCACGGGGTGCGATAGCTCCTGTGTTGAACCC
Harpullia pendula	GGGGGGTGCACGGGGTGCGAGAGCTCCTGTGTCGGCCCC
Loxodiscus	GGGGGGTGCACGGGGTGAGA--GCTCCGGTGACATCCCC
Cossinia p	GGGGGGGAGCGCGGGTGCAA--GCTACCGCGACACCCTC

Phylogenetic reconstruction within Dodonaea

The phylogeny of *Dodonaea* presented here (fig. 5.3) is an unbalanced ladderised tree with a series of generally species poor sister lineages (which are by definition equal in age to the species rich lineage) at every node, which can confound interpretations of the direction of ancestral trait evolution within the genus (Crisp and Cook, 2005). The generally long branches leading to clades in the lower half of the phylogeny (for example, clade I and II – fig. 5.3) and also to the single species

lineages (*D. petiolaris* and *D. glandulosa*) are a potential indication for extinction of evolutionary lineages along these branches, and resulting misinterpretation of the potential derived morphologies (apomorphies) of the extant species.

The results of this molecular investigation into phylogenetic relationships within *Dodonaea* show some disparity with previous hypotheses of morphological evolution within *Dodonaea* which have grouped taxa primarily on a combination of leaf, capsule and seed characters, and postulated evolutionary trends in these characters. While some portions of the species groups from the most recent infrageneric classification of West (1984) are supported, generally the group classification needs substantial revision. As in previous morphological investigations, species with compound-leaves occur in several clades, generally interdispersed among species with simple leaves (for example, *D. humilis* is the only species in clade I with imparipinnate leaves). Variation of breeding system is equally diffuse across the phylogeny, and while a majority of species are dioecious there does appear to be some departure from this condition; however, this requires further study, since the is unclear in some taxa. For example, the generally regarded dioecious species, *D. polyandra*, has been observed in garden grown specimens to produce a few sets of female flowers on a staminate shrub (M. Harrington pers. obs.), while apomixis has also been reported for two other isolated dioecious species, *D. hexandra* and *D. microzyga* (Mueller, 1862; Keighery, 1982), and also for *D. viscosa* (West, 1980). Although a majority of genera in Sapindaceae are dioecious, those most closely related to *Dodonaea* in the phylogeny presented here (*Diplopeltis*, *Diplopeltis stuartii*, and *Cossinia*) are monoecious. It has also been noted that while the normal breeding condition in *Harpullia* is dioecism a couple of species have also been noted to be monoecious (Leenhouts and Vente, 1982), or potentially seasonally variable or polygamous (Mark Harrington pers. obs.). It appears that there is a general trend in breeding system evolution across the phylogeny towards strict dioecy, the polygamous condition potentially being intermediate or alternately a partial reversal.

The monophyly of *Dodonaea* as redefined here to include all species of *Distichostemon* is unequivocally supported by the molecular data (100% PP) and the morphological synapomorphies of petalless flowers with a highly reduced intrastaminal disk which is absent in staminate flowers. There do not appear to be any obvious evolutionary trends in the morphological characters (leaf and capsule

form, presence or absence of aril, or breeding system fig. 5.3B) that have been previously used to group taxa. However there are some morphological characters that may be useful to delineate some of the clades recovered in this molecular study. The presence or absence of an aril has been used as an associated character to define species groups by both West (1984) and Radlkofer (1933), while there has also been some discussion as to the presence of a small funicular outgrowth in some exarillate species examined by Corner (1976). I have examined representative species (25 species) from all the main clades of *Dodonaea* and also two species of *Diplopeltis* and have recognised similar small funicular arils in all samples (fig. 5.10). It has been noted that the seeds of *D. viscosa* ssp. *angustissima*, with their very small funicular aril, are prolifically harvested by *Pheidole* sp. of ants and deposited in middens outside the nest after the elaiosome has been consumed (Harrington and Driver, 1995).

The sister to all other ingroup taxa, *D. petiolaris*, is widely distributed throughout the arid zones of central Australia and has a generally unspecialised morphology compared to other *Dodonaea* (simple leaves, dioecious), except for its 3-winged greatly-inflated capsules and spherical seeds with raised funiculus (fig. 5.10).

While the exact sister relationships of *D. glandulosa* are yet to be determined, it is the only species of the genus besides *D. humilis* (clade I) to have glandular hairs on their capsules and a single glandular hair at the apex of each anther sac and is unique in having glandular hairs on its leaves, stamens on distinctively long filiaments, and distinctive two-valved capsules.

Clade I

An example of the difference between the historical morphological groups and the molecular phylogeny is found in Clade I. It contains four species from Radlkofer (1933) Series I, Cyclopterae, subseries I (West 1984 – Group 1) with an additional four species (only three known to Radlkofer) from Radlkofer (1933) Series III Aphanopterae subseries 6 (West 1984 – Group 5b and 5c). These eight species are distinguished from all other *Dodonaea* by the shiny lenticular black seeds that have a hyaline layer lifting generally only across the margin of the seed (fig. 5.10 a, b, c) and leaves lacking glands.

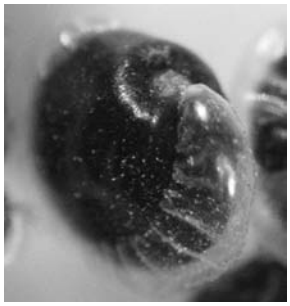
Clade II - Distichostemon

A genus of six species of generally low shrubs, endemic to northern tropical Australia, *Distichostemon* (false hopbushes) has long been recognized as closely related to *Dodonaea*. In his original description, Mueller (1857) differentiated *Distichostemon* from *Dodonaea* on stamen number and the disjunct distribution for the newly described species from the then known species of *Dodonaea*. Bentham and Hooker (1862b) substantiated the close relationship by placing the two genera in series in the Suborder Dodonaeae. Similarly, in his mammoth treatment of Sapindaceae, Radlkofer (1933) placed *Distichostemon* next to *Dodonaea* in the tribe Dodonaeeae. It was distinguished from *Dodonaea* primarily based on its hairy inflated capsules and the number of stamens, which for the specimens that Radlkofer had at his disposal were usually 20 or more, in two or more rows.

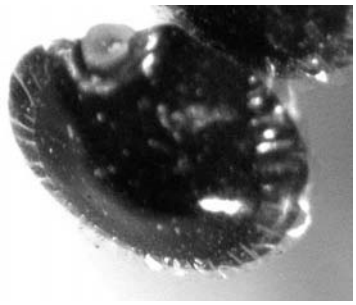
Mueller and Leenhouts (1976) identified similarities in pollen morphology between *Dodonaea* and *Distichostemon*, and noted that the infratectal layer is granular/columellate in both genera, which may relate to wind pollination. In this study all species of *Distichostemon* form a strongly supported clade (fig. 5.3 - clade II) derived within *Dodonaea*, and supported by a synapomorhphic single nucleotide insertion (indel 6 – table 5.1). They have densely-hairy simple leaves, flowers with up to 80 stamens, generally in two rows, and a distinctive lenticular seed that has a conspicuous funiculus with raised annular rim around the hilum (fig. 5.10 h). While plants of *Distichostemon* may be hermaphrodite or dioecious they are the only species of *Dodonaea* sens. lat. in which monoecism has also been recorded.

Figure 5.10: Seed types for *Dodonaea* discussed in text and selection of outgroups.

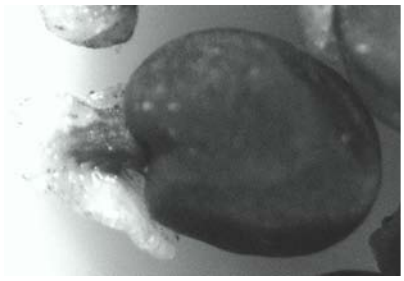
a) *Dodonaea lanceolata*



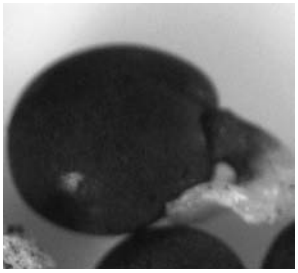
b) *Dodonaea bursariifolia*



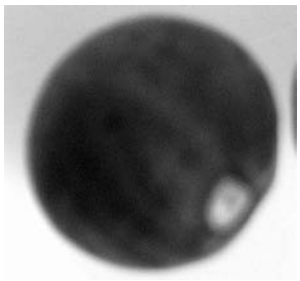
c) *Dodonaea triquetra*



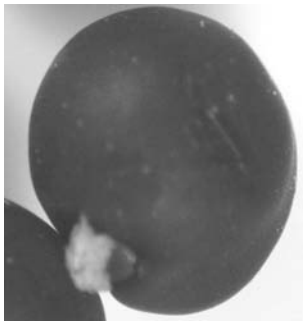
d) *Dodonaea pinifolia*



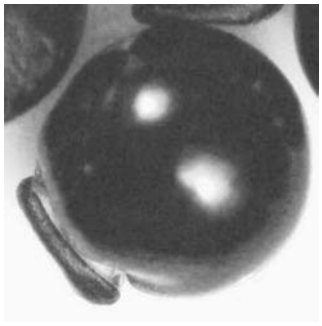
e) *Dodonaea divaricata*



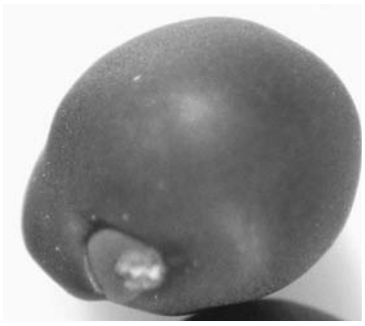
f) *Dodonaea petiolaris*



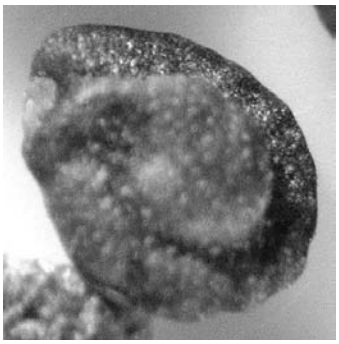
g) *Dodonaea viscosa*



h) *Distichostemon dodecandrus*

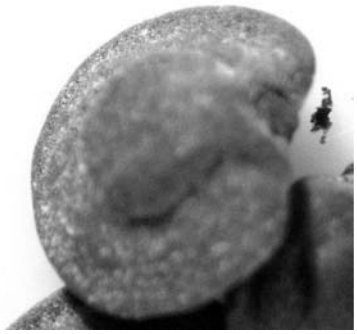


i) *Dodonaea polyandra*

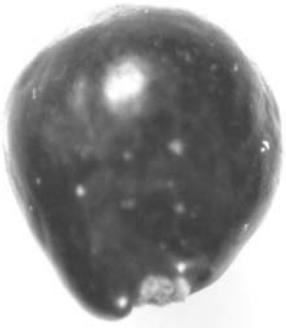


j) *Dodonaea peduncularis*

k) *Dodonaeaa ptarmicaefolia*



l) *Dodonaea hirsuta*



m) *Diplopeltis huegilii*

n) *Diplopeltis eriocarpa*

o) *Harpullia sp. Noah Creek*

Node A

Within the large clade described by node A, relationships between seven taxa (grey square in fig. 5.3) are in the main poorly resolved along with their relationship to the remaining taxa that branch off from the well-supported node B. Included within these seven taxa is *D. madagascariensis*, which is placed in only 55% of the Bayesian PP trees as sister to a strongly supported *D. polyzyga* and *D. physocarpa.* Similarities, primarily in their imparipinnate leaf morphology led both Radlkofer (1933) and Leenhouts (1983) to postulate that *D. madagascariensis* was most closely related to *D. polyzyga* from the Victoria River Downs and Kimberly Districts of north-western Australia.

Node B

This node encompasses a number of well supported clades, but unequivocal patterns of relationship between them and also between many of the species contained within them are not supported. Morphological synapomorphies are difficult to discern for any of these clades, however, across the whole clade there is a tendency for species from the same biome to occur in the same clade (fig. 5.3). The strongly supported clade III contains two well supported sister clades (IIIa and IIIb – fig. 5.3). Clade IIIa which is supported by a single nucleotide insertion in their ETSf sequences (indel 14 – table 5.1), contains all species from groups 5a and 6a of West (1984), and the species from series III, Aphanopterae, subseries 5, *Appendiculatae* of Radlkofer (1933), except *D. aptera*. All of these species have a distinctive protrusive post-floral funicular outgrowth on the seed (fig. 5.10); however, a parallel evolutionary seed condition is also found in *D. aptera*, which the molecular data place in clade VI. Clade IIIb is molecularly defined by a single nucleotide deletion in ITS2 (indel 9 – table 5.1), and contains species with disparate capsule morphologies (transverse wings, rounded wings, or globose without appendages) and lenticular, black or grey, shiny or dull seeds.

The prostrate shrub, *D. camfieldii*, is sister to species of the *D. viscosa* species complex (fig. 5.3 clade iv - discussed in detail in the next chapter), regarded by West (1984) as having the greatest combination of primitive morphological features in the genus. The molecular data indicate that *D. viscosa* is relatively recently derived within *Dodonaea*. Both clades vi and vii contain species with distinctive lenticular black seeds

that are compressed at the margins (fig. 5.9), and also capsules with rounded wings. Clade vii species generally have capsules with transverse wings, except for rounded wings in *D. sinuolata*, and small lobed capsule appendages in *D. aptera*.

It is difficult to identify or trace back key morphological or physiological innovations that have arisen in the evolutionary history of *Dodonaea*, and it may be easier to attribute the relatively rapid diversification of *Dodonaea* across continental Australia, especially in the more arid regions, to increased ecological opportunity provided by the drying climate and rainforest decline, and the subsequent increased availability of new habitats and absence of competition.

Evolutionary history of hopbushes and pepperflowers

Evidence from a recent molecular dating study (Harrington et al. in review) suggests that the dry and temperate adapted genera of Sapindaceae (*Diplopeltis* species and *Dodonaea* including *Distichostemon* species) are relatively recent radiations in the Australian flora, and are most likely to be a response to increased aridity and seasonality from the Late Miocene14 Mya to Recent (Bowler, 1982). The tempo of the radiation may have been altered by the drastic move towards severe aridity in the Late Pilocene 2-4 Mya (Fujioka et al., 2005). The respective Australian sister genera to these dry adapted frost tolerant genera are *Cossinia*, which in Australia is represented by a single endangered species, *Cossinia australiana* (http://www.environment.gov.au/cgi-bin/sprat/public/publicspecies.pl?taxon_id=3066), presently confined to a small area of relict scrub woodland in the Wide Bay District of central Queensland, and *Harpullia*, with eight species in Australia presently restricted to or found on the margins of rainforest on the eastern coast of Australia. The drying of continental Australia has restricted the distributions of the potential habitats currently preferred by the extant species of the sister genera and favoured the virtual continent-wide distribution and speciation of *Dodonaea*, while *Diplopeltis* is presently restricted to a 70 kilometre wide coastal strip in the arid zone of Western Australia.

It can only be speculated as to what favoured *Dodonaea*, in terms of extant species number, over *Diplopeltis* (70 verses 4). Is the difference due to an increased speciation rate in *Dodonaea*, or has extinction dramatically reduced extant species numbers in *Diplopeltis*, or both, or is it just due to chance events? Both genera are

adapted to drying conditions by having viscid resin exuded on young leaves and stems, which can play an ecological role in deterring herbivores or acting as antidesiccants (Langenheim, 2003), and both also have seeds with water-impermeable testa (physical dormancy) (Baskin et al., 2004; Turner et al., 2006; Cook et al., 2008) that allow their long-term maintenance in the soil-seed bank, and which also support regeneration after fire (Hodgkinson and Griffin, 1982).

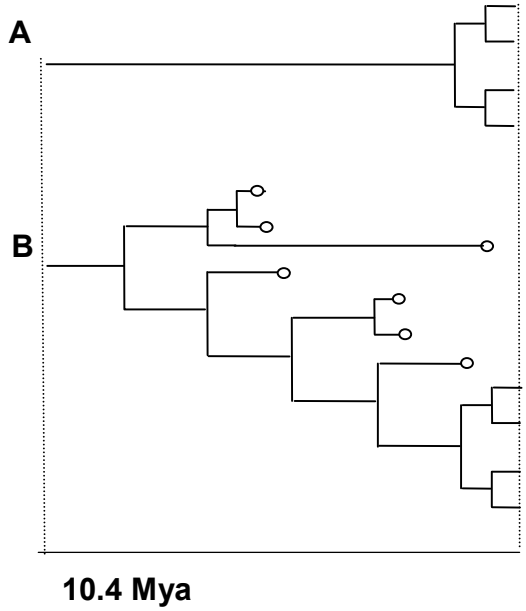
While it has been suggested that, like *Dodonaea*, *Diplopeltis* has some pollen features that may favour wind pollination (Adema et al., 1994), the presence of showy petals and an extrastamin(oid)al nectar disk suggest that they are also insect pollinated or at least maintain an intermediate condition (ambophily – wind and biotic pollination), which has been suggested as a possible transitional pathway to anemophily (Culley et al., 2002). Nectar-seeking scorpionflies (*Harpobittacus* sp.) have been reported visiting male flowers of *Diplopeltis huegelii* (Armstrong, 1979). The transition to an increasingly dry climatic regime and potential resultant change in biotic pollination vectors may have initiated the evolutionary shift to wind pollination in *Dodonaea* involving loss of petals and reduction or loss of the nectar disk along with a myriad of other morphological and pollen features of the genus that favour wind pollination (West, 1980; 1993).

In the absence of any reliable fossil dates, reliance on ages extrapolated from a relative inference of divergence dates means that a cautious approach must be taken to interpreting divergence-time estimations across *Dodonaea* and *Diplopeltis*. The mutational processes that resulted in the loss of up 29 nucleotides from helix 1 ITS1 occurred on the stem leading to the split between the four species of *Diplopeltis* and *Dodonaea* between 10.4 – 12.1 Mya (mean range from tip to tip of stem).

Another interesting feature of the evolutionary history of the sister genera is the long branch (speciation minus extinction over time = one) leading to the crown of *Diplopeltis* (when converted to time, mean 8.1 million years), compared with the stem of *Dodonaea* (2.3 million years). There are two possible evolutionary scenarios evinced by this period of time: i) evolutionary stasis in the *Diplopeltis* lineage, with a single species maintained for a long period, that eventually speciated during the severe aridification in the Late Pliocene-early Pleistocene; or alternately ii) any number of

speciation and extinction events on the stem leading to the four extant species (fig. 5.11).

Figure 5.11: Contrasting interpretations of the evolutionary history of crown group *Diplopeltis* (four species) with a long stem from split with *Dodonaea* (mean 10.4 Mya) from the Bayesian relaxed clock analyses. A) Single common ancestor (stem lineage) of extant taxa until recent speciation. B) Multiple speciation and extinctions (branches ending in circles) leading to contemporaneous species.



Similar conflicting interpretations can also be intimated for other long branches in the chronogram, particularly those leading to *Diplopeltis stuartii*, *D. petiolaris*, *D. glandulosa* and *Distichostemon*. It is difficult to speculate which scenario is correct given the non-existent macrofossil record, and an assessment of the palynological record that is questionable based on the molecular evidence presented here and in Harrington et al. in review. Currently *Diplopeltis stuartii* and *D. petiolaris* have virtually overlapping distributions throughout arid Australia and *D. glandulosa* is only known from two localities in the southwestern corner of Australia, and all represent the extant representatives of lineages that have developed specific ecological tolerances that have allowed them to survive drying environmental conditions. *Diplopeltis stuartii* is a dry

tolerant perennial herbs that can grow in sandy or lateritic soils, while *D. petiolaris* are dry tolerant shrubs that grow on virtually all arid zone soils, with its distribution only constrained by the deep sands of the various deserts in central and Western Australia. The small hairy leaves of *D. glandulosa* allow reduction in moisture evaporation by restricting air movement and reflecting sunlight. Potentially the extreme arid periods of the Pleistocene and the accompanying shifts in the extent of the central Australian deserts may have been responsible for extinction of 'at risk' narrow endemic species, leaving only those species that survived the climatic instability (Crisp et al. 2001).

Within crown *Distichostemon*, which are restricted to the monsoonal tropics, there is another example of an extant clade (six species) on a long stem (mean 4.5 million years) which also dates to the Late Pliocene-Early Pleistocene. It would be easy to speculate that there was formerly only one wide-ranging species across the top of Australia that has been regionally fragmented by the numerous wet-dry cycles of the Quaternary, but the alternative as shown in fig. 5.11B is also plausible. Further confliction scenarios can be interpreted for the well supported allopatric species pair of *Distichostemon arnhemicus* from Arnhem Land in the Northern Territory and *Distichostemon malvaceus* from Cape York Peninsula in far northern Queensland that has a mean age of 0.5 Mya (0.2-0.9 Mya). These two species are currently separated by the Gulf of Carpentaria which has been postulated as an isolation barrier due to fluctuations in sea levels and the resultant moving coastlines during the glacial and interglacial cycles of the Pleistocene (Cracraft, 1986), or alternately the current distribution of the two species may represent expansion of the descendant species into special niches following speciation.

Variations in sea levels in the Pleistocene are probably also responsible for the occurrence of a number of mainland species on offshore islands. For example, *D. lanceolata* is found on islands off the east and west coasts, *Distichostemon filamentosus* is on the Tiwi Islands, and *D. platyptera* and *D. oxyptera* are also found on islands in the Gulf of Carpentaria. Given the present occurrence of *D. polyandra* on many islands in the Torres Strait, and since its seeds do not float (Mark Harrington, unpub. data), it is envisaged that in the Pleistocene there may have been a continuous distribution of this

species from Cape York Peninsula, across the Torres Strait to the Western District of Papua New Guinea.

The temporal diversification of contemporary *Dodonaea* as a function of speciation minus extinction (figs. 5-6, and γ-statistic) indicates that, while there is the general appearance of a constant rate with no punctuated bursts of evolution, there has been significant slowing of the speciation rate towards the present. The divergence of *Dodonaea* from its sister genus *Diplopeltis* occurred between 8-13 Mya – mean 10.4 (table 5.3). Between 10.4 Mya and 5 Mya there were only five extant lineages, while within the next million years this number doubled. Bowler (1982) has hypothesised that this time in Australia was the transition from previous humid conditions to increased seasonality indicative of savannah conditions (warm, wet summers, dry winters) from the beginning of the Pliocene to 2.5 Mya, with the most torrid conditions occurring on the Nullarbor Plain. While not supported by the molecular data (PP ≥ 95%) there are potentially multiple disjunctions in *Dodonaea* between geographical species groups (allopatric sister clades southwestern-southeastern in clade i, iiia and vii) that diverged and speciated in this period (figs. 5.3 and 5.5). Similar vicariant disjunctions have also been found in a number of other plant lineages on either side of the Nullarbor Plain that relate to this period of aridification (Crisp and Cook, 2007). The next vector of *Dodonaea* diversification was the change to a winter rainfall regime in southern Australia from the Late Pliocene to the mid-Pleistocene, and periods of dramatic climatic fluctuations that have continued to the present.

Transoceanic disjunctions and dispersal biology

Within the phylogeny presented here there is a number of striking examples of trans-oceanic sister species relationships. The four species of *Cossinia* provide an extraordinary example of a relatively recent dispersal of a genus across two of the great oceans of the world. Since the current sampling only includes two of the species the overall direction of speciation or trans-oceanic dispersal within the genus cannot be assessed. *Cossinia pinnata* is found on Mauritius and Reunion Islands in the Mascarenes. These are a volcanic island chain with the existing islands of relatively recent origin: Mauritius 7.8 Mya and Reunion 2 Mya (Emerick and Duncan, 1982). The split between *Cossinia australiana* and *Cossinia pinnata* (crown *Cossinia*) in this

phylogeny is dated at mean 3.3 Mya (95% HPD 1.2 - 5.6), which is within the emergence of the islands, and probably represents a one-off long distance dispersal (LDD) most likely from Australia, and possibly the result of a rare stochastic event. Given the current restricted distribution of *Cossinia australiana* on the east coast of Australia, LDD in a westerly direction may be an indication that this species had a much wider distribution in the past.

Within *Dodonaea* there is a further relatively recent example of a unique 5400 km westward dispersal across the Indian Ocean most likely from northern Australia to Madagascar. A conservative estimation (using crown node A age - nearest node with ≥95% PP) for the divergence of the endemic *Dodonaea madagascariensis* from MRCA is 5 Mya (95% HPD 4 – 6.1). This species is now confined to the uplands of central Madagascar. Given the current species and at least one of its apparent sisters (*D. polyzyga*) both have seeds that sink in water (Mark Harrington, unpub. data), it is presumed that chance dispersal of viable propagules was by bird or floating debris assisted by the prevailing easterly winds across the Indian Ocean and is part of a continuous supplementation of the Malagasy flora by LDD (Schatz, 1996). Another similarly dated example of LDD of plant propagules comes from baobab trees (*Adansonia* spp. Bombacaceae), where there is a divergence age of 5-23 Mya for taxa in Africa/Madagascar and Australia: however, in this case the direction of the dispersal is still inconclusive (Baum et al., 1998).

In *Dodonaea* we also have a ubiquitous species that is equal to one of the world's great dispersers and colonisers. The polymorphous *D. viscosa* diverged from its sister lineage, *D. camfieldii* (see next chapter), and subsequently dispersed around the world within two million years (stem *D. viscosa* 2 Mya 95% HPD 1 – 3.2). The transoceanic dispersal of *D. viscosa* is enhanced by its ability to grow as a strandline shrub, often associated with mangroves. The seeds of *D. viscosa* are prolifically produced, resistant to salinity and able to float for extended periods (79% germination after six months soaking – West 1980), suggesting LDD dispersal by rafting or oceanic drift. They are also able to move from strand to other lowland and upland environments. For example, in Hawaii *D. viscosa* is an early coloniser of lava fields (Wagner et al.,

1990), and it is included amongst some of the earliest colonisers of the Krakatau Islands after their "sterilisation" in 1883 (Whittaker et al., 1989).

In general the seeds of *Dodonaea* are small (2-3 mm) (fig. 5.10) and bear either a prominent or small virtually-indistinguishable aril. While it has already been pointed out that seeds of *Dodonaea* species are moved small distances by ants, there are examples of other animals as probable vectors for overland dispersal. The seeds of *D. bursariifolia* (827 seeds - 87% of crop contents) have been found in the crop of a fox-killed mallee fowl (*Leipoa ocellate*), indicating possible dispersal by grainivorous birds (Booth, 1986). Unspecified *Dodonaea* seeds are part of the diet of common bronzewings crested pigeons, *Phaps chalcoptera* (Frith et al., 1974). The brush-tailed rock-wallaby (*Petrogale penicillata*) in western Victoria eats flowers and capsules of *Dodonaea viscosa* (Wakefield, 1971). West (1980) has also reported that capsules of *D. humilis*, which have dense glandular hairs have been found attached to the underbelly fur of a house mouse, *Mus musculus*. Capsules can also be dispersed by wind or by overland water flow (West, 1980).

Rate heterogeneity over time

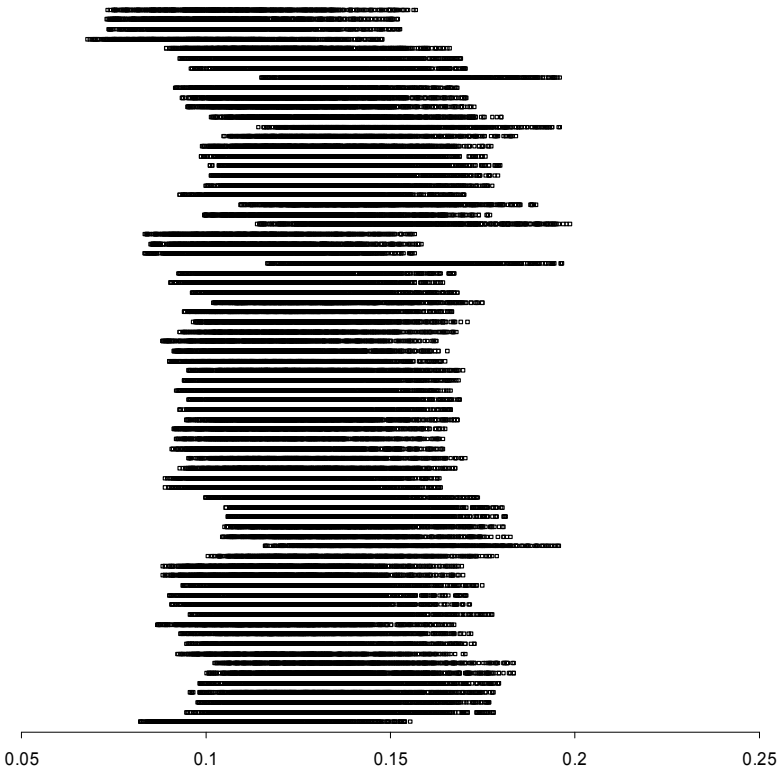
The mean node height of 0.088 substitutions per site (95% HPD 0.07-0.11) converted to units of time for the split between *Dodonaea* and the four species of *Diplopeltis* (table 5.3) resulted in the range for substitutions per site per million years (s/s/my) of 6.7 – 10 x 10^{-9} (mean 8.5). Given the qualification that this rate range is the result of using a 4 x GTR model for the combined ITS and ETSf data partition and not using a covariation model, and also the further proviso that substitution rates for the two spacers may not be selectively neutral, this rate range is amongst the highest reported for a nuclear ribosomal spacer (Kay et al., 2006). There is some limited evidence that life history traits such as growth form and the associated shorter generation time are related to evolutionary rate - herbaceous annual/perennial higher than woody perennials (Kay et al. 2006 - but see Whittle and Johnston 2003 for a report of no link between generation time and mutation rate). Although *Dodonaea* are regarded as perennial woody shrubs or small trees, the habitats that they occupy are generally fire prone and regular or even annual burning is not uncommon in some areas. While *Dodonaea* has some resprouting ability, it is generally the soil seed bank that restocks populations after fire. Many

species display an early reproductive maturity (Hodgkinson and Griffin, 1982) which enables them to survive such high frequency fire regimes: for example, garden grown *D. lanceolata* produced viable seed eight months after sprouting (Mark Harrington, unpubl. data). A recent study on the influence of a range of potential factors (shade, salt, and drought tolerance, pollen:ovule ratio, and seed bank persistence) on molecular evolutionary rates in plants found a significant positive association with drought tolerance and seed bank persistence (Whittle, 2006). The increased heritable mutation rate can be due to a number of factors that raise the level of DNA damage in both the seed and in the plant (see Whittle 2006 for further discussion). These findings indicate that potentially the molecular evolutionary rate and ultimately the diversification of *Dodonaea* may be driven by adaptations to drying environment, in particular increased seed longevity and drought tolerance.

The Bayesian relative rate test of the combined data indicates that all species of *Dodonaea* and its sister clade of *Diplopeltis* have rates of molecular evolution that fall within an overlapping range (branch lengths 95% CI distributions from MRCA of ingroup *Dodonaea* – fig. 5.12), and so are not significantly different from each other. Because terminal branches are the only ones that do not include a shared history between taxa, the relative rate test does not preclude possible concerted rate shifts (accelerations and/or decelerations) along some of the branches that contribute to the overall rate for each lineage.

An examination of the chronogram from the Bayesian reconstruction of the relative ancestral rates of molecular evolution (fig. 5.4) shows evidence of the stochastic nature of the evolutionary processes occurring in the ribosomal spacers. There are a number of taxa showing substantial rate shifts, predominantly on branches of extant species (terminal branches) and also within lineages (stem *Cossinia*) over time. The rate on the long stem of *Diplopeltis* (mean 0.74 95% HPD 0.4-113) is half that on the shorter stem of *Dodonaea* (mean 1.43, 95% HPD 0.5-2.84) and may be indicative that increases in speciation rate is associated with increases in the rates of molecular evolution (Barraclough and Savolainen, 2001).

Figure 5.12: Bayesian relative rates test 95% credibility intervals of the posterior probability distributions of summed branch lengths from the most recent common ancestor (MRCA) of split between *Dodonaea* and *Diplopeltis*. Rates of evolution are considered to be significantly different between two taxa if their 95% credibility intervals do not overlap. Taxa are in the same order as in fig. 5.3 with *D. truncatiales* at the top and *Diplopeltis petiolaris* at the bottom.



There are a number of species pairs where one species has at least twice the mean substitution rate of the terminal branch to the other; however, there is only one of these sister pairs where there is no overlap in the 95% HPD. The monsoon tropics species, *D. oxyptera* (mean 2.36, 95% HPD 1.33 – 3.6), has a far higher evolutionary rate than its sister, *D. macrossanii* (mean 0.69, 95% HPD 0.28 – 1.18), which extends from the northern extremity of the southeastern temperate biome. Compared to the stem of the sister pair (mean 1.1, 95% HPD 0.3 – 2.1), there has been an increase in *D. oxyptera* and a decrease in *D. macrossanii*. These two species diverged 2 million years ago (95% HPD 1.2 – 2.9 Mya). It is difficult to speculate as to the cause of the severe rate heterogeneity. There is morphological variation in leaf and capsule characters between the two species, but there are no gross differences that would change biological

processes that have been suggested to alter evolutionary rates (such as different metabolic rate or generation times). Nonetheless molecular change can occur independently of morphology. The rate heterogeneity may be due to distinct ecological and biogeographical factors, but for plants there has been shown to be a positive relationship between temperature and rate of molecular evolution (Davies et al., 2004b; Wright et al., 2006).

Within this phylogeny there is some limited evidence that monsoonal biome restricted species generally have faster evolutionary rates on their terminal branches. Of the five terminals within *Dodonaea* that have mean evolutionary rates greater than 2 subs/site (mean across phylogeny set to one), three (*D. oxyptera*, *D. platyptera*, and *Distichostemon malvaceus*) are restricted to the monsoonal tropics. These high rates of change may be a reflection of high fire frequencies combined with the soil seed bank effect mentioned previously. When compared to 15 other fire-intolerant species at a study site in north Queensland, *D. oxyptera* had the best reproductive success (measured percentage of new plant flowering after three years from fire – 87%) of all sampled species (Williams et al., 2006). The other two species with rates >2 subs/site (*D. hakettiana*, *D. ericoides*) are both restricted to a small number of populations, and genetic drift may be responsible for the faster evolutionary rate. While other monsoonal restricted species also have relatively fast terminal rates (*D. physocarpa*, *D. polyandra*, and *Distichostemon hispidulus*), there are some inconsistencies in the rates of the other species of *Distichostemon* which are difficult to explain biologically. These inconsistent rates may reflect the relatively young age of these lineages and/or the taxonomic concepts in this complex group of taxa with overlapping distinguishing morphological characters (Reynolds, 1984). Another species pair that shows a similar disjunct distribution to the *D. oxyptera* / *D. macrossanii* example is *D. lanceolata* and *D. serratifolia*, but in this case both terminal branches have rates that are not elevated: *D. lanceolata* mean 1.1 (95% HPD 0.43-1.8), *D. serratifolia* mean 1.1 (95% HPD 0.52-1.9).

Diversification and geographical phylogenetic structure

Species of *Dodonaea* are found in all Australian biomes except the aseasonally wet biome. There are two major areas of species concentration in *Dodonaea*, one centred on

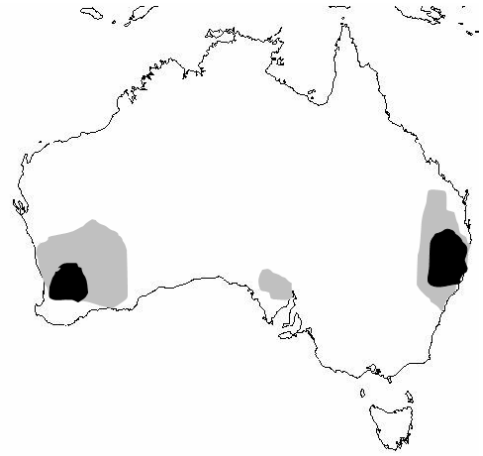
a semi-arid area that spans the boundary between the south western temperate and the eremean biomes, and the other at the northern extremity of the southeastern temperate biome with a small extension into the southern monsoon biome (fig. 5.13). Both of these regions have been identified as extant areas of endemism in the Australian flora (Crisp et al., 2001). There is only limited congruence in the patterns over time in the supported (≥95% PP) sister species relationships within the phylogenies presented here. Species from these two regions of endemism generally group together in all of the main clades of *Dodonaea* except for *Distichostemon* that are solely restricted to the monsoon tropics, and also clades iv and v which contain no species restricted to the southwest (fig. 5.3). The temporal pattern of speciation in these two regions is reflected in the biome lineage accumulation plots, given the proviso that species historical distributions are not necessarily congruent with their present distributions.

There is an extreme example of phylogeographic disjunction between a narrowly distributed species from the monsoon tropics (*D. uncinata*) and the sister species in the south western biome (*D. hackettiana*), and another between a species restricted to Queensland's monsoon tropics (*D. stenophylla*) and a species restricted to the western regions of the Eremean (*D. pachyneura*).

There has been little attempt to quantify the tempo of species diversification for arid and temperate adapted Australian plant genera. Such estimations are able to be attempted since the phylogeny is relatively robust and well resolved, with virtually complete sampling. The net diversification of the arid and temperate lineages of Australian Sapindaceae (*Dodonaea* and *Diplopeltis*) is relatively high ($r_{0.9}$ = 0.16 - 0.26 net species per million years, table 5.4, fig. 5.8) and is presumably facilitated by the ecological opportunities offered by the increased seasonality and drying of continental Australia that would have caused large numbers of extinctions in other moist adapted plant lineages and also fostered fragmentation of distributional ranges (Markgraf et al., 1995). The speciation rate is even higher ($r_{0.9}$ = 0.21 - 0.3) when measured from the stem leading to the node identified by the relative cladogenesis statistic (fig. 5.4), where there was a significant shift in the rate of lineage accumulation within *Dodonaea*. It may have been a rapid radiation that is now obscured by extinction, but the most likely scenario is that it was an initial gradual adaptation to increasingly dry environments

followed by dispersal of species out of the eremean/southwestern temperate biome into other biomes.

Figure 5.13: Geographical zonation of *Dodonaea* depicting areas with similar numbers of species. Black areas have more than 20 species, grey 10-19 species, and white 1-9 species (except for the west coast of Tasmania where there are no species) (based on West 1982).



The tempo of the radiation exceeds the mean for Eudicots ($r_{0.9}$ = 0.08) and approaches that of the fastest rate for angiosperm orders (Asterales, $r_{0.9}$ = 0.27) (Magallón and Sanderson, 2001), but does not reach the rates for South African semi-desert ice plants ($r_{0.9}$ = 0.58-1.32) (Klak et al., 2004), or the explosion of species in multiple genera generated by the final uplift of the northern Andes (see Hughes and Eastwood, 2006, for references). For the sister genera, the stem analysis indicates that at the same high extinction rate the net diversification in *Dodonaea* is more than twice that of *Diplopeltis*, whose rate is close to the mean for the angiosperms as a whole ($r_{0.9}$ = 0.077, Magallón and Sanderson, 2001). The current rate of lineage accumulation for the crown of *Diplopeltis*, which is dated to mean 2.2 Mya, is considerably higher ($r_{0.9}$ = 0.38) than that of crown *Dodonaea* ($r_{0.9}$ = 0.38), and may indicate that after a long period of evolutionary stasis or no net speciation, new species are now being derived in the Eremean zone. A similar scenario is envisaged for the monsoon tropics, where the *Distichostemon* clade has six species that have been derived from a MRCA in the Pleistocene.

The deceleration towards the present in the apparent diversification rate as measured by the γ-statistic can only result from a decrease in the rate of speciation and

not an increase in the rate of extinction (Pybus and Harvey, 2000). The same trend is visible in the two major biomes (in terms of species number) with both the southwestern and southeastern temperate biomes appearing to have a increase in speciation rate from four to one million years ago with a decline from the mid Pleistocene (fig. 5.7). Since only a single sample per species have been used as terminals in this study there is the potential for underestimates of evolutionary species in this last period due to cryptic "species" that may be genetically distinct but not known by taxonomists. It could also be an indication of incipient speciation. Alternatively the capacity of these biomes to maintain a constant rate of diversification in the last million years may have been hindered by ecological and biological space being at a premium. The potential for regional species saturation would appear more pronounced in the southwestern temperate biome where the LTT levels out considerably in the last million years. While considering that there may be contemporary regional species saturation, it is necessary to note that two new species of *Dodonaea* from the southwest have just been described (Shepherd et al., 2007), and a further species from the southeast is in the process of being described (Ian Telford pers com.). None of these are included in this study. There are also a number of widespread species with disjunct distributions and potentially reproductively isolated populations that have not been sampled. Inclusion of all these could influence assessments of diversification in *Dodonaea*. For example, there are five species whose distributions span the Nullarbor Plain, but only one of them (*D. stenozyga*) actually grows on the Nullarbor Plain.

Conclusion

Hopbushes and Pepperflowers are relatively recent additions to the diversity of the Australian flora. Within the Australian Sapindaceae, they represent the transition from the more typically aseasonally wet adapted species to species that are presently adapted to a wider range of environmental conditions, in particular the harsh conditions of the arid zone. Both lineages began their diversification during the Late Miocene in a period of climate change to increased seasonality and dryness. It could be presumed given current species distributions and that of the likely sister lineage, *Diplopeltis stuartii*, that the geographical zone of origin was the Eremean. However, the wetter conditions of the

period mean that the environmental conditions currently associated with this biome may not have existed (Truswell and Harris, 1982).

There has not been a constant rate of speciation and extinction within these lineages: rates of net diversification appear to have slowed towards the present. It is also difficult to understand why net diversification (speciation minus extinction) has been so strikingly unbalanced within the earliest branching events within *Dodonaea*. We can look towards morphological changes such as the loss of petals and nectar disk, and the move to wind pollination, when relating the success (in terms of species numbers) of *Dodonaea* over its sister *Diplopeltis*. However, there are no obvious morphological or biogeographical correlates, except possibly dispersal out of the region of origin that characterise the more successful lineages over their poorer sisters within *Dodonaea*. Alternatively, it is difficult to assess biologically why extinction rates could be so significantly unequal between sister lineages in the radiation of *Dodonaea*.

Since the beginning of the Pliestocene (~1.8 Mya), at least 42 species have been derived within *Diplopeltis* and *Dodonaea*, marking them as extremely successful components of the Australian flora in the face of the continued aridification of the continent.

Chapter 6: A species well travelled – the *Dodonaea viscosa* (Sapindaceae) complex based on phylogenetic analyses of nuclear ribosomal ITS and ETSf sequences

DIOECIA. OCTANDRIA.

Dodonaea viscosa. -- We noticed this under-shrub only at the Society Islands, where it is called *apiri*. The leaves are lanceolate, and covered with a viscid matter of agreeable balsamic odour. The flowers are arranged in panicles at the extremity of the branches; they were invariably *dioecious* in the examples we obtained at these islands, although Forster remarks, that the species is hermaphrodite in New Zealand.

This plant is the laurel of the Tahitian warriors: its branches being selected to adorn the brows of those who return victorious from war.

NARRATIVE

OF A

WHALING VOYAGE

ROUND THE GLOBE,

FROM THE YEAR 1833 TO 1836.

COMPRISING SKETCHES OF

POLYNESIA, CALIFORNIA, THE INDIAN ARCHIPELAGO,

ETC.

WITH AN ACCOUNT OF

SOUTHERN WHALES, THE SPERM WHALE FISHERY,

AND

THE NATURAL HISTORY OF THE CLIMATES VISITED.

BY

FREDERICK DEBELL BENNETT, ESQ. F.R.G.S.

FELLOW OF THE ROYAL COLLEGE OF SURGEONS, LONDON.

IN TWO VOLUMES.

VOL. II.

LONDON:
RICHARD BENTLEY, NEW BURLINGTON STREET,

Publisher in Ordinary to Her Majesty.

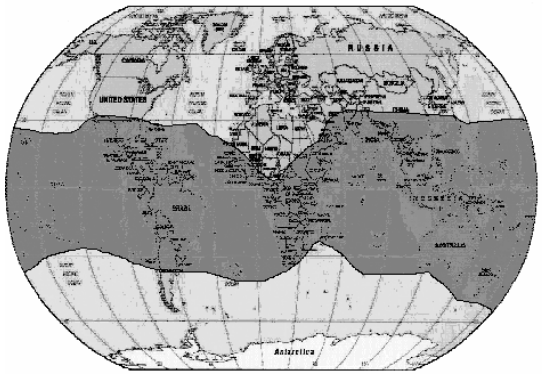
1840.

How is their vast distribution to be explained – are they 'super-nomads' of actually great vagility, or stagnant bradytelic taxa from the earliest times? (Webb et al., 1986)

These are the ponderous questions proposed by Webb et al. (1986) about the possible correlations between distribution and degree of dispersability of the twenty or so ubiquitous species of the Indo-Malayan flora. Included in this list of species is the subject of this study *Dodonaea viscosa*.

Within the largest genus of the Sapindaceae in Australia, there is one species with an omnipresent reputation. The cosmopolitan species *D. viscosa* (Hop Bush, Varnish Tree), has been the subject of taxonomic and ecological enquiry for over 150 years. Many of the world's greatest taxonomists and naturalists have wondered and postulated about its pan-tropical distribution and extensive polymorphism.

Figure 6.1: Worldwide distribution of *Dodonaea viscosa* (dark shading) and climatic factors of habitat (adapted from Liu and Noshiro, 2003)



		Range
Mean monthly temp (oC)	Max	16.9 – 34.1
	Min	5.8 – 27.5
Annual range of temp.		0.4 – 22.1
Mean monthly prec. (mm)	Max	25 – 391
	Min	3 – 167
Yearly precipitation		197 – 2748

Dodonaea has ca. 70 species restricted to continental Australia, one far northern Queensland tropical species shared with Papua New Guinea (*D. polyandra*) and one species endemic to Madagascar (*D. madagascariensis*). The only other species of the genus to spread out from mainland Australia, *D. viscosa*, has the distribution equal to some of the world's greatest transoceanic dispersers. It can be found on six continents, with a distribution extending from 44^{o} S (in the South Island of New Zealand) to 33^{o}N (in California and Arizona), and including a wide range of climatic and ecological tolerances (fig. 6.1). It is commonly found as a strandline shrub growing next to mangroves, as a prostrate shrub or small tree in temperate woodlands, and in desert

gullies and arid shrublands. It also grows at altitude (up to 4000m) in the Himalayan foothills, the interandean valleys of Peru and the mountains of Oman and Yemen. In Hawaii it has been noted that *D. viscosa* is an early coloniser of lava fields (Wagner et al., 1990), and it is included amongst some of the earliest colonisers of the Krakatau Islands after their "sterilisation" in 1883 (Whittaker et al., 1989). Within Australia, the centre of origin for the genus and for *D. viscosa* (Chapter 3), it is found in all States and Territories, but somewhat surprisingly there are few records north of 20^{o}S in the Northern Territory or Western Australia.

Across this vast distribution many different indigenous regional cultural and medicinal uses of virtually all parts of the plant have been described (for review see Ghisalberti, 1998; Prendergast and Pearman, 2001). For example, an infusion of crushed leaves is used by many rural coastal villagers in Yemen as a treatment for malaria (Ali et al., 2004), in Peru leaves have been chewed as a substitute for coca leaves (Ghisalberti, 1998), while Aboriginal Australians in north Queensland used the chewed leaves and associated juice to apply to stings of stonefish and stingrays (Cribb, 1981). Worldwide it has been the subject of an enormous number of pharmacological and phytochemical investigations (for a few examples see Khan et al., 1992; Ahmad et al., 1994; Ghisalberti, 1998; Siddiqui, 1998), and also ecological and physiological studies – for example investigations into potential latitudinal trends in wood anatomy (Liu and Noshiro, 2003), seed dormancy studies (Burrows, 1995; Baskin et al., 2004; Phartyal et al., 2005), adaptations to fire and drought (Hodgkinson and Griffin, 1982; Hodgkinson, 1992; Mishio, 1992; Harrington and Driver, 1995), allelopathic potential (Maraschin-Silva and Aqüila, 2005), pollen productivity and dispersal (Reddi et al., 1980; Cambon et al., 1992) and leaf stomatal conductance (Barradas et al., 2004). Besides the major taxonomic reviews that have included *D. viscosa* (see below), it has also been the subject of botanical investigation into breeding system (Rivers, 1971), leaf gland structure (Collins, 1920) and parthenocarpy – development of a fruit without fertilisation or seed production (Joshi, 1938). It is also of horticultural interest, with the purple-leaved form of *D. viscosa* that originated in New Zealand's South Island being a popular garden ornamental (West, 1984).

Since *D. viscosa* is so widely distributed and extremely polymorphic, especially in leaf and capsule morphology, it has had a long and convoluted taxonomic history. Bentham (1863) described three varieties, one of which was further subdivided by Radlkofer (1900) into three forms. Bentham (1863) further believed that most of the extra-Australian taxa could be assigned to one of the three Australian varieties, with the only potential exceptions being possible species or varieties from the Sandwich Islands, South Africa and Mexico. The infrageneric classification of *Dodonaea* by Radlkofer (1933) places *D. viscosa* (with 'D. stenoptera' from Hawaii) in Series 1 Cyclopterae, Subseries 2 Oospermae, closely related to *D. camfieldii*. Further development of the *D. viscosa* complex was spawned by Sherff (1945; 1947), who reduced some varieties and expanded others: for example 18 varieties and 12 forms were described for the Hawaiian *D. eriocarpa* Sm. A taxonomic treatment of American *D. viscosa* by Lippold (1978) recognised five species in the complex for that region (*D. arizonica*, *D. bialata*, *D. elaeagnoides*, *D. linearifolia* and *D. viscosa*). In a revision of Sapindaceae in the south-eastern United States Brizicky (1963) commented on the extreme polymorphism in *D. viscosa* especially in size and shape of leaves, capsules and seeds, in some cases even on the same plant, leading him to remark that "in some instances two sheets of the same collection of *D. viscosa* have been cited as different varieties". In noting that the differences between species or varieties were weak, Brizicky (1963) only recognised one species, *D. viscosa*, for the region.

The Australian members of the *D. viscosa* species complex were treated by West (1984) in her revision of the genus, with classification within the complex based principally on a numerical analysis of leaf measurements (West and Noble, 1984) and limited use of capsule morphology. West (1984) divided *D. viscosa* in Australia into seven subspecies, which included reducing *D. angustissima* DC. and *D. cuneata* Smith to subspecies (table 6.1). She had examined many of the specimens viewed by Radlkofer and Sherff and concluded that the majority could be placed in ssp. *angustifolia*, ssp. *burmanniana* or ssp. *viscosa*, and that New Zealand material possibly belonged in ssp. *spatulata*. For the genus as a whole, West (1984) made predictions of phylogeny based on putative evolutionary trends in morphology, especially in capsule, seed inflorescence architecture, breeding system and leaf morphology. *Dodonaea*

viscosa was regarded as having the highest proportion of primitive characters for the genus.

Table 6.1: Seven subspecies of Australian *Dodonaea viscosa* from West (1984), with general description of habit and distribution.

D. viscosa ssp. *viscosa*
Spreading shrub with most constant morphology; only member of complex with distinctive 2 occasionally 3 winged capsules, relatively large elliptic leaves, spreading dense shrub to 2m of mostly coastal situations found in north-eastern Queensland (also Port Macquarie) and off-shore islands extending to Papua New Guinea. Also occurs in tropical regions of the Americas (Florida, West Indies, Venezuela and Brazil), Africa, Madagascar and Asia (Philippines, Celebes, Java, and Borneo).
D. viscosa ssp. *burmanniana*
Large shrub or small tree (2-6 m) of wet sclerophyll forest or woodland association, capsule 3-4 winged, generally leaves narrower than ssp. *viscosa*, distributed in Australia from Grafton to Cairns. Extra-Australian tropical regions of Mexico, Central America to South America (Peru, Brazil, Argentina), Kenya, Tanzania, China, India, Malaya, Philippines and Indonesia.
D. viscosa ssp. *angustifolia*
Compact shrub 1.5-3(-5) m of dry sclerophyll forest or woodland associations, capsule same as ssp. *burmanniana*, leaves linear-lanceolate, from south-eastern Queensland to far-eastern Victoria (Orbost). Florida, Mexico, Peru, Ecuador, Bolivia, West Indies, South Africa (probably Madagascar), the Middle East to India, China, the Philippines and Fiji.
D. viscosa ssp. *angustissima*
Erect multistemmed shrub (2-4 m) in arid areas of open woodland and sand plains, and on margins of sand dunes, widespread in the southern and central parts of Australia. Capsule same as ssp. *burmanniana*, leaves generally linear to narrow-oblong.
D. viscosa ssp. *cuneata*
Compact, spreading shrub (1-3 m) of semi-arid areas southeastern Australia. Capsule same as ssp. *burmanniana*, leaves obitriangular to obovate.
D. viscosa ssp. *mucronata*
Erect to spreading shrub (1.5-4 m), widespread but restricted to rocky hills and ranges in arid central Australia. Capsule same as ssp. *burmanniana*, leaves spathulate.

D. viscosa ssp. *spatulata*

Erect to spreading shrub (1.5-4 m), widespread in temperate and semi-arid southern Australia including Tasmania. Capsule same as ssp. *burmanniana*, leaves same as ssp. *mucronata* with attenuate base and shorter petiole.

Another major review of the *D. viscosa* complex was undertaken by Leenhouts (1983) as part of his treatment of Sapindaceae for the Flora Malesiana series. In his extensive re-examination of the historical literature and inspection of herbarium specimens, Leenhouts questioned the systems of Radlkofer and Sherff, and identified that the cause of the systematic confusion "was to all probability the insufficient understanding of the flower conditions". He proposed the recognition of three species in the *D. viscosa* complex (which he now called the *D. angustifolia* complex) – *D. angustifolia* (pantropical and subtropics - inland), *D. elalaeagnoides* (Florida and West Indies) and *D. viscosa* (pantropical - coastal). The three species were distinguished on leaf texture, breeding system (bisexual versus dioecious), presence or absence of staminodes, and capsule form. However, in conclusion he stated that while less confusion may now surround *D. viscosa*, he may have produced another taxonomic headache by creating a "distinctly more unmanageable" *D. angustifolia* complex (Leenhouts, 1983).

From an evolutionary perspective the difficulty with virtually all of these systems is that they include very few regionally endemic taxa, with a number of the species or subspecies or varieties being widely distributed. This suggests as an illustration, that strandline species in say, Madagascar, South Africa and Australia are more closely related to each other than to the species or subspecies or varieties that grow in upland regions of those same countries. Since the revision of Leenhouts (1983) a number of regional taxonomies have rejected the concept of inland and upland species or subspecies, and maintained a single species concept (*D. viscosa*) (Smith, 1985; Wagner et al., 1990; Davies and Verdcourt, 1998). For the Fijian region, where Leenhouts recognised two species, Smith (1985) concluded from an examination of 75 collections from coastal and forest situations up to 1,100 m. altitude, that "morphological variation is totally haphazard…and…it would seem apparent that any

two individuals, whether with bisexual or unisexual flowers, can produce progeny which flourish in any reasonably available habitat". For Hawaii, where there has been as many as four species, 22 varieties and eight forms accepted (Sherff, 1945; 1947), Wagner et al. (1990) reviewed over 500 Hawaiian collections, and while distinguishing three intergrading entities based on capsule morphology, none were formally recognised because there was no correlation with other morphological characters.

In this study analyses of molecular data have been used to evaluate the monophyly of the *Dodonaea viscosa* complex, and to develop hypotheses concerning the biogeographical distribution and evolution of the complex. Bayesian analyses were performed on nuclear ribosomal transcribed spacers (ITS1 and ITS2, together referred to as ITS), and partial external transcribed spacer (referred to as ETSf), regions often used to determine generic and species relationships (Soltis and Soltis, 1998). All chloroplast regions trialled for this study (*trnL* intron*, trnL-trnF* spacer, *psbA-trnH*, *ndhF*, and *rpl16*) had no sequence variation between samples.

Materials and methods

Study species biological profile

Members of the *D. viscosa* complex can be dioecious or polygamo-dioecious (♂ and bisexual on one individual, and ♀ and bisexual on another), prostrate or upright shrubs, or small trees to 9 m. tall (fig. 6.2). This variation in breeding system has also been recorded in Hawaiian plants, with variation also occurring on individual plants over time (Sakai et al., 1995). The usually viscous leaves are simple, sessile or petiolate, and extremely variable in shape from the small virtually linear form of ssp. *angustissima* to the large broad elliptic leaves of ssp. *viscosa*. The small unisexual or bisexual flowers have pedicels to 1.5 cm. long, generally with 3-4 sepals, and lack petals. Stamen numbers are variable (6-10, in the main 8); the 2-4 carpelate ovary is well adapted for wind pollination (Reddi et al., 1980). Prolific numbers of fruit are formed, which are 2-4 winged papery dehiscent capsules, which vary in size (body 0.8 – 2.3 cm high, wings 0.3 – 1.2 cm. wide) between the subspecies, or between upland or lowland regional forms. There are generally two seeds per locule (but not all are viable), which are hard, lenticular with a small funicular aril around the abscission scar (Brizicky, 1963; Corner,

1976). They also vary in size, usually correlated with capsule size. The seeds have physical dormancy (PY) due to a water-impermeable seed coat (Baskin et al., 2004). It has been noted that field grown seed can reach reproductive maturity in less than three years (Hodgkinson and Griffin, 1982). Chromosome numbers have been reported for over 50 Australian populations, and for several specimens from India, Japan, New Zealand and "*Dodonaea ericocarpa*" from Hawaii, all being gametic determinations of n = 14 and/or sporophytic counts 2n = 28 (Hair and Beuzenberg, 1959; Love, 1975; Love, 1976; Carr, 1978; Love, 1984; West, 1984; Gill et al., 1990; Oginuma et al., 1997). However, counts of 2n = 32 (Brizicky, 1963), and n = 13 (Sutaria, 1930), may represent evidence of cytological evolution (by dysploidy) within the complex; this needs to be established more firmly by further counts to rule out the possibility of them being erroneous counts.

Figure 6.2: Polygamo-dioecious (♀ and bisexual) *Dodonaea viscosa* ssp. *viscosa* on front dune next to mangroves Half Moon Bay Beach, Cairns, Australia, with a majority of 2-winged capsules. Photo: Mark Harrington 15/8/04 IMG0494JPG.

Sampling

The naming of all Australian specimens included in this study follows West (1984), while all other samples are labelled *D. viscosa* followed by country or island of origin. Members of the *D. viscosa* species complex display a wide variety of forms. Due to the vast worldwide distribution of *D. viscosa*, I was relying on other institutions and individuals for samples from outside Australia. As a result, it was not possible to adequately sample both the distribution and potential morphological variability of the species complex. As far as possible samples are drawn from a range of locations, and from both lowland and highlands (see appendix 6.1).

The 15 Australian samples include representatives from all seven subspecies recognised by West (1984). There are 33 extra-Australian samples plus five outgroup taxa as identified in the analyses presented in the previous chapter – *D. camfieldii*, *D. filiformis*, *D. macrossanii*, *D. oxyptera* and *D. vestita* (appendix 6.1). Also included are the currently recognised species, *D. biloba* and *D. procumbens*, since preliminary analyses indicated that they are potentially closely related to other members of the complex. There is little or no sampling from the Indonesian/southeast Asian region, from the multitude of smaller Pacific Islands, or from the Florida region of the United States of America. It is considered that sampling includes enough of the diversity to make some assessment of the evolutionary history of the complex.

Extraction of DNA, PCR amplification and sequencing protocols follow those described or referenced in previous chapters. Alignment of sequences was relatively easy as there were only a few insertions or deletions of nucleotides between samples. After removal of the invariant ribosomal 5.8S gene (164 base pairs), a secondary structure mask was added to the alignment following the procedures described in the previous chapter. All sequences have been lodged with GenBank (appendix 6.1) and the complete aligned data with secondary structure mask can be accessed from Treebase (file) or accompanying CD.

Model selection

The best models of nucleotide evolution for the separate partitions were selected by comparison of corrected Akaike Information Criterion (AIC_c) scores from the 56 available DNA nucleotide substitution models in MrAIC.pl 1.4 (Nylander, 2004) with

the trees and likelihood scores under each model estimated using PHYML (Guindon and Gascuel, 2003b). For both ITS and ETSf, the general-time-reversible (GTR) model with a correction for gamma-distributed rate heterogeneity (Γ) was selected. To determine whether secondary structure partitioned data could produce a better phylogenetic interpretation of the alignment than non-partitioned data or a partition by spacer region only, AIC_c scores were compared. Likelihoods were estimated using the Optimizer module of the PHASE software package for the best AIC_c topology generated by MrAIC.pl 1.4 for the combined dataset. Because of the low number of nucleotide changes in the separate stem partitions and a noticeable bias towards C↔T transitions in all partitions, the best-fit RNA specific paired sites model was not tested since it would potentially over-parameterise the already small data partitions (Sullivan and Joyce, 2005).

Phylogeny estimation

Bayesian estimation of the posterior probability distribution of phylogenies amongst the sampled taxa were estimated using the Metropolis-coupled Markov chain Monte Carlo algorithm of MrBayes-3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) on the concatenated alignment (ITS loops + ITS stems + ETSf loops + ETSf stems) with a GTR model plus gamma distribution applied to each of the four partitions. Phylogenetically informative loop insertions of nucleotides were coded as additional characters and added to the loop partition where they occurred. Estimation of all model parameters was unlinked across the partitions – MrBayes commands unlink shape=(all) pinvar=(all) statefreq=(all) revmat=(all); prset applyto=(all) ratepr=variable. Two replicate analyses were carried out, each starting with a randomly generated tree and run for two million generations with sampling of the phylogenetic hypothesis every 250 generations. A conservative approach to burnin estimation was taken by eliminating the first 2000 generated trees. The remaining trees (12,000) from the two runs were combined to produce a 50% majority rule consensus tree. All log files generated were analysed in Tracer v1.3 2003-2005 (http://evolve.zoo.ox.ac.uk/beast/) to confirm that the likelihood scores had maintained stability after the burnin and that the effective sample size (ESS – number of effectively independent uncorrelated samples from the posterior distribution

that the Markov chain is equivalent to) for each estimated parameter was greater than 200 (Drummond et al., 2006).

Relative evolutionary rates analysis

Bayesian Markov chain Monte Carlo (MCMC) phylogeny and molecular evolutionary rates estimations of the combined aligned matrix without indels and with *D. camfieldii* as the only outgroup, were performed using the BEAST package (Drummond and Rambaut, 2007). For the combined ITS and ETSf dataset the GTR model with gamma distributed rate variation and an assumed proportion of invariable sites were applied. The mean substitution rate was fixed to one to allow for an estimation of the relative substitution rates (estimated in substitutions per site) across the phylogeny (i.e. a ratogram). Substitution rates were estimated under an assumption of a relaxed clock with the rates in each branch independently drawn from an assumed log normal distribution (uncorrelated log normal model – UCLN). The degree of autocorrelation of substitution rate variation is estimated directly from the data (covariance statistic) and is not assumed *a priori*. The coalescent tree prior was assumed to follow a constant size demographic model. All model parameters, priors, operators, MCMC options and taxon group sets generated in the BEAUti utility are on the associated CD.

Two independent MCMC chains for the combined dataset were run for ten million generations sampling the topology every 10,000 generations and parameter values every 1000 generations. The output was examined using Tracer v1.4, and summarised (excluding the appropriate burn in) using Logcombiner v1.4.6. TreeAnnotator v1.4.6 was used to generate a maximum clade credibility chronogram scaled to mean node heights (all part of the BEAST package http://beast.bio.ed.ac.uk/). Because of a lack of reliable fossils (see below), divergence times were determined by scaling the relative node heights into time by setting the divergence of the entire *Dodonaea viscosa* complex from *Dodonaea camfieldii* to 2 million years ago (Mya), corresponding to the estimated mean date for this node from analyses presented in the previous chapter. The analysis was also performed on a dataset with no outgroups to determine whether the 95% CI coefficient of variation was significantly removed from zero for *Dodonaea viscosa* complex, which is indicative of sequences that are not evolving in a clock-like manner.

Fossils

There have been a number of fossils that have been attributed to *Dodonaea*. Early last century Berry (1914; 1916) reported over 20 described species of *Dodonaea* fossils based on leaves and/or fruits that could be assigned to the *D. viscosa* complex. The oldest was from the Palaeocene in northern England, with a number of other species described from the Eocene and Oligocene of North America. Berry (1916) expressed doubts about the European material, while a number of the other designations have also been challenged (Axelrod, 1939; Brown, 1960). Various species have also been described from Oligocene and Miocene fossil floras from North America (Axelrod, 1939; MacGinitie, 1953; Becker, 1961; 1966), but some of these have been subsequently assigned to other genera (Manchester and Donoghue, 1995). There have been no fossils from Australian Tertiary floras that have been attributed to *Dodonaea*. While some of the leaf fossils cannot be unequivocally rejected, relaxed-clock estimations presented in chapter 3 cast doubt over the Oligocene and Miocene fossils.

Results

For the *D. viscosa* complex (excluding outgroups and including *D. biloba* and *D. procumbens*) ITS 1 varied in length from 221-225 bp. Length variation was due to a cytosine deletion from the single strand terminal bulge of helix II in the sample from Peru, and a synapomorphic 3 bp insertion which is probably the result of a replication slippage event that causes a lateral bulge in helix II in *D. biloba*, *D. procumbens* (x2), *D. viscosa* from New Zealand (x4), *D. viscosa* ssp. *angustifolia*, *D. viscosa* ssp. *angustissima* (x2), *D. viscosa* ssp. *cuneata*, *D. viscosa* ssp. *mucronata*, and *D. viscosa* ssp. *spatulata*. ITS2 was generally 223 bp except for a cytosine deletion from the single strand region between helices III and IV in the two samples from New Caledonia and three samples from New Zealand.

The RNA transcript folding structure of Helix III in ITS2 also contains one couplet that is particularly informative of phylogenetic relationships in the *D. viscosa* complex: stem pair 351/382 has C:G in the majority of taxa: while a compensatory base change (U:A) is found in samples from China, Japan, and Taiwan (x2): and all samples from the Americas (Arizona x2, Bolivia, Brazil, Columbia, Mexico, and Peru) have the

non-canonical transitional U:G. It is interesting to note that this stem pairing at wider divergence levels, i.e. Sapindaceae alignment (chapter 3 – stem pair 383/415 dataset 3 CD) and *Dodonaea* alignment (chapter 4 – stem pair 380/411 dataset 4 CD), does not show evidence of being overly saturated (homoplasious with multiple substitutions arising in unrelated taxa). In the Sapindaceae alignment there is only one species of *Acer* that deviates from the constant C:G pair (*A. saccharum* U:G), while in the *Dodonaea* alignment there is also a U:G pair in the sister taxa *D. ericoides* and *D. divaricata*.

The partial ETSf is also generally conserved in length (449-453). Length variation in this region is also due to informative indels, the only exception being a four bp duplication CTACCTAC in *D. angustifolia*. The other two indels include a single guanine lateral bulge in helix VI in *D. procumbens* (x2), *D. viscosa* ssp. *angustifolia*, *D. viscosa* ssp. *cuneata* and *D. viscosa* ssp. *spatulata*, and a adenine-cytosine duplication in the unpaired nucleotides separating helices VI and VII in all samples from the Americas (taxa as above).

Model testing using comparison of AICc scores selected the four-way partition of ITS stems + ITS loops + ETSf stems + ETSf loops each with a separate GTR model with gamma distribution for rates amongst sites as the best fit for the combined dataset (table 6.2).

Table 6.2: Comparison of AICc scores for three different partitions of the ITS/ETSf concatenated alignment. n = number of model parameters, AICc = AIC + 2k(k +1)/n – k where AIC = -*L* + 2k - 1, with k the number of estimable parameters, *L* the likelihood estimate and n the number of nucleotide positions (combined = 903).

Partition	Model	n	-ln*L*	AIC_c
None	GTR + Γ	11	2752	5526
ITS + ETSf	2 x GTR + Γ	22	2735	5515
stems + loops	2 x GTR + Γ	22	2703	5451
ITS (stem + loops) + ETSf (stem + loops)	4 x GTR + Γ	44	2673	5438

The standard deviation of the split frequencies at the end of the simultaneous runs (0.006) indicated that they had converged onto stationary distribution (Ronquist et al., 2005). The relative mean substitution rate ratios (ETSloops/ITSloops = 0.24, ITSstems/ITSloops = 0.4, and ETSstems/ITSloops = 0.28, as estimated in PHASE version 2.0b 2005,

http://www.bioinf.man.ac.uk/resources/phase) indicate that ITS loops contain by far the most phylogenetic information, and an examination of the reversible substitution rate ratios (appendix 6.2) indicates the highest proportions of mutations in each data partition are transitions, with C↔T changes in the majority.

One of the 12,000 trees from the Bayesian analyses is shown in fig. 6.3. There is support for *D. camfieldii* (100% PP, posterior probability) as sister to a monophyletic *D. viscosa* complex that also includes two currently described species, *D. biloba* and *D. procumbens*. The *D. viscosa* clade divides into two strongly supported lineages (fig. 6.3, I and II), and within both of these subclades there are a number of strongly supported groupings of terminal taxa. However relationships within these subclades are largely unresolved. Despite the relatively sparse sampling there is some geographical structuring in the recovered phylogeny, with strongly supported groupings for all mainland American samples, all New Zealand samples, and all Hawaiian samples. Australian samples are found in both clades I and II, and where there are multiple samples of subspecies (ssp. *viscosa*, *angustissima* and *burmanniana*) they always group together with ≥95% PP.

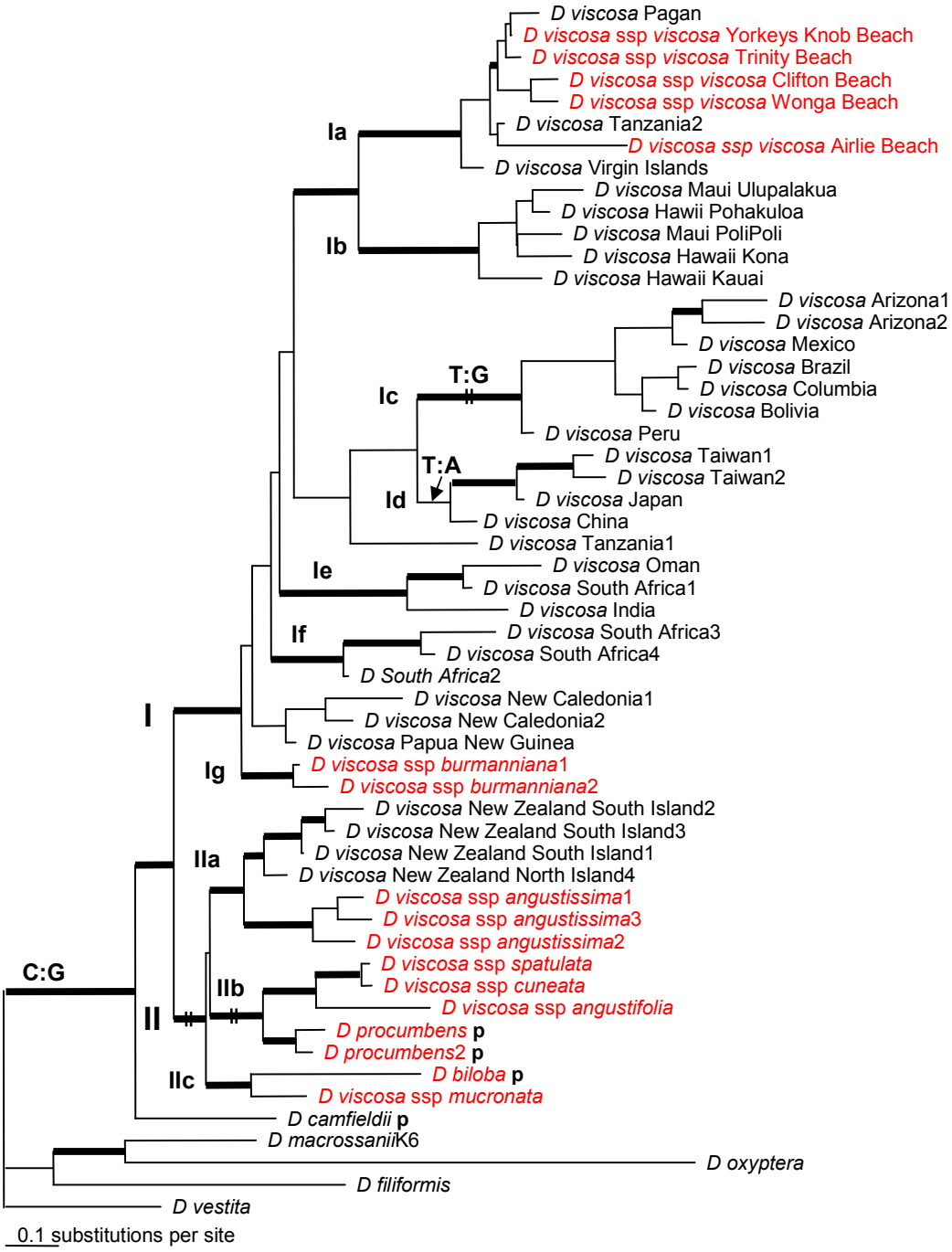
Evolutionary rates analyses

The estimated coefficient of variation of the branch rates (σ_r = the standard deviation divided by the mean) for a truncated dataset (no outgroups) was 0.39 (95% highest posterior density, HPD upper 0.7, 95% HPD lower 0.06) indicating significant rate heterogeneity among branches and that the spacer regions within the *D. viscosa* complex are not evolving in a clocklike manner. There was no substitution rate autocorrelation between related branches in the phylogeny, since the 95% CI of the covariance statistic was not significantly removed from zero (mean 0.03, HPD -0.1–0.17).

The ratogram with branch lengths extrapolated from the mean substitution rate per site to the dimension of time (substitutions per million years), based on the crown node of the split of *D. camfieldii* and *D. viscosa* (mean 2 Mya) taken from the analyses in the previous chapter, is shown in fig. 6.4.

Figure 6.3 (next page): One of the 12,000 trees from the Bayesian MCMC analyses of the combined dataset (ITS + ETS) estimated with the model GTR + I for each of the separate stem

and loop partitions. Branches with posterior probabilities ≥ 95 % are in bold. Taxa in red are samples from Australia. Synapomorphic indels are marked with double vertical lines. Nucleotide assignments for ingroup taxa for the stem pair 351/382 from the ITS secondary structure alignment are marked on branches. Intraspecific groups I and II and subclades are numbered, and **p** indicates taxa that are prostrate.



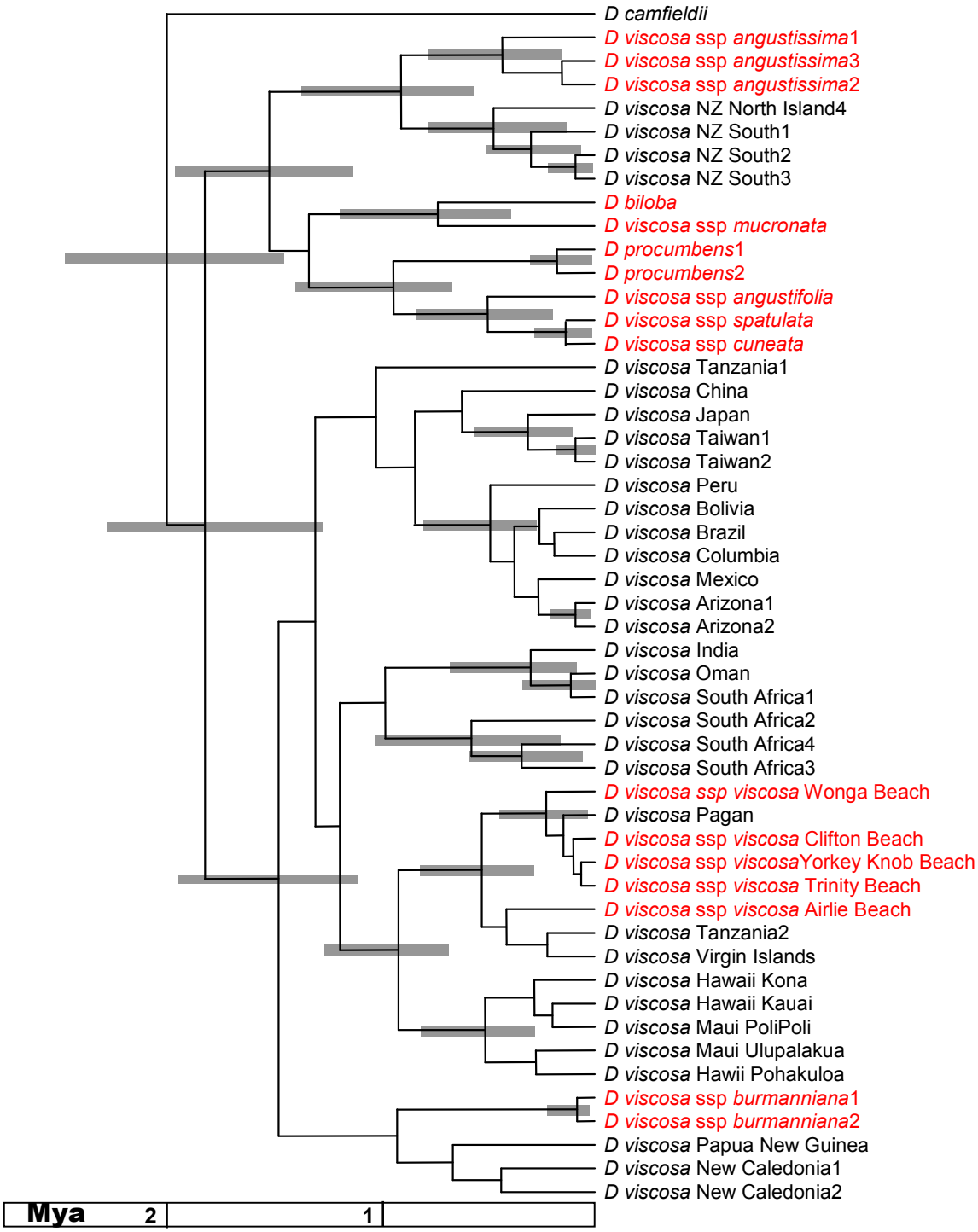

D camfieldii
D viscosa ssp angustissima1
D viscosa ssp angustissima3
D viscosa ssp angustissima2
D viscosa NZ North Island4
D viscosa NZ South1
D viscosa NZ South2
D viscosa NZ South3
D biloba
D viscosa ssp mucronata
D procumbens1
D procumbens2
D viscosa ssp angustifolia
D viscosa ssp spatulata
D viscosa ssp cuneata
D viscosa Tanzania1
D viscosa China
D viscosa Japan
D viscosa Taiwan1
D viscosa Taiwan2
D viscosa Peru
D viscosa Bolivia
D viscosa Brazil
D viscosa Columbia
D viscosa Mexico
D viscosa Arizona1
D viscosa Arizona2
D viscosa India
D viscosa Oman
D viscosa South Africa1
D viscosa South Africa2
D viscosa South Africa4
D viscosa South Africa3
D viscosa ssp viscosa Wonga Beach
D viscosa Pagan
D viscosa ssp viscosa Clifton Beach
D viscosa ssp viscosaYorkey Knob Beach
D viscosa ssp viscosa Trinity Beach
D viscosa ssp viscosa Airlie Beach
D viscosa Tanzania2
D viscosa Virgin Islands
D viscosa Hawaii Kona
D viscosa Hawaii Kauai
D viscosa Maui PoliPoli
D viscosa Maui Ulupalakua
D viscosa Hawii Pohakuloa
D viscosa ssp burmanniana1
D viscosa ssp burmanniana2
D viscosa Papua New Guinea
D viscosa New Caledonia1
D viscosa New Caledonia2
Mya
2
1

Figure 6.4 (previous page): The posterior distribution of relative branching times. Grey bars represent the 95% HPD interval for the relative (mean root height scaled to one) branching times from the Bayesian uncorrelated log normal relaxed clock estimation of phylogeny. Bars appear only on nodes that receive more than 95% PP. Scale represents extrapolation of branch lengths using crown node of *D. camfieldii* split from *D. viscosa* (mean 2 Mya) from the analyses in chapter 3. Taxa in red are samples from Australia.

Discussion

The generally low levels of nucleotide divergence in nuclear ribosomal spacers within a species means there are relatively few studies that have been able to use these markers for intraspecific studies (for examples see Dick et al., 2003; Noyes, 2006; Nettel and Dodd, 2007). This study presents an example for a single widespread species, using a combination of molecular data from ITS and ETSf ribosomal spacers to provide some insights into evolutionary relationships within the *D. viscosa* complex.

The widely distributed *D. viscosa* evolved in Australia from its most recent common ancestor (MRCA), which it shared with the prostrate shrub *D. camfieldii* that is presently confined to the central and south coast region of New South Wales. The divergence of these taxa is estimated to have occurred in the Late Pliocene to early-Pleistocene (2.7–1.4 million years ago Mya, 95% HPD). The molecular data support a monophyletic *D. viscosa* that includes two previously described prostrate species *D. procumbens* and *D. biloba*. Within *Dodonaea* this growth condition is also found in *D. humifusa* from southwestern Australia. There is enough nucleotide variation within the nuclear ribosomal spacers of *D. viscosa* to identify two geographically-based intraspecific lineages and a number of regional subclades that are also supported by specific molecular elements (fig. 6.3). Given the qualification of the currently limited geographical sampling, the following hypothesis of evolutionary history and intraspecific diversification of *D. viscosa* is tentatively formulated.

It is envisaged that the original diversification of *D. viscosa* into two intraspecific groups occurred in the Pleistocene 1.1–2.1 Mya (95% HPD), when they last shared a common ancestor. The spit is also molecularly characterised by a possible replication slippage event (a three base pair insertion) in the ITS1 sequences of all group II samples. Within Australia there is geographical structuring of the two intraspecific

groups. Group I are generally strandline shrubs or confined to wet sclerophyll forests in the coastal strip from northeastern Queensland to the New South Wales border. The only area of perceived present distributional overlap with group II (ssp. *angustifolia* and ssp. *cuneata*) is an area of southern Queensland that extends into New South Wales (fig. 6.5). It is not known whether members of the two lineages co-occur in this region. At the southern extremity of Queensland in the area of current overlap is the McPherson Range, which runs east to the Pacific Ocean as a spur from the north-south Great Dividing Range. The McPherson Range contains an extensive montane area of wet forest. It forms a natural barrier between temperate and tropical floras in one of the principle floristic zones of Australia known as the McPherson Macleay overlap (Burbridge, 1960). The range has been identified as an historical ecological barrier to gene flow during the Pleistocene in a species of elapid snake that principally occurs in open patches of rainforest and wet sclerophyll forest (Keogh et al., 2003), and has been claimed to be responsible for pre-Pleistocene vicariance in a open forest species of sedge frog (James and Moritz, 2000). While the current distribution of a species and its ecological tolerances do not necessarily reflect historical distributions and ecology, the fact that no current species of *Dodonaea* is found in mesic forests suggests that climate driven vicariance in the Pleistocene may have restricted gene flow between the northern tropical coastal populations of *D viscosa* and those of southern temperate and more arid areas, and this is reflected in groups I and II in figs. 6.3 and 6.4. While the current distributional overlap may indicate that the McPherson Range does not now constitute a barrier to gene flow for *D. viscosa*, no intergradation of morphologies or putative hybridisation between ssp. *burmanniana* and the other two subspecies that occur in the area, either ssp. *angustifolia* or ssp. *cuneata* has been reported (West, 1984).

Figure 6.5 (next page): Current distribution of *Dodonaea viscosa* in Australia. Adapted from information in Australia Virtual Herbarium and West (1984). Darker shading shows approximate distribution of *Dodonaea viscosa* subsp. *viscosa* and subsp. *burmanniana*, while lighter shading is the distribution of all the other subspecies. Circle represents the region of overlap in the two distributions, while the line indicates the approximate placement of the McPherson Range

Group I

The most plausible scenario is that group I originated in Australia. It is most likely that one Australian lineage in this clade (*D. viscosa* ssp. *burmanniana*) is the sister to the remainder of this group (PP ≤ 95%). One possibility is that the other Australian components of this clade (ssp. *viscosa*) may represent a recent reintroduction: i.e. a secondary dispersal back to mainland Australia. However such an origin is not supported by the molecular data (PP ≤ 95%). In Australia this subspecies is found only as a mangrove associate or on sand dunes close to the foreshore, along the north-eastern Queensland coast and offshore islands. Outside of this region of continuous distribution there may be occasional accidental dispersals to other regions as evidenced by herbarium collections from Port Macquarie and from the Hastings River in north-eastern New South Wales. There is a herbarium specimen (Latz, P. K. 3282 CANB) collected in 1972 from a swale of a stable coastal dune on the Wessel Islands off the north-eastern coast of Northern Territory, however, there are no other floristic records of this

subspecies from this region (Woinarski et al., 2000; Kerrigan and Albrecht, 2007), suggesting a chance dispersal to and subsequent extinction from this region. Other evidence of relatively recent dispersals of ssp. *viscosa* is seen with it only recently being found on an islet in the lagoon on Aldabra in the western Indian Ocean (Wickens, 1979). Similarly, in this study there is a sample collected in 2005 from Pagan however, regional floras report *D. viscosa* in the Northern Marianas as only found on Asuncion Island, on the open slope of a volcano between 250-500 m altitude, and it is reported as not being seen on lower slopes or on other islands in the group (Fosberg et al., 1975). There is also some evidence that this form growing as a mangrove associate or as a strand plant may be pantropically distributed (samples from Pagan, Tanzania, and Virgin Islands, fig. 6.3, subclade Ia), but this can only be tested with further sampling.

Group I contains all of the samples from outside of Australia except New Zealand. It is presumed that the range expansion of group I is the result of long distance dispersal (LDD) from the northeastern coast of Australia with subsequent isolation and regional differentiation of populations fostered by the dramatic climatic cycling and associated environmental change of the Quaternary. There are a number of genetically divergent lineages that can be identified in this clade.

All five samples from the Hawaiian Islands of Maui and Hawaii form a monophyletic group (subclade Ib, fig 6.3.) which last shared a common ancestor with subclade Ia (the *D. viscosa* ssp. *viscosa* group) in the mid-Pleistocene 0.5–1.2 Mya (fig. 6.4). There was a single successful establishment of *D. viscosa* on the Hawaiian Islands by LDD, as hypothesised by Carlquist (1974) either directly from Australia or from another currently unsampled location. In Hawaii *Dodonaea viscosa* occurs on all the main islands except Kaho'olawe, and is a prominent component of the flora from sea level to the sub-alpine zone. Despite high levels of morphological variation amongst the populations Wagner (1990) considered that all the specimens that he had examined (including those previously described as *D. eriocarpa*, *D. sandwicensis* and *D. stenoptera*) belonged to one intergrading entity that cannot be distinguished from *D. viscosa*.

The genetic divergence of all the samples across the vast distribution from North and South America (subclades Ic) is molecularly marked by a two bp insertion (AC

duplication, fig. 6.3) and a hemi-CBC (compensatory base change on one side of a helix pairing) to a transitional noncannonical stem-pairing (U:G). The dating of these molecular events can go back as far as the crown of group I (0.8-1.7 Mya, fig. 6.4), which is the nearest lineage splitting node with ≥95% PP. It is envisaged that the genetic differentiation of this lineage is the result of the movement and subsequent isolation of populations from initial strandline or littoral environments into dry upland habitats due to Pleistocene climatic cycling.

In tropical and southern Africa two varieties of *D. viscosa* have been recognised (Davies and Verdcourt, 1998; Pearman, 2002): var. *viscosa* occurs along the coasts of West Africa (from Senegal to Nigeria) and East Africa (from Kenya to Mozambique), whereas var. *angustifolia* occurs naturally in mainly inland areas above 1,000 m. from the Democratic Republic of Congo in the west, to Ethiopia and Somalia in the east, and South Africa in the south. The estimate of the phylogeny in figs. 6.3 and 6.4 places the sample from Hillcrest on the east coast of South Africa closer to samples from upland India and Oman than to samples from the upland region of the West Cape Province of South Africa in subclades If. Similarly the upland sample, Tanzania 1, does not group with the coastal sample Tanzania 2 that is placed in subclades Ia with Australian ssp. *viscosa*. Given the current limited regional sampling it is not possible to provide scenarios for the apparent genetic disjunctions, other than to suggest that there may have been multiple foundings of small genetically isolated populations from different sources followed by genetic drift.

This estimate of the phylogeny (fig. 6.3) indicates that group I in Australia contains the relatively large lanceolate to narrow-elliptic leaves of *D. viscosa* ssp. *burmanniana* and ssp. *viscosa*, while throughout the extra-Australian distribution of *D. viscosa* there are a multitude of forms of leaves (table 6.1), including the form that has been previously attributed to D. angustifolia (Leenhouts, 1983; West 1984). Within Australia, however, the narrow lanceolate–linear form only occurs in group II.

Group II

The distribution of the subspecies of *D. viscosa* included in this clade covers virtually all of Australia, compared with the rather restricted distribution of group I. Within this clade there are at least three strongly-supported evolutionary lineages (IIa, b and c in fig.

6.3) that also have generally overlapping distributions. West (1984) noted that there is morphological intergradation between the subspecies from this clade, particularly in the higher rainfall, southern temperate areas of Australia, but not where they overlap in the arid zone (table 6.3). She has also reported putative hybridisation between the prostrate *D. procumbens* and *D. viscosa* in two populations, one in western Victoria and the other in central South Australia, suggesting that there may be ongoing gene flow between members of this clade. While low growing (<0.5 m) plants from group I can sometimes be found in various coastal situations, the truly prostrate habit (where stems root at nodes) has independently arisen twice in group II (fig. 6.3). This clade provides evidence of a further dispersal in the mid-Pleistocene (0.5–1.2 Mya, fig. 6.4) of *D. viscosa* from mainland Australia to New Zealand. The New Zealand samples are a strongly supported monophyletic group (100% PP) sister to a monophyletic *D. viscosa* ssp *angustissima*; however, there is no morphological synapomorphy that diagnoses the subclade.

Table 6.3: Morphological intergradation between subspecies of *Dodonaea viscosa*, and the regions where it has been recorded (from West 1984).

Intergradation of subspecies	occurrence
ssp. spatulata—*ssp*. mucronata	Western New South Wales, south-western Queensland and western Australia.
ssp. spatulata—*ssp*. angustissima	Southern South Australia, Victoria, New South Wales, southern Queensland and Western Australia.
ssp. spatulata—*ssp*. cuneata	Areas of New South Wales and Victoria where the two subspecies are sympatric.
ssp. cuneata—*ssp*. spatulata—*ssp*. angustissima	New South Wales

Dispersal

Amongst the late 19th and early 20th century botanists and naturalists there was much speculation as to the agents of dispersal for *D. viscosa*. Hillebrand (1888) speculated that the glutinous capsules would stick to the plumage of birds, while Radlkofer (1895) and Engler (1921) considered that wind was the main agent of fruit dispersal. However, one earlier report of an investigation of Indo-Malayan stand flora (Schimper, 1891) indicated that while the capsules were too large to be transported by winds across a broad tract of sea, the seeds floated in seawater for from 10 to 60 days. Given this time

period and a sea surface current of 1.8 km per hour (1 knot), the seeds could float for anywhere from 430 to 2,600 km. He also indicated that the floation is derived from unoccupied space in the cavity of the seed due to the folded cotyledons incompletely filling the seed. Guppy (1906) also performed floating experiments but found that only half the seeds floated in seawater, which indicated the "purely accidental nature" of the imperfect filling of only some of the seeds. Over many years of observation on Pacific islands, he claimed never to have seen the seeds or capsules in floating or stranded seed-drift and "that if we placed the agencies of dispersal in their order of effectiveness they would be, first granivorous birds, then currents, and lastly man" (Guppy 1906).

Another treatment of the dispersal capabilities of *D. viscosa* is provided by Ridley (1930), who suggests that the inland forms are losing or have lost their floating ability, and this may account for the only moderate success rate in Guppy's Hawaiian floating experiments, where most of the plants are inland forms. Ridley (1930) also dismisses the suggestion that capsules are capable of sticking to birds and considerers it "most improbable that birds would eat this dry, tasteless capsule". While he too had not seen capsules or seed in the sea-shore drift of the Malay region, his most acceptable modes of dispersal were ocean currents assisted by land-based wind dispersal. West (1980) also conducted floatation experiments with *D. viscosa* capsules in still sea water and found that 30% were still afloat after 100 days, while 79% of seeds germinated after soaking in sea water for six months, indicating the potential of successful establishment after LDD dispersal.

Unspecified *Dodonaea* seeds are part of the diet of common bronzewings crested pigeons, *Phaps chalcopteva* (Frith et al., 1974). The brush-tailed rock-wallaby (*Petrogale penicillata*) in western Victoria eats flowers and capsules of *Dodonaea viscosa* (Wakefield, 1971). It is not known whether the seeds are likely to survive digestion or promote regurgitation. Capsules can also be dispersed by wind or by overland water flow, and floatation may be assisted by the viscous nature of the fruits reducing water permeability (West, 1980).

It is most likely that the efficacy of present-day transoceanic dispersal of *D. viscosa* is enhanced by seeds that have PY and the ability to float and subsequently establish, coupled with populations that have a strandline or coastal distribution. Further

sampling, especially from the multitude of Pacific Islands, may indicate whether dispersal follows prevailing wind or ocean currents mediated by West Wind Drift and the Antarctic Circumpolar Current (Sanmartın et al., 2007). Dispersal to other habitats, particularly to upland environments, is probably animal assisted. Due to the widespread cultural use of *D. viscosa*, current distributions may also reflect human-assisted dispersal.

One species or many?

Taxonomists have long struggled to accommodate widespread species that have high levels of intraspecific morphological variation, especially when that variation can intergrade geographically or ecologically repeatedly across different regional biotas. An extreme example can be seen in another genus from Sapindaceae, *Allophylus* L., which at any one time has had recognised up to 255 species, many based on regional taxonomies that have not considered the repeated gradation of morphological characters across its vast tropical distribution and ecological range. A comprehensive review of the species complex led Leenhouts (1967) to draw a provisional taxonomic conclusion to treat *Allophyllus* as a single polymorphic species. For another sapindaceous species, *Pometia pinnata* J. R. & G. Forst., that also has reticulate morphological variation across its range, Whitmore (1976) reintroduced the term ochlospecies, originally proposed by White (1962) for the African *Diospyros mespiliformis* A.DC. (Ebenaceae). Similarly, *D. viscosa* exhibits morphological variation that does not correlate with geography, and can be considered a complex ochlospecies (sensu. White 1962). This may be more satisfactory than attempting to recognise formal intraspecific taxa when the genetically distinct elements cannot be defined by morphological synapomorphies.

I leave the concluding remarks on this subject to the Pacific naturalist H. B. Guppy. Commenting on plant dispersal and distributions in the Pacific Ocean and particularly on the flora of the Hawaiian islands, Guppy (1906; 1917) referred to *D. viscosa* as part of what he called the “Malayan era on non-endemic genera”. These are genera not entirely represented by endemic species, but he suggests that the polymorphic *D. viscosa* is in the earliest stages of “that process by which a solitary widely-ranging species, alone representing its genus, becomes ultimately in each group the parent of a number of peculiar species”.

Conclusion

Morphological markers that allow the consistent recognition of the distinct entities identified by the sequence data in *D. viscosa* do not exist. While certain morphological forms have been recognised that are placed together in the estimate of phylogeny, these do not correlate with the actual lineages in the species. Hence it is not a useful exercise to try to define morphological subunits of *D. viscosa*, and the term ochlospecies is useful in communicating the variable nature of the species throughout its range.

In *Dodonaea viscosa* we have an example of a highly vagile species within a genus, that has transversed the worlds oceans east and west, and through LDD has initiated an Australian component in a diverse worldwide array of modern biotas, all within the confines of Quaternary climatic changes.

Conclusion

This thesis is grounded in the angiosperm family Sapindaceae. The hierarchical structure of this study and the incorporation of a temporal framework has allowed for the study of the tempo of evolution from the broad family level to insight into the evolutionary history of a particularly remarkable species complex. Also significantly, this study has presented systematic investigations using models of evolution that more appropriately represent the evolutionary history of the sequences to generate generally well-supported phylogenetic hypotheses.

Within Sapindaceae s.l a new and more robust inference of phylogenetic relationships has been provided by the use of molecular data from the chloroplast genome. A four subfamily arrangement (Sapindoideae, Hippocastanoideae, Dodonaeoideae and Xanthoceroideae) has been proposed to replace the previous hypotheses of a two-subfamilial arrangement within Sapindaceae based on morphological and anatomical criteria. Tribal concepts within Hippocastanoideae and Dodonaeoideae have been redefined due to the near complete sampling within these subfamilies. Ongoing collaborative work now sees over 70% of genera sampled, allowing further redefining of the tribal concepts within Sapindaceae, particularly in the large Cupanieae. Further studies are still required to determine the relationship of Sapindaceae to other families in the Sapindales.

The timing of divergences within Sapindaceae has identified a case of seemingly irreconcilable age estimates of *Acer* and *Aesculus* derived from fossil evidence verses molecular dating. Although such contradictions are nothing unusual per se, the case presented in this study is of wider interest as it suggests that either a considerable number of morphologically unambiguous fossils have nevertheless been wrongly assigned or that some lineages experienced biologically hardly plausible changes in their mean evolutionary rates. Further exploration of alternative datasets and the fossil record might address and perhaps solve the striking discrepancy.

A relatively novel approach for the alignment and analysis of small RNAs (ITS and ETSf) that considers secondary structure has been adopted in this thesis. The analyses of nuclear ribosomal spacers for Dodonaeoideae identified that models that

account for RNA secondary structure are more appropriate than traditional 4-state substitution models. The derived phylogeny for Dodonaeoideae was largely congruent with the plastid data. The secondary structure alignment generated for Dodonaeoideae also provides the methodological framework for similar analyses of *Dodonaea* and allied genera.

Previous phylogenetic hypotheses based on distribution and macro- and micro-morphology have deemed that *Dodonaea* has had a relatively long evolutionary history in Australia. However Bayesian relaxed clock analyses of nuclear ribosomal spacers, and incorporation of dates from similar analyses of Sapindaceae suggest an origin of the genus in the Late Miocene, and subsequent radiation associated with the increased continental aridity from the the Late Miocene to the Recent. Phylogenetic analyses of *Dodonaea* and *Diplopeltis* incorporating RNA specific models has identified that all species of *Distichostemon* should be transferred to *Dodonaea*, and that a new monotypic genus should be erected for *Diplopeltis stuartii*. Further analyses to resolve all intrageneric groups within *Dodonaea* should include the three newly described species and sequencing of other nuclear and chloroplast markers.

The current molecular investigations into *D. viscosa* have identified that while there are least two distinct evolutionary lineages within the complex, they do not correlate with any distinct morphological subunits. Subsequently it is proposed that *D. viscosa* be recognized as an ochlospecies. There are many other possible avenues for further studies that could range from population to regional landscape level analyses. Further intensive sampling is needed to advance our understanding of patterns of diversification over time of this wide-ranging species.

Bibliography

Acevedo-Rodriguez, P. 1993. Systematics of *Serjania* (Sapindaceae) Part I: A revision of *Serjania* Sect. Platycoccus. Memoirs of the New York Botanical Garden 67:2-91.

Acevedo-Rodriguez, P. 2003. Melicocceae (Sapindaceae): *Melicoccus* and *Talisia*. Flora Neotropica Monograph 87:1-179.

Ackerly, D. D., and M. J. Donoghue. 1998. Leaf size, sapling allometry, and Corner's rules: a phylogenetic study of correlated evolution in maples (*Acer*). American Naturalist 152:767-791.

Adema, F., P. W. Leenhouts, and P. C. van Welzen. 1994. Sapindaceae. Flora Malesiana ser. I, 11:419-768.

Ahmad, I., M. Ahmad, and A. Ahmad. 1994. Antimicrobial activity of Dodonaea viscosa oil. Fitoterapia 65:167-168.

Albach, D. C., and M. W. Chase. 2004. Incongruence in Veroniceae (Plantaginaceae): evidence from two plastid and a nuclear ribosomal DNA region. Molecular Phylogenetics and Evolution 32:183-197.

Ali, A. N., K. Al-rahwi, and U. Lindequist. 2004. Some medicinal plants used in Yemeni herbal medicine to treat malaria. African Journal of Traditional, Complementary and Alternative Medicines 1:71-76.

Alvarez, I., and J. F. Wendel. 2003. Ribosomal ITS sequences and plant phylogenetic inference. Molecular Phylogenetics and Evolution 29:417-434.

Anderson, C. L., K. Bremer, and E. M. Friis. 2005. Dating phylogenetically basal eudicots using rbcL sequences and multiple fossil reference points. American Journal of Botany 92:1737-1748.

APG. 1998. An ordinal classification for the families of flowering plants. Annuals of Missouri Botanical Gardens 85:531-553.

APG. 2003. An update of the angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. Botanical Journal of the Linnean Society 141:399-436.

Armstrong, J. A. 1979. Biotic pollination mechanisms in the Australian flora - a review. New Zealand Journal of Botany 17:467-508.

Axelrod, D. I. 1939. Contributions to paleontology. A Miocene flora from the western border of the Mohave Desert. Carnegie Institute of Washington 516:1-129.

Bailey, C. D., T. G. Carr, S. A. Harris, and C. E. Hughes. 2003. Characterisation of angiosperm nrDNA polymorphism, paralogy and pseudogenes. Molecular Phylogenetics and Evolution 29:435-455.

Baldwin, B. G., and M. J. Sanderson. 1998. Age and rate of diversification of the Hawaiian silversword alliance (Compositae). Proceedings of the National Academy of Sciences USA. 95:9402-9406.

Barraclough, T. G., and V. Savolainen. 2001. Evolutionary rates and species diversity in flowering plants. Evolution 55:677-683.

Barraclough, T. G., and A. P. Vogler. 2002. Recent diversification rates in North America tiger beetles estimated from a dated mtDNA phylogenetic tree. Molecular Biology and Evolution 19:1706-1716.

Barradas, V. L., A. Ramos-Vázquez, and A. Orozco-Segovia. 2004. Stomatal conductance in a tropical xerophilous shrubland at a lava substratum. International Journal of Biometeorology 48:119-127.

Baskin, J. M., B. H. Davis, C. C. Baskin, S. M. Gleason, and S. Cordell. 2004. Physical dormancy in seeds of *Dodonaea viscosa* (Sapindales, Sapindaceae) from Hawaii. Seed Science Research 14:81-90.

Baum, D. A., R. L. Small, and J. F. Wendel. 1998. Biogeography and floral evolution of baobabs (*Adansonia*, Bombacaceae) as inferred from multiple data sets. Systematic Biology 47:181-207.

Becerra, J. X. 2003. Evolution of Mexican *Bursera* (Burseraceae) inferred from ITS, ETS, and 5S nuclear ribosomal DNA sequences. Molecular Phylogenetics and Evolution 26:300-309.

Becker, H. F. 1961. Oligocene plants from the upper Ruby River Basin, southwestern Montana. Geological Society of America Memoirs 82:1-127.

Becker, H. F. 1966. Additions to and revision of the Oligocene Ruby paper shale flora of southwestern Montana. Contributions from the Museum of Paleontology. The University of Michigan 20:89-119.

Bena, G., M.-F. Jubier, I. Olivieri, and B. Lejeune. 1998. Ribosomal external and internal transcribed spacers: combined use in the phylogenetic analysis of *Medicago* (Leguminosae). Journal of Molecular Evolution 46:299-306.

Bentham, G. 1863. Flora Australiensis. Lorel Reeve and Co, London.

Bentham, G., and J. D. Hooker. 1862a. Genera Plantarum. Lorel Reeve and Co, London.

Bentham, G., and J. D. Hooker. 1862b. Sapindaceae. Pages 388-413 *in* Genera Plantarum Lorel Reeve and Co, London.

Benton, M. J., and P. C. Donoghue. 2007. Paleontological evidence to date the tree of life. Molecular Biology and Evolution 24:26–53.

Berry, E. W. 1914. The affinities and distribution of the lower Eocene flora of southeastern North America. Proceedings of the American Philosophical Society 53:129-250.

Berry, E. W. 1916. The lower Eocene floras of southeastern North America. United States Geological Survey, Professional Paper 91:1-481.

Biffin, E., M. G. Harrington, M. D. Crisp, L. A. Craven, and P. A. Gadek. 2007. Structural partitioning, paired-site models and evolution of the ITS rDNA in *Syzygium* and Myrtaceae. Molecular Phylogenetics and Evolution 43:124-139.

Bindewald, E., and B. A. Shapiro. 2006. RNA secondary structure prediction from sequence alignments using a network of k-nearest neighbor classifiers. RNA 12:342-352.

Booth, D. T. 1986. Crop and gizzard contents of two mallee fowl. Emu 86:51-53.

Boulter, M. C., J. N. Benfield, H. C. Fisher, D. A. Gee, and M. Lhotak. 1996. The evolution and global migration of the Aceraceae. Philosophical Transactions of the Royal Society of London, Series B. 351:589-603.

Bowler, J. M. 1982. Age, origin and landform expression of aridity in Australia. Pages 35-46 *in* Evolution of the Flora and Fauna of Arid Australia (W. R. Barker, and P. J. Greenslade, eds.). Peacock Publications, Adelaide.

Bremer, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. Evolution 42:795–803.

Brizicky, G. E. 1963. The genera of Sapindales in the southeastern United States. Journal of the Arnold Arboretum 44:462-501.
Brown, R. W. 1960. Corkwood in the Eocene Flora of the southeastern United States. Journal of Paleontology:429-432.
Buckler, E. S., A. Ippolito, and T. P. Holtsford. 1997. The evolution of ribosomal DNA divergent paralogues and phylogenetic implications. Genetics 145:821-832.
Buckley, T. R. 2002. Model misspecification and probabilistic tests of topology: evidence from empirical data sets. Systematic Biology 51:509-523.
Buijsen, J. R. 1995. Leaf anatomy of *Harpullia*, *Majidea* and *Conchopetalum* (Sapindaceae). Blumea 40:345-361.
Buijsen, J. R., P. C. van Welzen, and R. W. van der Ham. 2003. A phylogenetic analysis of *Harpullia* (Sapindaceae) with notes on historical biogeography. Systematic Botany 28:106-117.
Burbridge, N. T. 1960. The phytogeography of the Australian region. Australian Journal of Botany 8:75-212.
Burnham, K. P., and D. R. Anderson. 2003. Model selection and inference: a practical information-theoretic approach, 2 edition. Springer-Verlag, New York.
Burrows, C. J. 1995. Germination behaviour of the seeds of six New Zealand woody plant species. New Zealand Journal of Botany 33:365-377.
Cambon, G., J. C. Ritchie, and P. Guinet. 1992. Marker pollen of long-distance transport in the atmosphere of southern Ontario, Canada. Canadian Journal of Botany 70:2284-2293.
Carlquist, S. 1974. Island biology. Columbia University Press, New York.
Carr, G. D. 1978. Chromosome numbers of Hawaiian flowering plants and the significance of cytology in selected taxa. American Journal of Botany 65:236-242.
Chase, M. W., and V. A. Albert. 1998. A perspective on the contribution of plastid rbcL DNA sequences to angiosperm phylogenetics. Pages 488-507 *in* Molecular systematics of plants II: DNA sequencing (D. E. Soltis, P. S. Soltis, and J. J. Doyle, eds.). Kluwer Academic, Dordrecht.
Chase, M. W., M. F. Fay, and V. Savolainen. 2000. Higher-level classification in the angiosperms: new insights from the perspective of DNA sequence data. Taxon 49:685-704.
Coleman, A. W., and J. C. Mai. 1997. Ribosomal DNA ITS-1 and ITS-2 sequence comparisons as a tool for predicting genetic relatedness. Journal of Molecular Evolution 45:168-177.
Collins, M. I. 1920. On the structure of the resin-secreting glands in some Australian plants. Proceedings of the Linnean Society N.S.W. 45:329-336.
Cook, A., S. R. Turner, J. M. Baskin, C. C. Baskin, K. J. Steadman, and K. W. Dixon. 2008. Occurrence of physical dormancy in seeds of Australian Sapindaceae: A survey of 14 species in nine genera. Annals of Botany doi:10.1093/aob/mcn043.
Corbett, S. L., and S. R. Manchester. 2004. Phytogeography and fossil history of *Ailanthus* (Simaroubaceae). International Journal of Plant Sciences 165:671-690.
Corner, E. J. 1976. The seeds of dicotyledons. Cambridge University Press, Cambridge.
Cracraft, J. 1986. Origin and evolution of continental biotas: speciation and historical congruence within the Australian avifauna. Evolution 40:977-996.

Crane, P. R., S. R. Manchester, and D. L. Dilcher. 1990. A preliminary survey of fossil leaves and well-preserved reproductive structures from the Sentinel Butte Formation (Paleocene) near Almont, North Dakota. Fieldiana, Geology, New Series 20:1-63.

Cribb, A. B. 1981. Wild medicine in Australia. Collins, Sydney.

Crisp, M. D., and L. G. Cook. 2005. Do early branching lineages signify ancestral traits? Trends in Ecology and Evolution 20:122-128.

Crisp, M. D., and L. G. Cook. 2007. A congruent molecular signature of vicariance across multiple plant lineages. Molecular Phylogenetics and Evolution 43:1106-1117.

Crisp, M. D., L. G. Cook, and D. A. Steane. 2004. Radiation of the Australian flora: what can comparisons of molecular phylogenies across multiple taxa tell us about the evolution of diversity in present-day communities? Philosophical Transactions of the Royal Society of London, Series B. 359:1551-1571.

Crisp, M. D., S. Laffan, L. H. P., and A. Monro. 2001. Endemism in the Australian flora. Journal of Biogeography 28:183-198.

Cronquist, A. 1988. The Evolution and Classification of Flowering Plants, 2 edition. New York Botanic Gardens, New York.

Culley, T. M., S. G. Weller, and A. K. Sakai. 2002. The evolution of wind pollination in angiosperms. Trends in Ecology and Evolution 17:361-369.

Dahlgren, G. 1989. An updated system of classification. Botanical Journal of the Linnean Society 100:197-203.

Davies, F. G., and B. Verdcourt. 1998. Flora of tropical East Africa - Sapindaceae. A. A. Balkema, Rotterdam.

Davies, T. J., T. G. Barraclough, M. W. Chase, P. S. Soltis, D. E. Soltis, and V. Savolainen. 2004a. Darwin's abominable mystery: Insights from a supertree of the angiosperms. Proceedings of the National Academy of Sciences USA. 101:1904-1909.

Davies, T. J., V. Savolainen, M. W. Chase, J. Moat, and T. G. Barraclough. 2004b. Environmental energy and evolutionary rates in flowering plants. Proceedings of the Royal Society of London B Series 271:2195-2200.

Davis, C. C., C. O. Webb, K. J. Wurdack, C. A. Jaramillo, and M. J. Donoghue. 2005. Explosive radiation of Malpighiales supports a mid-Cretaceous origin of modern tropical rain forests. The American Naturalist 165:E36-E65.

de Jong, P. C. 2002. World maple diversity. Pages 2-11 *in* Proceedings of the 2002 International Maple Symposium held at Westonbirt Arboretum and the Royal Agricultural College in Gloucestershire, England (S. J. Wiegrefe, H. Angus, D. Otis, and P. Gregorey, eds.). Westonbirt Arboretum.

Delphino, F. 1890. Note ed osservazioni botaniche. Decuria seconde. Malpighia 4:3-33.

Denduangboripant, J., and Q. C. Cronk. 2001. Evolution and alignment of the hypervariable arm 1 of *Aeschynanthus* (Gesneriaceae) ITS2 nuclear ribosomal DNA. Molecular Phylogenetics and Evolution 20:163-172.

Dick, C. W., K. Abdul-Salim, and E. Bermingham. 2003. Molecular systematic analysis reveals cryptic Tertiary diversification of a widespread tropical rain forest tree. The American Naturalist 162:691-703.

Donoghue, M. J., R. G. Olmstead, J. F. Smith, and J. D. Palmer. 1992. Phylogenetic relationships of Dipscales based on rbcl sequences. Annals of the Missouri Botanic Garden 79:333-345.

Doyle, J. A. 1992. Revised palynological correlations of the lower Potomac Group (USA) and the Cocobeach sequence of Gabon (Barremian-Aptian). Cretaceous Research 13:337–349.

Doyle, J. A., and C. L. Hotton. 1991. Diversification of early angiosperm pollen in a cladistic context. Pages 69–195 *in* Pollen and spores: patterns of diversification (S. Blackmore, and S. H. Barnes, eds.). Clarendon Press, Oxford, U.K.

Doyle, J. J., and J. L. Doyle. 1990. Isolation of plant DNA from fresh tissue. Focus 12:13-15.

Drummond, A. J., S. Y. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biology 4:e88.

Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7:214.

Edwards, K. J., and P. A. Gadek. 2001. Evolution and biogeography of *Alectryon* (Sapindaceae). Molecular Phylogenetics and Evolution 20:14-26.

Emerick, C. M., and R. A. Duncan. 1982. Age progressive volcanism in the Comores Archipelago, western Indian Ocean and implications for Somali plate tectonics. Earth and Planetary Science Letters 60:415-428.

Endlicher, S. F. 1837. *Diplopeltis*. Pages 12-13 *in* Enumeratio plantarum quas in Novae Hollandiae ora austro-occidentali ad Fluvium Cygnorum et in Sinu Regis Georgii collegit Carolus liber baro de Hügel (S. F. Endlicher, G. Bentham, E. Fenzi, and H. Schott, eds.). F. Beck, Vienna.

Engler, A. 1921. Sapindaceae*in* Vegetation der Erde. IX. Die Pflanzenwelt Afrikas insbesondere seiner tropischen Gebiete 3(2):265-288 (A. Engler, and O. Drude, eds.). W. Engelmann Leipzig.

Eriksson, O., and B. Bremer. 1992. Pollination systems, dispersal modes, life forms, and diversification rates in angiosperm families. Evolution 46:258-266.

Eriksson, T. 1998. AutoDecay version 4.0.2 Department of Botany, Stockholm University, Stockholm, Sweeden.

Farris, J. S., M. Kallersjo, A. G. Kluge, and C. Bult. 1995. Constructing a significance test for incongruence. Systematic Biology 44:570-572.

Fay, M. F., C. Bayer, W. S. Alverson, A. Y. de Bruijn, S. M. Swensen, and M. W. Chase. 1998. Plastid *rbcL* sequences indicate a close affinity between *Diegodendron* and *Bixa*. Taxon 47:43-50.

Felsenstein, J. 1978. Cases in which parsimony or compatibilty will be positively misleading. Systematic Zoology 27:401-410.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791.

Felsenstein, J. 2003. Inferring phylogenies. Sinauer Associates, Massachusetts.

Forest, F., R. Grenyer, M. Rouget, T. J. Davies, R. M. Cowling, D. P. Faith, A. Balmford, J. C. Manning, S. Proches, M. van der Bank, G. Reeves, T. A. Hedderson, and V. Savolainen. 2007. Preserving the evolutionary potential of floras in biodiversity hotspots. Nature 445:757-760.

Forest, F., J. N. Orouin, R. Charest, L. Brouillet, and A. Bruneau. 2001. A morphological phylogenetic analysis of *Aesculus* L. and *Billia* Peyr. (Sapindaceae). Canadian Journal of Botany 79:154-169.

Fosberg, F. R., M. V. Falanruw, and M.-H. Sachet. 1975. Vascular flora of the northern Marianas Islands. Smithsonian Institution Press, Washington.

Frith, H. J., B. K. Brown, and R. D. Barker. 1974. Food of the crested and common bronzewing pigeons in inland New South Wales. Australian Wildlife Research 1:129-144.

Fujioka, T., J. Chappell, M. Honda, I. Yatsevich, K. Fifield, and D. Fabel. 2005. Global cooling initiated stony deserts in central Australia 2-4 Ma, dated by cosmogenic Ne-21-Be-10. Geology 33:993-996.

Gadek, P. A., E. S. Fernando, C. J. Quinn, S. B. Hoot, T. Terrazas, M. C. Sheahan, and M. W. Chase. 1996b. Sapindales: Molecular delimitation and infraordinal groups. American Journal of Botany 83:802-811.

Galtier, N. 2004. Sampling properties of the bootstrap support in molecular phylogeny: influence of nonindependence among sites. Systematic Biology 53:38-46.

Gardner, P. P., and R. Giegerich. 2004. A comprehensive comparison of comparative RNA structure prediction approaches. BMC Bioinformatics 5:140-158.

Gautheret, D., D. Konings, and R. R. Gutell. 1995. G·U base pairing motifs in ribosomal RNA. RNA 1:807-814.

George, A. S., and G. Erdtman. 1969. A revision of the genus *Diplopeltis* Endl. (Sapindaceae). Grana Palynologica 9:92-109.

Ghisalberti, E. L. 1998. Ethnopharmacology and phytochemistry of *Dodonaea* species. Fitotereapia 69:99-113.

Gianoli, E. 2004. Evolution of a climbing habit promotes diversification in flowering plants. Proceedings of the Royal Society, London, Series B 271:2011-2015.

Gill, B. S., V. K. Singhal, Y. S. Bedi, and S. S. Bir. 1990. Cytological evolution in the woody taxa of Pachmarhi Hills. Journal of Cytology and Genetics 25:308-320.

Gillespie, J. J. 2004. Characterising regions of ambiguous alignment caused by the expansion and contraction of hairpin-stem loops in ribosomal RNA molecules. Molecular Phylogenetics and Evolution 33:936-943.

Givnish, T. J. 1997. Adaptive radiation and molecular systematics: issues and approaches. Pages 1-54 *in* Molecular evolution and adaptive radiation (T. J. Givnish, and K. J. Sytsma, eds.). Cambridge University Press, Cambridge.

Goertzen, L. R., J. J. Cannone, R. R. Gutell, and R. K. Jansen. 2003. ITS secondary structure derived from comparative analysis: implications for sequence alignment and phylogeny of the Asteraceae. Molecular Phylogenetics and Evolution 29:216-234.

Goldman, D. H. 2001. Phylogenetics of Arethuseae (Orchidaceae) based on plastid *matK* and *rbcL* sequences. Systematic Botany 26:670-695.

Gottschling, M., H. H. Hilger, M. Wolf, and N. Diane. 2001. Secondary structure of the ITS1 transcript and its application in the reconstruction of the phylogeny of Boraginales. Plant Biology 3:629-636.

Gowri-Shankar, V., and M. Rattray. 2006. On the correlation between composition and site-specific evolutionary rate: implications for phylogenetic inference. Molecular Biology and Evolution 23:352-364.

Grimm, G. W., S. S. Renner, A. Stamatakis, and V. Hemleben. 2006. A nuclear ribosomal DNA phylogeny of *Acer* inferred with maximum likelihood, splits graphs, and motif analysis of 606 sequences. Evolutionary Bioinformatics Online 2:1-16.

Guerin, G. R., and R. S. Hill. 2006. Plant macrofossil evidence for the environment associated with the Riversleigh fauna. Australian Journal of Botany 54:717-731.

Guindon, S., and O. Gascuel. 2003a. A simple, fast, and accurate algorithm to estimate phylogenies by maximum likelihood. Systematic Biology 52:696-704.

Guindon, S., and O. Gascuel. 2003b. A simple, fast, and accurate algorithm to estimate phyogenies by maximum likelihood. Systematic Biology 52:696-704.

Guppy, H. B. 1906. Observations of a naturalist in the Pacific between 1896 and 1899. Vol 2 Plant -Dispersal. Macmillan and Co, London.

Guppy, H. B. 1917. Plants, seeds, and currents in the West Indies and Azores. Williams and Norgate, London.

Gutell, R., J. Cannone, Z. Shang, Y. Du, and M. Serra. 2000. A story: unpaired adenosine bases in ribosomal RNAs. Journal of Molecular Biology 304:335-354.

Gutell, R. R., J. C. Lee, and J. J. Cannone. 2002. The accuracy of RNA comparative structure models. Current Opinion in Structural Biology 12:301-310.

Hair, J. B., and E. J. Beuzenberg. 1959. Contributions to a chromosome atlas of the New Zealand flora-2. Miscellaneous families. New Zealand Journal of Science (Wellington) 2:148–156.

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

Han, K., and B. Yanga. 2003. PseudoViewer2: visualization of RNA pseudoknots of any type. Nucleic Acid Research 31:3432-3440.

Hardin, J. W. 1957. A revision of the American Hippocastanaceae. Brittonia 9:145-171.

Harrington, G. N., and M. A. Driver. 1995. The effect of fire and ants on the seed-bank of a shrub in a semi-arid grassland. Australian Journal of Ecology 20:538-547.

Harrington, M. G., E. Biffin, and P. A. Gadek. 2009. Comparative study of the evolution of nuclear ribosomal spacers incorporating secondary structure analyses within Dodonaeoideae, Hippocastanoideae and Xanthoceroideae (Sapindaceae). Molecular Phylogenetics and Evolution 50:364-375.

Harrington, M. G., E. Biffin, and P. A. Gadek. in review. The tempo of evolution in *Acer* (Sapindaceae) and *Aesculus* (Sapindaceae).

Harrington, M. G., K. J. Edwards, S. A. Johnson, M. W. Chase, and P. A. Gadek. 2005. Phylogenetic inference in Sapindaceae *sensu lato* using plastid *matK* and *rbcL* DNA sequences. Systematic Botany 30:366-382.

Harrington, M. G., and P. A. Gadek. 2004. Molecular systematics of the *Acmena* alliance (Myrtaceae): phylogenetic analyses and evolutionary implications with reference to Australian taxa. Australian Systematic Botany 17:63-72.

Hasebe, M., T. Ando, and K. Iwatsuki. 1998. Intrageneric relationships of maple trees based on the chloroplast DNA restriction fragment length polymorphisms. Journal of Plant Research 111:441-451.

Heimsch, C. J. 1942. Comparative anatomy of the secondary xylem in the "Gruinales" and "Terebinthales", of Wettstein with reference to taxonomic grouping. Lilloa 8:83-198.

Hendy, M. D., and D. Penny. 1989. A framework for the quantitative study of evolutionary trees. Systematic Zoology 38:297-309.

Hershkovitz, M. A., and E. A. Zimmer. 1996. Conservation patterns in angiosperm rDNA ITS2 sequences. Nucleic Acid Research 24:2857-2867.

Hershkovitz, M. A., E. A. Zimmer, and W. J. Hahn. 1999. Ribosomal DNA sequences and angiosperm systematics. Pages 268-326 *in* Molecular systematics and plant evolution (P. M. Hollingsworth, R. M. Bateman, and R. J. Gornall, eds.). Taylor and Francis, London.

Higgs, P. G. 2000. RNA secondary structure: physical and computational aspects. Quarterly Reviews of Biophysics 33:199–253.

Hillebrand, W. 1888. Flora of the Hawaiian Islands. Carl Winter, Heidelberg.

Hilu, K. W., and L. A. Alice. 1999. Evolutionary implications of *mat*K indels in Poaceae. American Journal of Botany 86:1735-1741.

Hirose, S., S. Tanda, L. Kiss, B. Grigaliunaite, M. Havrylenko, and S. Takamatsu. 2005. Molecular phylogeny and evolution of the maple powdery mildew (*Sawadaea*, Erysiphaceae) inferred from nuclear rDNA sequences. Mycological Research 8:912-922.

Hodges, S. A. 1997. A rapid adaptive radiation via a key innovation in *Aquilegia* Pages 391-405 *in* Molecular evolution and adaptive radiation (T. J. Givnish, and K. J. Sytsma, eds.). Cambridge University Press, Cambridge.

Hodgkinson, K. C. 1992. Water relations and growth of shrubs before and after fire in a semi-arid woodland. Oecologia 90:467-473.

Hodgkinson, K. C., and G. F. Griffin. 1982. Adaptation of shrub species to fires in the arid zone. Pages 145-152 *in* Evolution of the flora and fauna of arid Australia (W. R. Barker, and P. J. Greenslade, eds.). Peacock, Adelaide.

Hoot, S. B., J. W. Kadereit, F. R. Blattner, K. B. Jork, A. E. Schwarzbach, and P. R. Crane. 1997. Data congruence and phylogeny of the Papaveraceae s. l. based on four data sets: *atpB* and *rbcL* sequences, *trnK* restriction sites, and morphological characters. Systematic Botany 22:575-590.

Hudelot, C., V. Gowri-Shankar, H. Jow, M. Rattray, and P. G. Higgs. 2003. RNA-based phylogenetics methods: Application to mammalian mitochondrial RNA sequences. Molecular Phylogenetics and Evolution 28:241-252.

Huelsenbeck, J. P., and B. Rannala. 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. Systematic Biology 53:904-913.

Huelsenbeck, J. P., and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogeny. Bioinformatics 17:754-755.

Hughes, C., and R. Eastwood. 2006. Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. Proceedings of the National Academy of Sciences USA. 103:10334-10339.

Hughes, N. F., and A. B. McDougall. 1990. Barremian-Aptian angiospermid pollen records from southern England. Review of Palaeobotany and Palynology 65:145–151.

Hutchinson, J. 1926. The families of flowering plants. Macmillian & Co, London.
James, C. H., and C. Moritz. 2000. Intraspecific phylogeography in the sedge frog *Litoria fallax* (Hylidae) indicates pre-Pleistocene vicariance of an open forest species from eastern Australia. Molecular Ecology 9:349-385.
Johnson, L. A., and D. E. Soltis. 1995. Phylogenetic inference in Saxifragaceae sensu stricto and *Gilia* (Polemoniaceae) using *mat*K sequences. Annuals of the Missouri Botanic Gardens 82:149-175.
Joshi, A. C. 1938. Parthenocarpy in *Dodonaea viscosa*. Journal of the Indian Botanical Society 17:97-99.
Jow, H., C. Hudelot, M. Rattray, and P. G. Higgs. 2002. Bayesian phylogenetics using an RNA substitution model applied to early mammalian evolution. Molecular Biology and Evolution 19:1591-1601.
Judd, W. S., R. W. Sanders, and M. J. Donoghue. 1994. Angiosperm family pairs: preliminary phylogenetic analyses. Harvard Papers in Botany 5:1-51.
Kay, K. M., J. B. Whittall, and S. A. Hodges. 2006. A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effect. BMC Evolutionary Biology 6:36.
Keighery, G. J. 1982. Pollination syndromes and breeding systems. Pages 167-172 *in* Evolution of the flora and fauna of arid Australia (W. R. Barker, and P. J. Greenslade, eds.). Peacock, Adelaide.
Kellogg, E. A., and H. P. Linder. 1995. Phylogeny of Poales. Pages 511-542 *in* Monocotyledons: systematics and evolution (P. J. Rudall, P. J. Cribb, D. F. Cutler, and C. J. Humphries, eds.). Royal Botanic Gardens, Kew.
Keogh, S. J., I. A. Scott, M. Fitzgerald, and R. Shine. 2003. Molecular phylogeny of the Australian venomous snake genus *Hoplocephalus* (Serpentes, Elapidae) and conservation genetics of the threatened *H. stephensii*. Conservation Genetics 4:57-65.
Kerrigan, R. A., and D. E. Albrecht. 2007. Checklist of NT Vascular Plant Species. http://www.nt.gov.au/nreta/wildlife/plants/pdf/200701nt_checklist.pdf Northern Territory Government. Accessed 20th, February, 2008.
Khan, M. S. Y., K. Javed, and M. Hasnain Khan. 1992. Constituents of the flowers of *Dodonaea viscosa*. Fitoterapia 63:83-84.
Kjer, K. M. 2004. Aligned 18S insect phylogeny. Systematic Biology 53:506-514.
Klaassen, R. K. 1999. Wood anatomy of the Sapindaceae. IAWA Journal Supplement 2.
Klak, C., G. Reeves, and T. Hedderson. 2004. Unmatched tempo of evolution in Southern African semi-desert ice plants. Nature 427:63-65.
Kolaczkowski, B., and J. W. Thornton. 2004. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. Nature 431:980-984.
Kvaĉek, Z. 1994. Connecting links between the arctic Paleocene and European Tertiary floras Pages 251-266 *in* Cenozoic plants and climates of the arctic (M. C. Boulter, and H. C. Fisher, eds.). Springer-Verlag.
Langenheim, J. H. 2003. Plant resins: chemistry, evolution, ecology, and ethnobotany. Timber Press, Portland, Cambridge.
Lee-Chuan, C. Y., and J. C. Lee. 1992. Structure analysis of the 5' external transcribed spacer of the precursor ribosomal RNA from *Saccharomyces cerevisiae*. Journal of Molecular Biology 228:827-838.

Leenhouts, P. W. 1967. A conspectus of the genus *Allophylus* (Sapindaceae). Blumea 15:301-358.
Leenhouts, P. W. 1983. Notes on the extra-Australian species of *Dodonaea* (Sapindaceae). Blumea 28:271-289.
Leenhouts, P. W. 1985. An attempt towards a natural system of *Harpullia* (Sapindaceae). Blumea 31:219-234.
Leenhouts, P. W., and M. Vente. 1982. A taxonomic revision of *Harpullia* (Sapindaceae). Blumea 28:1-51.
Lewis, P. O. 2001. Phylogenetic systematics turns over a new leaf. Trends in Ecology and Evolution 16:30-37.
Li, J., J. Yue, and S. Shoup. 2006. Phylogenetics of *Acer* (Aceroideae, Sapindaceae) based on nucleotide sequences of two chloroplast non-coding regions. Harvard Papers in Botany 11:101-115.
Linder, H. P., C. R. Hardy, and F. Rutschmann. 2005. Taxon sampling effects in molecular clock dating: An example from the African Restionaceae. Molecular Phylogenetics and Evolution 35:569-582.
Lippold, H. 1978. Die gattung *Dodonaea* Miller (Sapindaceae) in Amerika. Bietr. Phytotaxon 6:79-126.
Liu, J.-S., and C. L. Schardl. 1994. A conserved sequence in internal transcribed spacer 1 of plant nuclear rRNA genes. Plant Molecular Biology 26:775-778.
Liu, J., and S. Noshiro. 2003. Lack of latitudinal trends in wood anatomy of *Dodonaea viscosa* (Sapindaceae), a species with a worldwide distribution. American Journal of Botany 90:532-539.
Losos, J. B., and D. B. Miles. 2002. Testing the hypothesis that a clade has adaptively radiated: Iguanid lizard clades as a case study. The American Naturalist 160:147-157.
Love, A. 1975. IOPB chromosome reports L. Taxon 24:671-678.
Love, A. 1976. IOPB chromosome reports LVI. Taxon 25:631-649.
Love, A. 1984. IOPB chromosome reports LXXXIV. Taxon 33:536-538.
MacGinitie, H. D. 1953. Fossil plants of the Florissant beds, Colorado. Carnegie Institute of Washington Publication 599:1-198.
Magallón, S., and M. J. Sanderson. 2001. Absolute diversification rates in angiosperm clades. Evolution 55:1762-1780.
Mai, J. C., and A. W. Coleman. 1997. The internal transcribed spacer 2 exhibits a common secondary structure in green algae and flowering plants. Journal of Molecular Evolution 44:258–271.
Manchester, S. R. 1999. Biogeographical relationships of North American Tertiary flora. Annals of the Missouri Botanical Garden 86:472-522.
Manchester, S. R. 2001. Leaves and fruits of *Aesculus* (Sapindales) from the Paleocene of North America. International Journal of Plant Sciences 162:985-998.
Manchester, S. R., and M. J. Donoghue. 1995. Winged Fruits of Linnaeeae (Caprifoliaceae) in the Tertiary of Western North America: *Diplodipelta* gen. nov. International Journal of Plant Sciences 156:709-722.
Manchester, S. R., and B. H. Tiffney. 2001. Intergration of paleobotanical and neobotanical data in the assessment of phytogeographic history of holarctic angiosperm clades. International Journal of Plant Sciences 162:S19-S28.

Maraschin-Silva, F., and M. E. Aqüila. 2005. Potencial alelopático de *Dodonaea viscosa* (L.) Jacq. Iheringia Série Botânica, Porto Alegre 60:91-98.
Markgraf, V., M. McGlone, and G. Hope. 1995. Neogene paleoenvironmental and paleoclimatic change in southern temperate ecosystems - a southern perspective. Trends in Ecology and Evolution 10.
Martin, H. A. 1994. Australian Tertiary phytogeography: evidence from palynology. Pages 104-142 *in* Australian vegetation history, Cretaceous to recent (R. S. Hill, ed.) Cambridge University Press, Cambridge.
Martin, H. A. 1997. The use of ecological tolerances for the reconstruction of Tertiary palaeoclimates. Australian Journal of Botany 45:475-492.
Mathews, D. H., M. D. Disney, J. L. Childs, S. J. Schroeder, M. Zuker, and D. H. Turner. 2004. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proceedings of the National Academy of Sciences USA. 101:7287-7292.
Mayol, M., and J. A. Rossello. 2001. Why nuclear ribosomal DNA spacers (ITS) tell different stories in *Quercus*. Molecular Phylogenetics and Evolution 19:167-176.
McClain, A. M., and S. R. Manchester. 2001. *Dipteronia* (Sapindaceae) from the Tertiary of North America and implications for the phytogeographic history of the Aceroideae. American Journal of Botany 88:1316-1325.
Metcalfe, C. R., and L. Chalk. 1950. Anatomy of the Dicotyledons. Clarendon Press, Oxford.
Meyer, F. G. 1977. *Sinoradlkofera*: a new genus of Sapindaceae. Journal of the Arnold Arboretum 58:182-189.
Miquel, F. A. 1844. Observations de plantis Novae Hollandiae et Novae Zeelandiae. Linnaea 18:83-95.
Mishio, M. 1992. Adaptations to drought in five woody species co-occurring on shallow- soil ridges. Australian Journal of Plant Physiology 19:539-553.
Moore, B. R., K. M. Chan, and M. J. Donoghue. 2004. Detecting diversification rate variation in supertrees. Pages 487-533 *in* Phylogenetic supertrees: combining information to reveal the tree of life (O. R. Bininda-Emonds, ed.) Kluwer Academic, Dordrecht.
Mueller, F. 1857. On some new genera of Australian plants. Hooker's Journal of Botany and Kew Gardens Miscellany 9:302-310.
Mueller, F. 1862. The plants indigenous to the colony of Victoria, I. Thalamiflorae. Government Printer, Melbourne.
Mueller, F. 1863a. Additamenta ad volumen tertium. Fragmenta Phytographiae Australiae 3:167.
Mueller, F. 1863b. Sapindaceae. Fragmenta Phytographiae Australiae 3:12-13.
Muller, J. 1985. Pollen morphology and evolution of the genus *Harpullia* (Sapindaceae - Harpulieae). Blumea 31:161-218.
Muller, J., and P. W. Leenhouts. 1976. A general survey of pollen types in Sapindaceae in relation to taxonomy. Pages 407-445 *in* The evolutionary significance of the exine (I. K. Ferguson, and J. Muller, eds.). Academic Press, London.
Muse, S. V. 1995. Evolutionary analysis of DNA sequences subject to constraints on secondary structure. Genetics 139:1429–1439.

Musters, W., K. Boon, F. M. van der Sande, H. van Heerikhuizen, and R. J. Planta. 1990. Functional analysis of transcribed spacers of yeast ribosomal DNA. EMBO Journal 9:3989-3996.

Nettel, A., and R. S. Dodd. 2007. Drifting propagules and receding swamps: genetic footprints of mangrove recolonization and dispersal along tropical coasts. Evolution 61:958-971.

Nowicke, J. 1976. Pollen morphology. In: F. G. Mayer, A revision of the genus *Koelreuteria* (Sapindaceae). Journal of the Arnold Arboretum 57:134-136.

Noyes, R. D. 2006. Intraspecific nuclear ribosomal DNA divergence and reticulation in sexual diploid *Erigeron strigosus* (Asteraceae). American Journal of Botany 93:470-479.

Nylander, J. A. 2004. MrAIC.pl. Program distributed by the author.

Nylander, J. A., F. Ronquist, J. P. Huelsenbeck, and J. L. Nieves-Aldrey. 2004. Bayesian phylogenetic analysis of combined data. Systematic Biology 53:47-67.

Oginuma, K., Y. Kuroki, H. Li, and K. Kondo. 1997. Karyomorphology of the dioecious plant *Dodonaea viscosa* (L.) Jacq. (Sapindaceae). Chromosome Science 1997:117--119.

Pagel, M., C. Venditti, and A. Meade. 2006. Large punctuational contribution of speciation to evolutionary divergence at the molecular level. Science 314:119-121.

Pearman, G. 2002. *Dodonaea viscosa* Jacq.. Record from Protabase. Oyen, L.P.A. & Lemmens, R.H.M.J. (Editors). PROTA (Plant Resources of Tropical Africa / Ressources végétales de l'Afrique tropicale), Wageningen, the Netherlands. http://database.prota.org/PROTAhtml/Dodonaea%20viscosa_En.htm. Accessed 02/01/08.

Pfosser, M. F., J. Guzy-wrobelska, B.-Y. Sun, T. F. Stuessy, T. Sugawara, and N. Fugii. 2002. The origin of species of *Acer* (Sapindaceae) endemic to Ullung Island, Korea. Systematic Botany 27:351-367.

Phartyal, S. S., J. M. Baskin, C. C. Baskin, and R. C. Thapliyal. 2005. Physical dormancy in seeds of *Dodonaea viscosa* (Sapindaceae) from India. Seed Science Research 15:59-61.

Plunkett, G. M., D. E. Soltis, and P. S. Soltis. 1997. Clarification of the relationship between Apiaceae and Araliaceae based on *mat*K and *rbc*L sequence data. American Journal of Botany 84:565-580.

Polanco, C., and M. Perez de La Vega. 1994. The structure of the rDNA intergenic spacer of *Avena sativa* L.: a comparative study. Plant Molecular Biology 25:751-756.

Posada, D., and T. R. Buckley. 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.

Prendergast, H. D., and G. Pearman. 2001. Comparing uses and collections - the example of *Dodonaea viscosa* Jacq. (Sapindaceae). Economic Botany 55:184-186.

Purvis, A., S. Nee, and P. H. Harvey. 1995. Macroevolutionary inferences from primate phylogeny. Proceedings of the Royal Society of London B Series 260:329-333.

Pybus, O. G., and P. H. Harvey. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. Proceedings of the Royal Society of London B Series 267:2267-2272.

Pybus, O. G., and A. Rambaut. 2002. GENIE: estimating demographic history from molecular phylogenies. Bioinformatics 18:1404-1405.

Radlkofer, L. 1890. Ueber die gliederung der familie der Sapindaceen. Sitzungsberichte der koniglich bayerischen Akademie der Wissenschaften 20:105-379.

Radlkofer, L. 1895. Sapindaceae. Pages 277-366, 460-462 *in* Die naturlichen Pflanzenfamilien, Vol. III, 5 (A. Engler, and K. Prantl, eds.). Engelmann, Berlin.

Radlkofer, L. 1900. Flora Brasiliensis 13:639-645.

Radlkofer, L. 1933. Sapindaceae. Pages 1-1539 *in* Das Pflanzenreich IV, 165 (Heft 98a-h) (A. Engler, ed.) Verlag von Wilhelm Engelmann, Leipzig.

Rambaut, A., and A. Drummond. 2004. TRACER v1.1 MCMC Trace Analysis Tool. University of Oxford.
Software available at: http://evolve.zoo.ox.ac.uk/software.html.

Rambaut, A. P., P. H. Harvey, and S. Nee. 1997. End-Epi: an application for reconstructing phylogenetic and population processes from molecular sequences. Computer Applications in the Biosciences 13:303-306.

Reddi, C. S., A. J. Bai, E. U. Reddi, and K. V. Raju. 1980. Pollen productivity, release and dispersal in *Dodonaea viscosa* (Linn.) Jacq. Proceedings of the Indian National Science Academy B46:184-190.

Rehder, A. 1935. *Handeliodendron*, a new genus of the Sapindaceae. Journal of the Arnold Arboretum 16:65-67.

Renner, S. S., L. Beenken, G. W. Grimm, A. Kocyan, and R. E. Ricklefs. 2007. The evolution of dioecy, heterodichogamy, and labile sex expression in *Acer*. Evolution 61:2701-2719.

Reynolds, S. T. 1984. Notes on Sapindaceae, III. Austrobaileya 2:29-64.

Richardson, J. E., R. T. Pennington, T. D. Pennington, and P. M. Hollingsworth. 2001. Rapid diversification of a species-rich genus of neotropical rain forest trees. Science 293:2242-2245.

Ridley, H. N. 1930. The dipersal of plants throughout the world. L. Reeve & Co., Ltd., Ashford, Kent.

Rivers, E. R. 1971. Variation in dioecism in cultivated *Dodonaea viscosa*. New Zealand Journal of Botany 9:549-554.

Ronquist, F., and J. P. Huelsenbeck. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

Ronquist, F., J. P. Huelsenbeck, and P. van der Mark. 2005. MrBayes 3.1Manual, Draft 5/26/2005, online at <http://mrbayes.csit.fsu.edu/manual.php/> (last accessed 10.01.07).

Ronse Decraene, L. P., E. F. Smets, and D. Clinckemaillie. 2000. Floral ontogeny and anatomy in *Koelreuteria* with special emphasis on monosymmetry and septal cavities. Plant Systematics and Evolution 223:91-107.

Sakai, A. K., W. L. Wagner, D. M. Ferguson, and D. R. Herbst. 1995. Origins of dioecy in the Hawaiian flora. Ecology 76:2517-2529.

Salamin, N., M. W. Chase, T. R. Hodkinson, and V. Savolainen. 2003. Assessing internal support with large phylogenetic DNA matrices. Molecular Phylogenetics and Evolution 27:528-539.

Sanders, K. L., and M. S. Lee. 2007. Evaluating molecular clock calibrations using Bayesian analyses with soft and hard bounds. Biology Letters 3:275-279.

Sanderson, M. J. 2003. Molecular data from 27 proteins do not support a Precambian origin of land plants. Ameican Journal of Botany 90:954-956.

Sanmartın, I., L. Wanntorp, and R. C. Winkworth. 2007. West Wind Drift revisited: testing for directional dispersal in the Southern Hemisphere using event-based tree fitting. Journal of Biogeography 34:398-416.

Savill, N. J., D. C. Hoyle, and P. G. Higgs. 2001. RNA sequence evolution with secondary structure constraints: comparison of substitution rate models using maximum likelihood methods. Genetics 157:399-411.

Savolainen, V., M. W. Chase, S. B. Hoot, C. M. Morton, D. E. Soltis, C. Bayer, M. F. Fay, A. Y. deBruijn, S. Sullivan, and Y.-L. Qiu. 2000a. Phylogenetics of flowering plants based on combined analysis of plastid *atp*B and *rbc*L gene sequences. Systematics Biology 49:306-362.

Savolainen, V., M. F. Fay, D. C. Albach, A. Backlund, M. van der Bank, K. M. Cameron, S. A. Johnson, M. D. Lledó, J.-C. Pintaud, M. Powell, M. C. Sheahan, D. E. Soltis, P. S. Soltis, P. Weston, W. M. Whitten, K. J. Wurdack, and M. W. Chase. 2000b. Phylogeny of the eudicots: a newly complete familial analysis based on *rbc*L gene sequences. Kew Bulletin 55:257-309.

Schatz, G. E. 1996. Malagasy/Indo-australo-malesian phytogeographic connections. Pages 73-83 *in* Biogéographie de Madagascar (W. R. Lourenço, ed.) Editions de l'ORSTOM, Paris

Schimper, A. F. W. 1891. Die indo-malaische strandflora. Botanische Mittheilungen aus den Tropen 3, Jena.

Schoniger, M., and A. von Haeseler. 1999. Toward assigning helical regions in alignments of ribosomal RNA and testing the appropriateness of evolutionary models. Journal of Molecular Evolution 49:691–698.

Schultz, J., S. Maisel, D. Gerlach, T. Müller, and M. Wolf. 2005. A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. RNA 11:361-364.

Shepherd, K. A., B. L. Rye, R. A. Meissner, and J. G. West. 2007. Two new Western Australian species of Dodonaea (Sapindaceae) from northern Yilgarn ironstones. Nuytsia 17:375-384.

Sherff, E. E. 1945. Some additions to the genus *Dodonaea* L. (Sapindaceae). American Journal of Botany 32:202-214.

Sherff, E. E. 1947. Further studies on the genus *Dodonaea* L. (family Sapindaceae). Field Museum Natural History Botanical Series 23:269-317.

Siddiqui, A. A. 1998. Chemical and Pharmacological Evaluation of Dodonaea Viscosa Linn. Asian Journal of Chemistry 10:14-16.

Smith, A. D., T. W. Lui, and E. R. Tillier. 2004. Empirical models for substitution in ribosomal RNA. Molecular Biology and Evolution 21:419-427.

Smith, A. G. 1985. Flora Vitiensis Nova Vol 3. Pacific Tropical Botanical Garden, Hawaii.

Soltis, D. E., R. K. Kuzoff, E. Conti, R. Gornall, and K. Ferguson. 1996. *mat*K and *rbc*L gene sequence data indicate that *Saxifraga* (Saxifragaceae) is polyphyletic. American Journal of Botany 83:371-382.

Soltis, D. E., and P. S. Soltis. 1998. Choosing an approach and appropriate gene for phylogenetic analysis. Pages 1-42 *in* Molecular systematics of plants II: DNA sequencing (D. E. Soltis, P. S. Soltis, and J. J. Doyle, eds.). Kluwer Academic Publishers, Boston.

Soltis, D. E., P. S. Soltis, M. W. Chase, M. E. Mort, D. C. Albach, M. Zanis, V. Savolainen, W. H. Hahn, S. B. Hoot, M. F. Fay, M. Axtell, S. M. Swensen, L. M. Prince, W. J. Kress, K. C. Nixon, and J. S. Farris. 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbc*L, and *atp*B sequences. Botanical Journal of the Linnean Society 133:381-461.

Soltis, D. E., P. S. Soltis, P. K. Endress, and M. W. Chase. 2005. Phylogeny and evolution of angiosperms. Sinauer Associates, Sunderland, Massachusetts.

Subbotin, S. A., D. Sturhan, N. Vovlas, P. Castillo, J. T. Tamber, M. Moens, and J. G. Baldwin. 2007. Application of the secondary structure model of rRNA for phylogeny: D2-D3 expansion segments of the LSU gene of plant-parasitic nematodes from the family Hoplolaimidae Filipjev, 1934. Molecular Phylogenetics and Evolution 43:881-890.

Suh, Y., K. Heo, and C.-W. Park. 2000. Phylogenetic relationships of maples (*Acer* L; Aceraceae) implied by nuclear ribosomal ITS sequences. Journal of Plant Research 113:193-202.

Sullivan, J., and P. Joyce. 2005. Model selection in phylogenetics. Annual Review of Ecology and Systematics 36:445-466.

Sun, Y., D. Z. Skinner, G. H. Liang, and S. H. Hulbert. 1994. Phylogenetic analysis of *Sorghum* and related taxa using internal transcribed spacers of nuclear ribosomal DNA. Theory of Applied Genetics 89:26-32.

Sutaria, R. N. 1930. Meiosis in *Dodonaea viscosa* L. Proceedings of the Indian National Congress Association 17:297-298.

Swofford, D. L. 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Sytsma, K. J., A. Litt, M. L. Zjhra, J. C. Pires, M. Nepokroeff, E. Conti, J. Walker, and P. G. Wilson. 2004. Clades, clocks, and continents: historical and biogeographical analysis of Myrtaceae, Vochysiaceae, and relatives in the Southern Hemisphere. International Journal of Plant Sciences 165:85-105.

Takhtajan, A. 1987. Systema Magnoliophytorum. Soviet Sciences Press, Leningrad.

Telford, M. J., M. J. Wise, and V. Gowri-Shankar. 2005. Consideration of RNA secondary structure significantly improves likelihood-based estimates of phylogeny: examples from the Bilateria. Molecular Biology and Evolution 22:1129-1136.

Thorne, R. F. 2000. The classification and geography of the flowering plants: dicotyledons of the class angiospermae. Botanical Review 66:441-647.

Tian, X., Z.-H. Guo, and D.-Z. Li. 2002. Phylogeny of Aceraceae based on ITS and trnL-F data sets. Acta Botanica Sinica 44:714-724.

Truswell, E. M., and W. K. Harris. 1982. The Cainozoic palaeobotanical record in arid Australia: fossil evidence for the origins of an arid-adapted flora. Pages 67-76 *in*

Evolution of the flora and fauna of arid Australia (W. R. Barker, and P. J. Greenslade, eds.). Peacock, Adelaide.
Turland, and Watson. 2004. (242-257) Report of the special committee on suprageneric names. Taxon 53:1081-1089.
Turner, H. 1995. Cladistic and biogeographic analyses of *Arytera* Blume and *Mischarytera* gen. nov. (Sapindaceae) with notes on methodology and a full taxonomic revision. Blumea Supplement No. 9:1-230.
Turner, S. R., D. J. Merritt, J. M. Baskin, C. C. Baskin, and K. W. Dixon. 2006. Combinational dormancy in seeds of the Western Australian endemic species *Diplopeltis huegelii* (Sapindaceae). Australian Journal of Botany 54:565-570.
Umadevi, I., and M. Daniel. 1991. Chemosystematics of the Sapindaceae. Feddes Repert. 102:607-612.
van der Ham, R. W. 1990. Nephelieae pollen (Sapindaceae): form, function, and evolution. Leiden Botanical Series 13:1-255.
van der Ham, R. W., and A. Tomlik. 1994. *Serjania* pollen and the origin of the tribe Paullinieae (Sapindaceae). Review of Palaeobotany and Palynology 83:43-53.
van Welzen, P. C. 1989. *Guioa* Cav. (Sapindaceae): Taxonomy, phylogeny and historical biogeography. Leiden Botanical Series 12:1-315.
Varani, G., and W. H. McClain. 2000. The G-U wobble base pair: a fundamental building block of RNA structure crucial to RNA function in diverse biological systems. EMBO Reports 1:18-23.
Venditti, C., A. Meade, and M. Pagel. 2006. Detecting the node-density artefact in phylogeny reconstruction. Systematic Biology 55:637-643.
Volkov, R. A., N. Y. Komarova, I. I. Panchuk, and V. Hemleben. 2003. Molecular evolution of rDNA external transcribed spacer and phylogeny of sect. Petota (genus *Solanum*). Molecular Phylogenetics and Evolution 29:187–202.
Wagner, W. L., D. R. Herbst, and S. H. Sohmer. 1990. Manual of the flowering plants of Hawai'i, Volume 2. University of Hawii Press, Honolulu.
Wakefield, N. A. 1971. The brush-tailed rock-wallaby (*Petrogale penicillata*) in western Victoria. Victorian Naturalist 88:92-102.
Webb, L. J., J. G. Tracey, and L. W. Jessup. 1986. Recent evidence for autochthony of Australian tropical and subtropical rainforest floristic elements. Telopea 2:575-589.
Webster, A. J., R. J. Payne, and M. Pagel. 2003. Molecular phylogenies link rates of evolution and speciation. Science 301:478.
Weckerle, C. S., and R. Rutishauser. 2003. Comparative morphology and systematic position of *Averrhoidium* within Sapindaceae. International Journal of Plant Sciences 164:775-793.
Wen, J. 1999. Evolution of eastern asian and eastern North American disjunct distributions in flowering plants. Annual Review of Ecology and Systematics 30:421-455.
West, J. G. 1980. A taxonomic revision of *Dodonaea* (Sapindaceae) in Australia. Ph.D. thesis University of Adelaide.
West, J. G. 1982a. Radiation and adaptation of *Dodonaea* (Sapindaceae) in arid Australia. Pages 329-333 *in* Evolution of the flora and fauna of arid Australia (W. R. Barker, and P. J. Greenslade, eds.). Peacock, Adelaide.

West, J. G. 1982b. Radiation and adaption of *Dodonaea* (Sapindaceae) in arid Australia. Pages 329-333 *in* Evolution of the flora and fauna of arid Australia (W. R. Barker, and P. J. Greenslade, eds.). Peacock, Adelaide.
West, J. G. 1984. A revision of *Dodonaea* Miller (Sapindaceae) in Australia. Brunonia 7:1-194.
West, J. G. 1985. *Dodonaea*. Pages 114-163 *in* Flora of Australia. Vol 25 (S. T. Reynolds, ed.) Australian Government Publishing Service, Canberra.
West, J. G. 1987. *Dodonaea biloba* (Sapindaceae), a new species from south-eastern Queensland, Australia. Brunonia 10:219-223.
West, J. G. 1993. *Hopbushes on the wind*. Australian Plants 17:225-226.
West, J. G. 1998. Floristics and Biodiversity Research in Australia: the 21st Century. Australian Systematic Botany 11:161-174.
West, J. G., and I. R. Noble. 1984. Analysis of digitised leaf images of the *Dodonaea viscosa* complex in Australia. Taxon 33:595-613.
Whelan, S., P. Liq, and N. Goldman. 2001. Molecular phylogenetics: state-of-the art methods for looking into the past. Trends in Genetics 17:262-272.
White, F. 1962. Geographic variation and speciation in Africa with particular reference to *Diospyros*. Pages 71-103 *in* Taxonomy and geography (D. Nichols, ed.) The Systematics Association, London.
White, T. J., T. Bruns, S. Lee, and J. Taylor. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pages 315-322 *in* PCR protocols: a guide to methods and applications (M. D. Innis, D. Gelfand, J. Sninsky, and T. White, eds.). Academic Press, San Diego.
Whitmore, T. C. 1976. Natural variation and its taxonomic treatment within tropical tree species as seen in the Far East *in* Tropical trees. Variation, breeding and conservation (J. Burley, and B. T. Styles, eds.). Academic Press, London.
Whittaker, R. J., M. B. Bush, and K. Richards. 1989. Plant recolonisation and vegetation succession on the Krakatau Islands. Ecological Monographs 59:59-123.
Whittle, C. A. 2006. The influence of environmental factors, the pollen:ovule ratio and seed bank persistence on molecular evolutionary rates in plants. Journal of Evolutionary Biology 19:302-308.
Whittle, C. A., and M. O. Johnston. 2003. Broad-scale analysis contradicts the theory that generation time affects molecular evolutionary rates in plants. Journal of Molecular Evolution 56:223-233.
Wickens, G. E. 1979. Speculations on seed dispersal and the flora of the Aldabra Archipelago. Philosophical Transactions of the Royal Society of London, Series B. 286:85-97.
Wikstrom, N., V. Savolainen, and M. W. Chase. 2001. Evolution of the angiosperms: calibrating the family tree. Proceedings of the Royal Society, London, Series B 268:2211-2220.
Wilcox, T. P., F. J. Garcia de Leon, D. A. Hendrickson, and D. M. Hillis. 2004. Convergence among cave catfishes: long-branch attraction and a Bayesian relative rates test. Molecular Phylogenetics and Evolution 31:1101-1113.
Williams, P., E. Collins, D. Mason, J. Prince, and G. Anchen. 2006. Variation in the age at first flowering for seedlings of 15 fire-killed shrubs and trees on sandstone

outcrops and sand plains in central and north-western Queensland. Ecological Management & Restoration 7:61-63.

Woinarski, J. C., K. Brennan, I. Cowie, A. Fisher, P. K. Latz, and J. Russell-Smith. 2000. Vegetation of the Wessel and English Company Islands, North-eastern Arnhem Land, Northern Territory, Australia. Australian Journal of Botany 48:115-141.

Wolf, M., J. Friedrich, T. Dandekar, and T. Muller. 2005. CBCAnalyzer: inferring phylogenies based on compensatory base changes in RNA secondary structures. Bioinformatics.

Wolfe, J. A. 1981. Vicariance biogeography of angiosperms in relation to paleobotanical data. Pages 413-445 *in* Vicariance biogeography: a critique (G. Nelson, and D. E. Rosen, eds.). Columbia University Press, New York.

Wolfe, J. A., and T. Tanai. 1987. Systematics, phylogeny and distribution of *Acer* (Maples) in the Cenozoic of western North America. Journal of the Faculty of Sciences Hokkaido University 22:1-246.

Wright, S., J. Keeling, and L. Gilman. 2006. The road from Santa Rosalia: a faster tempo of evolution in tropical climates. Proceedings of the National Academy of Sciences USA. 103:7718-7722.

Wright, S. D., C. G. Yong, S. R. Wichman, J. W. Dawson, and R. C. Gardner. 2001. Stepping stones to Hawaii: a trans-equatorial dispersal pathway for *Metrosideros* (Myrtaceae) inferred from nrDNA (ITS + ETS). Journal of Biogeography 28:769-774.

Xiang, Q.-Y., D. J. Crawford, A. D. Wolfe, Y. C. Tang, and C. W. DePamphilis. 1998a. Origin and biogeography of *Aesculus* L. (Hippocastanaceae): a molecular phylogenetic perspective. Evolution 52:988-997.

Xiang, Q.-Y., D. E. Soltis, and P. S. Soltis. 1998b. Phylogenetic relationships of Cornaceae and close relatives inferred from *matK* and *rbcL* sequences. American Journal of Botany 85:285-297.

Yang, Z., and B. Rannala. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Molecular Biology and Evolution 23:212–226.

Yule, G. U. 1924. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F. R. S. Philosophical Transactions of the Royal Society of London, Series B. 213:21-87.

APPENDIX 2.1

Vouchers/source for taxa used in analyses and GenBank accession numbers for the sequences. Where there is only one voucher the same DNA has been used for both *rbcL* and *matK*. Tribal assignments follow Muller and Leenhouts (1976): Cos – Cossinieae, Dod – Dodonaeeae, Dor – Doratoxyleae, Har – Harpullieae, Koel – Koelreuterieae, Mel – Melicocceae, Lep – Lepisantheae, Sap – Sapindeae, Cup – Cupanieae, Sch – Schleichereae, Nep – Nephelieae, Tho – Thouinieae, Pau – Paullinieae, Acer - Aceraceae. Samples beginning with lc are accession numbers from living collections. All species with GenBank accession number beginning with AY724 are from (Harrington, 2005 551), all other unreferenced species are new samples.

Species, Tribe, Voucher (citation), GenBank accession no., *rbcL, matK* .

Sapindaceae

Acer saccharum L., Acer, Chase 106 (NCU) (Chase, 1993), L01881, AY724265, *Acer palmatum* Thunb ex A.E.Murray, Acer, , , , *Aesculus chinensis* Bunge, Har?, Xiang 305 (OS) (Xiang, 1998), -, AY724267 , *Aesculus hippocastanum* L., Har?, Kew living collection 69.11289-263 (Xiang, 1998), -, AY724266 , *Aesculus pavia* Castigl., Har?, Chase 503 (K) (Gadek, 1996), U39277, - , *Alectryon connatus* (F. Muell.) Radlk., Nep, Chase 2047 (K); Edwards KE79 (JCT) (Edwards, 2001), AY724341, AF314788 , *Alectryon coriaceus* (Benth.) Radlk., Nep, Edwards KE59 (JCT) (Edwards, 2001), -, AF314783 , *Alectryon excelsus* ssp. *excelsus* Gaert., Nep, Edwards KE254 (JCT) (Edwards, 2001), , AF314797 , *Alecrtyon excelsus* ssp. *grandis* (Cheeseman) de Lange et E. K. Cameron, Nep, AKU 1983 083 (JCT) (Edwards, 2001), , AF314796 , *Allophylus cobbe* (L.) Blume, Tho, Williams 44A (CBG) , AY724342, - , *Allophylus javensis* Blume, Tho, Chase 2121 (K) , AY724343, - , *Allophylus natalensis* (Sond.) DeWinter, Tho, Edwards KE227 (JCT) , -, AY724268 , *Allophylus* sp. , Tho, Edwards KE273 (JCT) , -, AY724269 , *Arfeuillea arborescens* Pierre, Har, Chase 2122 (K), , , *Arytera divaricata* F.Muell., Cup, Edwards KE010 (JCT) , -, AY724271 , *Arytera litoralis* Blume, Cup, Chase 2123 (K) , AY724344, - , *Arytera microphylla* (Benth.) Radlk., Cup, Edwards KE60 (JCT) , -, AY724270 , *Atalaya alata* (Sim) H.Forbes, Sap, Chase 1126 (K); Edwards KE228 (JCT) , AY724345, AY724274 , *Atalaya angustifolia* S.Reyn., Sap, West 5349 (ANH) , -, AY724273 , *Atalaya capensis,* Sap, , , , *Atalaya salicifolia* (A.DC.) Blume, Sap, Edwards KE58 (JCT) , -, AY724272 , *Averrhoidium dalyi* Acev.-Rodr. & Ferrucci, Dor, , , , *Billia hippocastanum* Peyr., Har?, Pennington & Zamora 604 (K) (Savolainen, 2000), AJ402929, - , *Billia* sp. , Har?, Hammel 20075 (OS) (Xiang, 1998), -, AY724275 , *Blighia sapida* Koenig, Cup, Chase 2124 (K); Edwards KE86 (JCT) , AY724346, AY724277 , *Blighia unijugata* Baker, Cup, Edwards KE274 (JCT) , -, AY724276 , *Bridgesia incisifolia* Bert. Ex Cambess., Tho, Killip & Pisano 39778 (K) , AY724347, - , *Cardiospermum grandiflorum* Sw., Pau, Chase 2869 (K); Edwards KE207 (JCT) , AY724348, AY724278 , *Castanospora alphandii* (F.Muell.) F.Muell., Mel, Edwards KE88 (JCT) , -, AY724279 , *Chytranthus prieurianus* Baill., Lep, Edwards KE272 (JCT) , -, AY724280 , *Conchopetalum brachysepalum* Capuron, Har, Radenantoandro 674 (MO), , , *Cossina pinnata* Commerson ex Lam, Cos, lc950595(NTBG), , , *Cubilia cubili* (Blanco) Adelb., Nep, Chase 2125 (K) , -, AY724281 , *Cupaniopsis anacardiodes* (A.Rich.) Radlk., Cup,

Chase 217 (K) (Chase et al. 1993); Edwards KE47 (JCT) , AF035903, AY724283 , *Cupaniopsis flagelliformis* (Bailey) Radlk. v. *flagelliformis,* Cup, Edwards KE42 (JCT) , -, AY724282 , *Deinbollia borbonica* Scheff.*,* Sap, Edwards KE197 (JCT) , -, AY724284 , *Deinbollia oblongifolia* (E.Mey. ex Arn.) Radlk.*,* Sap, Edwards KE233 (JCT) , -, AY724285 , *Delavaya yunnanensis* Franch*,* Har, Chase 3831, , , *Diatenopteryx sorbifolia* Radlk.*,* Tho, Tressens et al. 3504 (K) (Savolainen, 2000) , AJ402943, - , *Dimocarpus australianus* Leenh.*,* Nep, Edwards KE34 (JCT) (Edwards, 2001) , -, AF314799 , *Dimocarpus longan* (Lour.) Fl. Coch. Leenh.*,* Nep, Chase 1351 (K); Edwards KE502 (JCT) , AY724349, AY724286 , *Diplokeleba floribunda* N. E. Br., Cup, Acevedo 11130 (US), , , *Diploglottis campbelli* Cheel*,* Cup, Chase 2048 (K), AY724350, - , *Diploglottis diphyllostegia* (F.Muell.) Bailey, Cup, Edwards KE001 (JCT) , -, AY724287 , *Diploglottis smithii* S.Reyn.*,* Cup, Gray BG838 (CSIRO) , -, AY724288 , *Diplopeltis huegelii* Endl., Dod, Chase 2192 (K) (Savolainen, 2000 263), AJ402944, - , *Diplopeltis stuartii* var. *stuartii* F.Muell., Dod, Kendrick, P. s.n. (CANB), , , *Dipteronia sinensis* Oliv.*,* Acer, Chase 502 (K) (Gadek, 1996) , U39268, AY724289 , *Distichostemon hispidulus* (Endl.) Baillon v. *hispidulus ,* Dod, Purdie 3405 (CBG) (Edwards, 2001), -, AF314804 , *Dodonaea lanceolata* v. *subsessifolia* F.Muell.*,* Dod, Edwards KE120 (JCT) , -, AY724290 , *Dodonaea triquetra* Wendl., Dod, Adam 21164 (UNSW) (Gadek, 1996), U38922, - , *Dodonaea viscosa* subsp. *angustifolia* (L.f.) J.West *,* Dod, Edwards KE59 (JCT) (Edwards, 2001), -, AF314803 , *Dodonaea viscosa* New Zealand*,* Dod, , , , *Elattostachys megalantha* S.Reyn.*,* Cup, Irvine IRV507 (CSIRO) , -, AY724291 , *Elattostachys microcarpa* S.Reyn.*,* Cup, Edwards KE98 (JCT) , -, AY724292 , *Erioglossum rubiginosum* (Roxb.) Blume = *Lepisanthes rubiginosa* (Roxb.) Leenh.*,* Lep, Chase 1350 (K) , AY724351, AY724293 , *Erythrophysa aesculina* Baill., Koel, Randrianasolo 625 (MO), , , *Erythrophysa transvaalensis*, Koel, , , , *Exothea diphylla* (Standl.) Lundell*,* Dor, Acevedo 12233 (US) , , , *Exothea paniculata* (Juss) Radlk.*,* Dor, Acevedo 12177 (US), , , *Filicium decipiens* (Wight & Arn.) Thwaites*,* Dor, Chase 2128 (K); Edwards KE271 (JCT) , AY724352, AY724294 , *Filicum longifolium* (H.Perrier) Caparon*,* Dor, Rabenantroandro 1113 (MO), , , *Ganophyllum falcatum* Blume*,* Dor, Chase 2129 (K); Hyland BH9269 (CSIRO) , AY724353, AY724295 , *Guindilia trinervis* Gilles ex Hook. *,* Tho, Chase 802 (K) , AY724354, AY724296 , *Guioa acutifolia* Radlk.*,* Cup, Edwards KE14 (JCT) , -, AY724297 , *Guioa lasioneura* Radlk.*,* Cup, Gray BG1888 (CSIRO) , -, AY724298 , *Guioa semiglauca* (F.Muell.) Radlk.*,* Cup, Chase 2058 (K), AY724355, - , *Handeliodendron bodinieri* Levl. Rehd.*,* Har, QYXiang 302 (OS) (Xiang et al. 1998a), - , AY724299 , *Haplocoelopsis africana* F.G.Davies*,* Cup, Edwards KE276 (JCT) , -, AY724300 , *Haplocoelum foliolosum* subsp. *foliolosum* (Hiern) Bullock*,* Sch, Edwards KE195 (JCT) , -, AY724301 , *Haplocoelum gallaense* (Engl.) Radlk.*,* Sch, Edwards KE501 (JCT) , -, AY724302 , *Harpullia arborea* (Blanco) Radlk.*,* Har, Edwards KE38 (JCT) , AY724356, - , *Harpullia ramiflora* (Radlk.)*,* Har, Edwards KE6 (JCT) (Edwards and Gadek 2001), -, AF314805 , *Harpullia rhyticarpa* C.White & Francis*,* Har, Edwards KE003 (JCT) , -, AY724303 , *Hebecoccus ferrugineus* Radlk. = *Lepisanthes ferruginea* (Radlk.) Leenh.*,* Lep, Chase 1354 (K) , AY724357, AY724304 , *Hippobromus pauciflora* (L.f.) Radlk.*,* Dor, Edwards KE229 (JCT), -, AY724305 , *Hornea mauritana,* Sap, lc760230(NTBG), , , *Hypelate trifoliata* Griseb.*,* Dor, Rankin & Arias 72057 (K), AY724358, - , *Jagera javanica* subsp. *australiana* Leenh.*,* Cup, Edwards KE178 (JCT),

-, AY724307 , *Jagera javanica* (Blume) Blume ex Kalkman subsp. *javanica* , Cup, Chase 2130 (K), AY724359, - , *Jagera pseudorhus* var. *pseudorhus* f. *pilosiuscula* Radlk., Cup, Edwards KE41 (JCT), -, AY724306 , *Koelreuteria paniculata* Laxm., Koel, Chase 115 (K) (Gadek, 1996 509); Wilson 1476 (RBG), U39283, AY724308 , *Koetreuteria elegans (Seem.) A.C.Sm subsp. formosana,* Koel, , , , *Llagunoa mollis,* Cos, Jaramillollejia et al. 3199 (K), , , *Llagunoa nitida* Ruiz & Pav., Cos, Acevedo 11130 (US), , , *Lecaniodiscus fraxinifolius* Baker, Sch, Edwards KE194 (JCT), -, AY724309 , *Lepiderema hirsuta* S.Reyn., Cup, Edwards KE36 (JCT), -, AY724310 , *Lepidopetalum xylocarpum* Radlk., Cup, Edwards KE139 (JCT), -, AY724311 , *Lepisanthes alata* (Blume) Leenh., Lep, Chase 1355 (K), AY724360, AY724312 , *Litchi chinensis* Sonn., Nep, Chase 2131 (K); Edwards KE212 (JCT) (Edwards and Gadek 2001), AY724361, AF314800 , *Loxodiscus coriaceus,* Har, Biffin EAB03/142 (CANB), , , *Magonia pubescens* St. Hil., Har, , , , *Majidea fosteri* Radlk., Har, Living plant collection 19695474 (NBGB), , , *Matayba* sp., Cup, Chase 2132 (K) , AY724362, - , *Mischarytera macrobotrys* (Merr. & Perry) H. Turner, Cup, Hyland BH6631 (CSIRO), -, AY724313 , *Mischocarpus exangulatus* (F.Muell.) Radlk., Cup, Edwards KE30 (JCT), -, AY724314 , *Mischocarpus grandissimus* (F.Muell.) Radlk., Cup, Edwards KE37 (JCT), -, AY724315 , *Mischocarpus pyriformis* (F.Muell.) Radlk., Cup, Chase 2059 (K), AY724363, - , *Nephelium lappaceum* L., Nep, Edwards KE222 (JCT) (Edwards and Gadek 2001), -, AF314801 , *Nephelium mutabile* Blume , Nep, Chase 2134 (K), AY724364, AY724316 , *Pancovia golungensis* (Hiern.) Excell & Mendonca, Lep, Edwards KE231 (JCT), -, AY724317, *Pappea capensis* Eckl. & Zeyh., Nep, Edwards KE232 (JCT) (Edwards and Gadek 2001), -, AF314798 , *Paranephelium macrophyllum* King, Cup, Chase 1356 (K) (Savolainen, 2000) , AJ403032, AY724318 , *Paranephelium xestophyllum* Miq., Cup, Edwards KE503 (JCT), -, AY724319 , *Paullinia pinnata* L., Pau, Edwards KE199 (JCT), -, AY724320 , *Paullinia venosa* L. Radlk., Pau, Chase 3312 (K), AY724365, - , *Pometia tomentosa* (Blume) Teijsm. & Binn. = *Pometia pinnata* Forst. & Forst., Nep, Chase 2135 (K) (Edwards and Gadek 2001), -, AF314802 , *Rhysotoechia mortoniana* (F.Muell.) Radlk., Cup, Edwards KE117 (JCT), -, AY724321 , *Rhysotoechia robertsonii* (F.Muell.) Radlk., Cup, Edwards KE277 (JCT), -, AY724322 , *Sapindus saponaria* L. , Sap, Chase 2136 (K), AY724366, AY724324 , *Sapindus trifoliatus* L., Sap, Edwards KE504 (JCT), -, AY724323 , *Sarcopteryx martyana* (F.Muell.) Radlk., Cup, Irvine IRV1810 (CSIRO), -, AY724326 , *Sarcopteryx reticulata* S.Reyn., Cup, Gray BG1137 (CSIRO), -, AY724325 , *Sarcotoechia serrata* S.Reyn., Cup, Edwards KE31 (JCT), -, AY724327 , *Sarcotoechia villosa* S.Reyn., Cup, Edwards KE102 (JCT), -, AY724328 , *Schleichera oleosa* (Lour.) Merr., Sch, Chase 2137 (K), AY724367, AY724329 , *Serjania communis* Cambess., Pau, Chase 2138 (K) (Savolainen, 2000), AJ403001, - , *Sinoradlkofera minor* (Hemsley) F.G.Meyer, Koel, , , , *Smelophyllum capense* Radlk., Nep, Edwards KE506 (JCT), -, AY724330 , *Stadmania oppositifolia* (Lam.) Poir., Nep, Edwards KE505 (JCT), -, AY724331 , *Synima cordieri* (F.Muell.) Radlk., Cup, Edwards KE29 (JCT), -, AY724333 , *Synima macrophylla* S.Reyn., Cup, Edwards KE19 (JCT), -, AY724332 , *Talisia nervosa* Radlk., Mel, Pennington 628 (K) (Savolainen, 2000), AJ403008, - , *Toechima daemelianum* (F.Muell.) Radlk., Cup, Clarkson JC66 (CSIRO), -, AY724334 , *Toechima erythrocarpum* (F.Muell.) Radlk., Cup, Edwards KE20 (JCT), -, AY724335 , *Toechima tenax* (Cunn. ex Benth.) Radlk., Cup, Chase 2046 (K), AY724368, - ,

Tristira triptera (Blanco) Radlk., Mel, Chase 2139 (K), AY724369, AY724336 , *Tristiropsis acutangula* Radlk., Mel, Chase 1358 (K), AY724370, AY724337 , *Ungnadia speciosa* Endl., Har, Chase 2854 (K) (Savolainen, 2000) , AJ403014, AY724338 , *Xanthoceras sorbifolium* Bunge, Har, Chase 2866 (K) , AJ403019, AY724339 , *Zanha africana* (Radlk.) Exell, Dor, , ,

Outgroups

Anacardiaceae *Buchanania latifolia* Roxb., , Terrazas 206 (CHAPA) (Gadek et al. 1996), U39275, , *Pleiogynium timorense* (DC.) Leenh., , Edwards KE50 (JCT), -, AY724340. Biebersteinaceae *Biebersteinia orphanidis,* , , , , Burseraceae *Bursera fagarioides,* , , , , *Commiphora habessinica,* , , , , Kirkiaceae *Kirkia wilmsii,* , , , , Meliaceae *Aglaia elliptica,* , , , , *Azadirachta indica,* , , , , *Cedrela odorata,* , , , , Nitrariaceae *Nitraria retusa,* , , , , *Peganum harmala* L., , Collenette 7/93 (K) (Gadek et al. 1996), U39279, AY177667 Rutaceae *Acronychia acidula,* , , , , *Flindersia australis,* , , , , Simbaroubaceae *Quassia amara,* , , , , *Simarouba glauca,* , , , , Malvales *Bixa orellana L.,* , , AF022128, , *Bombax buonopozense* P. Beaeu,, , AF022118, ,

Chase, M. W., D. E. Soltis, R. G. Olmstead, D. Morgan, D. H. Les, B. D. Mishler, M. R. Duvall, R. A. Price, H. G. Hills, Y. L. Qiu, K. A. Kron, J. H. Rettig, E. Conti, J. D. Palmer, J. R. Manhart, K. J. Sytsma, H. J. Michaels, W. J. Kress, K. G. Karol, W. D. Clark, M. Hedrén, B. S. Gaut, R. K. Jansen, K. J. Kim, C. F. Wimpee, J. F. Smith, G. R. Furnier, S. H. Strauss, Q. Y. Xiang, G. M. Plunkett, P. S. Soltis, S. M. Swensen, S. E. Williams, P. A. Gadek, C. J. Quinn, L. E. Eguiarte, E. Golenberg, G. H. Learn, S. W. Graham, S. C. Barrett, S. Dayanandan, and V. A. Albert. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. Annuals of Missouri Botanical Gardens 80:528-580.

Edwards, K. J., and P. A. Gadek. 2001. Evolution and biogeography of *Alectryon* (Sapindaceae). Molecular Phylogenetics and Evolution 20:14-26.

Gadek, P. A., E. S. Fernando, C. J. Quinn, S. B. Hoot, T. Terrazas, M. C. Sheahan, and M. W. Chase. 1996. Sapindales: Molecular delimitation and infraordinal groups. American Journal of Botany 83:802-811.

Harrington, M. G., K. J. Edwards, S. A. Johnson, M. W. Chase, and P. A. Gadek. 2005. Phylogenetic inference in Sapindaceae *sensu lato* using plastid *matK* and *rbcL* DNA sequences. Systematic Botany 30:366-382.

Savolainen, V., M. F. Fay, D. C. Albach, A. Backlund, M. van der Bank, K. M. Cameron, S. A. Johnson, M. D. Lledó, J.-C. Pintaud, M. Powell, M. C. Sheahan, D. E. Soltis, P. S. Soltis, P. Weston, W. M. Whitten, K. J. Wurdack, and M. W. Chase. 2000. Phylogeny of the eudicots: a newly complete familial analysis based on *rbc*L gene sequences. Kew Bulletin 55:257-309.

Xiang, Q.-Y., D. J. Crawford, A. D. Wolfe, Y. C. Tang, and C. W. DePamphilis. 1998. Origin and biogeography of *Aesculus* L. (Hippocastanaceae): a molecular phylogenetic perspective. Evolution 52:988-997.

APPENDIX 3.1

Vouchers/references for additional taxa included in this study that are not in appendix 2.1 and GenBank accession numbers for the sequences. Where there is only one voucher/reference the same DNA has been used for both *rbcL* and *matK*. A dash indicates that the region was not sampled.

Taxon, voucher or reference; Genbank accession numbers *rbcL*, *matK*.

Acer acuminatum Wall. ex D.Don, Renner et al. (2007), DQ978393, —. *Acer barbinerve* Maxim., Renner et al. (2007), DQ978395, —. *Acer buergerianum* Miq., Renner et al. (2007), DQ978396, —. *Acer campestre* L., Renner et al. (2007), DQ978399, —. *Acer carpinifolium* Siebold & Zucc., Renner et al. (2007), DQ978400, —. *Acer caudatum* Wall., Renner et al. (2007), DQ978401, —. *Acer circinatum* Pursh, Renner et al. (2007), DQ978403, —. *Acer davidii* Franch., Renner et al. (2007), DQ978406, —. *Acer diabolicum* Blume ex Koch, Renner et al. (2007), DQ978407, —. *Acer distylum* Siebold & Zucc., Renner et al. (2007), DQ978408, —. *Acer erianthum* Schwer., Renner et al. (2007), DQ978409, —. *Acer fabri* Hance, Renner et al. (2007), EF186772, —. *Acer kweilinense* Fang & Fang f., Renner et al. (2007), EF186773, —. *Acer laevigatum* Hu & Cheng, Renner et al. (2007), DQ978412, —. *Acer laurinum* Hassk., Renner et al. (2007), DQ978413, —. *Acer macrophyllum* Pursh, Renner et al. (2007), DQ978414, —. *Acer mono* Maxim., Renner et al. (2007), DQ978416, —. *Acer negundo* L., Renner et al. (2007), DQ978417, —. *Acer nigrum* Michx.f., Renner et al. (2007), DQ978431, —. *Acer nipponicum* Hara, Renner et al. (2007), DQ978418, —. *Acer oblongum* Wall. ex DC., Renner et al. (2007), DQ978419, —. *Acer pilosum* Maxim., Renner et al. (2007), DQ978423, —. *Acer pseudoplatanus* L., Renner et al. (2007), DQ978425, —. *Acer pycnanthum* K.Koch, Renner et al. (2007), DQ978427, —. *Acer rubrum* L., Renner et al. (2007), DQ978428, —. *Acer spicatum* Lamarck, Renner et al. (2007), DQ978434, —. *Acer stachyophyllum* Hiern, Renner et al. (2007), DQ987822, —. *Acer sterculiaceum* Wall., Renner et al. (2007), DQ978435, —. *Acer tataricum* L., Renner et al. (2007), DQ978436, —. *Acer trautvetteri* Medvedev, Renner et al. (2007), DQ978438, —. *Acmena smithii* (Poir.) Merr. & L.M.Perry, Conti et al. (1996), Biffin et al. (2006), U26315, DQ088545. *Aesculus glabra* Willd., Modliszewski et al. (2006), —, AY968671. *Aesculus flava* Ait., Renner et al. (2007), Modliszewski et al. (2006), DQ978441, AY968670, AY724266. *Aesculus parviflora* Walt., Renner et al. (2007), Modliszewski et al. (2006), DQ978448, AY968629. *Aesculus sylvatica* Bartr., Modliszewski et al. (2006), —, AY968666. *Aesculus turbinata* Blume, Modliszewski et al. (2006), —, AY968633. *Ailanthus altissima* (Mill.) Swingle, Muellner et al. (2003), AY128247, AY128208. *Alangium kurzii* Craib, unpublished, DQ340449, DQ341347. *Anacardium occidentale* L., Aguilar-Ortigoza et al. (2004), AY462008, —. *Asparagus capensis* L., Forest et al. (2007), AM234843, —. *Asparagus filicinus* Buch.-Ham. ex D.Don., Tamura (2000), — , AB029805. *Azadirachta indica* L., Muellner et al. (2003), AY128214, —. *Camptotheca acuminata* Decne., Xiang et al. (1993), Xiang et al. (1998), L11211,U96888. *Ceratophyllum demersum* L., Les et al. (1991), unpublished, M77030, AF543732. *Ceratophyllum submersum* L., Qiu et al. (1999), Hilu et al (2003), AF197599, AJ581400. *Cornus hongkongensis* Hemsl., unpublished, DQ340447, DQ341353. *Grevillea banksii* R. Br., Hilu et al. (2003), — , AF542583. *Grevillea robusta* A.Cunn. ex R.Br., unpublished, AF193973, — . *Metrosideros nervulosa* C.Moore & F.Muell., Savolainen et al. (2000), Biffin et al. (2006), AJ403028, DQ088535. *Papaver sp.* Goldblatt

12541, Forest et al. (2007), AM235045, ─. *Papaver triniifolium* Boiss., unpublished, ─ , AM396511. *Personia lanceolata* Michaux, unpublished, U79178, ─. *Persoonia katerae* P.H.Weston & L.A.S.Johnson, ─, AY437813. *Rhynchocalyx lawsonioides* Oliv., Conti et al. (1996), Wilson et al. (2001), U26336, AF368218. *Sanguinaria canadensis* L., Albert et al (1992), unpublished, L01951, DQ401350. *Schinus molle* L., Gadek et al. (1996), U39270, ─. *Schinus* sp. Chase 171, ─, AY491645. *Tetracentron sinense* Oliv., Qia et al. (1993), unpublished, L12668, AM396504. *Toona* sp. Chase 664, Muellner et al. (2003), AY128243, AY128201. *Trochodendron aralioides* Siebold & Zuccarini, Albert et al. (1992), unpublished, L01958, AF543751. *Tulipa turkestanica* Hort.van Tuberg., unpublished, AB037378, AB024386. *Turraea sericea* Sm., Muellner et al. (2003), AY128245, AY128203. *Vochysia hondurensis* Sprague, Conti et al. (1996), Sytsma et al. (2004), U26340, AY572446. *Walsura tubulata* Hiern., Muellner et al. (2003), AY128246, AY128204.

Aguilar-Ortigoza, C. J. and V. Sosa. 2004. The evolution of toxic phenolic compounds in a group of Anacardiaceae genera. Taxon 53:357-364.

Albert, V. A., S. E. Williams, and M. W. Chase. 1992. Carnivorous plants: phylogeny and structural evolution. Science 257:1491-1495.

Biffin, E., L. A. Craven, M. D. Crisp, P. A. Gadek. 2006. Molecular systematics of *Syzygium* and allied genera (Myrtaceae): evidence from the chloroplast genome. Taxon 55:79-94.

Conti, E., A. Litt, and K. J. Sytsma. 1996. Circumscription of Myrtales and their relationships to other rosids: Evidence from *rbc*L sequence data. American Journal of Botany 83:221-233.

Forest, F., R. Grenyer, M. Rouget, T. J. Davies, R. M. Cowling, D. P Faith, A. Balmford, J. C. Manning, S. Proches, M. van der Bank, G. Reeves, T. A. Hedderson, and V. Savolainen. 2007. Preserving the evolutionary potential of floras in biodiversity hotspots. Nature 445:757-760.

Gadek, P. A., E. S. Fernando, C. J. Quinn, S. B. Hoot, T. Terrazas, M. C. Sheahan, and M. W. Chase. 1996. Sapindales: Molecular delimitation and infraordinal groups. American Journal of Botany 83:802─811.

Hilu, K. W., T. Borsch, K. Muller, D. E. Soltis, P. S., Soltis, V.Savolainen, M. W. Chase, M. P. Powell, L. A. Alice, R. Evans, H .Sauquet, C. Neinhuis, T. A. Slotta, J. G. Rohwer, C. S. Campbell, and L. W. Chatrou. 2003. Angiosperm phylogeny based on *mat*K sequence information. American Journal of Botany 90:1758-1776.

Les, D. H., D. K. Garvin, and C. F. Wimpee. 1991. Molecular evolutionary history of ancient aquatic angiosperms. Proceeding of the National Academy of Science of USA 88:10119-10123.

Modliszewski, J. L., Thomas, D. T., Fan, C., Crawford, D. J., dePamphilis, C. W. and Xiang, Q.-Y. (2006) Ancestral chloroplast polymorphism and historical secondary contact in a broad hybrid zone of *Aesculus* (Sapindaceae). American Journal of Botany 93: 377─388

Muellner, A. N., R. Samuel, S. A. Johnson, M. Cheek, T. D. Pennington, and M. W. Chase. 2003. Molecular phylogenetics of Meliaceae (Sapindales) based on nuclear and plastid DNA sequences. American Journal of Botany 90:471-480.

Qiu, Y. -L., M. W. Chase, D. H. Les, and C. R. Parks. 1993. Molecular phylogenetics of the Magnoliidae: cladistic analysis of nucleotide sequences of the plastid gene *rbc*L. Annuals of Missouri Botanical Gardens 80:587-606.

Qiu,Y. L., J. Lee, F. Bernasconi-Quadroni, D. E. Soltis, P. S. Solis, M. Zanis, E. A. Zimmer, Z. Chen, V. Savolainen, and M. W. Chase. 1999. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. Nature 402:404-407.

Renner, S. S., Beenken, L., Grimm, G., Kocyan, A. and Ricklefs, R. E. (2007) The evolution of dioecy, heterodichogamy, and labile sex expression in *Acer*. Evolution 61:2701-2719.

Savolainen, V., M. F. Fay, D. C. Albach, A. Backlund, M. van der Bank, K. M. Cameron, S. A. Johnson, M. D. Lledó, J.-C. Pintaud, M. Powell, M. C. Sheahan, D. E. Soltis, P. S. Soltis, P. Weston, W. M. Whitten, K. J. Wurdack, and M. W. Chase. 2000. Phylogeny of the eudicots: a newly complete familial analysis based on *rbc*L gene sequences. Kew Bulletin 55:257-309.

Tamura, M. N. 2000. Molecular phylogeny of the Convallariaceae (Asparagales). Pages 387-400 *in* K. L. Wilson and D. A. Morrison eds. Monocots: systematics and evolution. CSIRO, Melbourne.

Wilson, P. G., M. M. O'Brien, P. A. Gadek, and C. J. Quinn. 2001. Myrtaceae revisited: a reassessment of intrafamilial groups. American Journal of Botany 88:2013-2025.
Xiang, Q. -Y., D. E. Soltis, D. R. Morgan, and P. S. Soltis. 1993. Phylogenetic relationships of *Cornus* L. sensu lato and putative relatives inferred from *rbc*L sequence data. Annals of Missouri Botanical Gardens 80:723-734.
Xiang, Q. -Y., D. E. Soltis, and P. S. Soltis. 1998. Phylogenetic relationships of Cornaceae and close relatives inferred from *mat*K and *rbc*L sequences. American Journal of Botany 85:285-297.

APPENDIX 4.1

Vouchers/source for taxa used in analyses and GenBank accession numbers for the sequences. Where there is only one voucher the same DNA has been used for both ITS and ETS. Samples beginning with lc are accession numbers from living collections.

Voucher/accession (citation), GenBank accession numbers ITS, ETS.

Sapindaceae

Acer saccharum L., ,AF020363, *Acer palmatum* Thunb ex A.E.Murray, ,AF020375, *Arfeuillea arborescens* Pierre, Chase 2122 (K), FJ375200, FJ372742, *Conchopetalum brachysepalum* Capuron, Radenantoandro 674 (MO), ,FJ372743 *Cossinia australiana* S.Reyn., G. P. Guymer et al. 1542 (CANB), ,FJ372745, *Cossina pinnata* Commerson ex Lam, lc950595 (NTBG), FJ375197, FJ372744, *Diplokeleba floribunda* N. E. Br., Acevedo 11130 (US), , FJ372746, *Diplopeltis eriocarpa* (Benth.) Hemsl., White, M.R.3 (CBG), FJ375190, FJ372735, *Diplopeltis intermedia* A.S.George, Bellairs, D.6158 (CANB), FJ375191, FJ372736, *Diplopeltis stuartii* F.Muell., Kendrick, P. s.n. (CANB), FJ375192, FJ372737, *Dipteronia dyeriana* Henry, ,AF401120, *Dipteronia sinensis* Oliv., ,AF020386, *Distichostemon filamentosus* S.Moore, Fryxell, P.A.et al 4924 (CANB), FJ375188, FJ372733, *Distichostemon hispidulus* (Endl.) Baill., J.G.West 5489 (CANB), FJ375189, FJ372734, *Dodonaea bursariifolia* F.Muell., Lyne, A.M.958 (CBG), FJ375180, FJ372725, *Dodonaea glandulosa* J.G.West, Newbey, K.9718 (CANB), FJ375181, FJ372726 *Dodonaea madagascariensis* Radlk., Birkinshaw C. & Raharison R. 1407 (MO), FJ375182, FJ372727, *Dodonaea petiolaris* F.Muell., LAC10562 (CANB), FJ375183, FJ372728, *Dodonaea physocarpa* F.Muell., Harrington MH293 (JCT), FJ375184, FJ372729, *Dodonaea polyandra* Merr. & L.M.Perry, Harrington MH324 (JCT), FJ375185, FJ372730, *Dodonaea triquetra* J.C.Wendl., Davies 240 (ANBG), FJ375186, FJ372731, *Dodonaea viscosa* Jacq. subsp. *viscosa*, Harrington MH325 (JCT), FJ375187, FJ372732, *Exothea diphylla* (Standl.) Lundell, Acevedo 12233 (US), , FJ372755, *Exothea paniculata* (Juss) Radlk., Acevedo 12177 (US), , FJ372756, *Filicium decipiens* (Wight & Arn.) Thwaites, Edwards KE271 (JCT), , FJ372752, *Ganophyllum falcatum* Blume, Harrington MH402 (JCT), ,FJ372753, *Harpullia arborea* (Blanco) Radlk. ,lc9300825 (CBG), FJ375194, FJ372739, *Harpullia*

rhyticarpa C.T.White & W.D.Francis , RIS126 (TFRC), FJ375195, FJ372741, *Harpullia* sp., Harrington MH403 (JCT), FJ375196, FJ372740, *Hippobromus pauciflora* (L.f.) Radlk., Edwards KE229 (JCT), FJ375203, FJ372751, *Hypelate trifoliata* Griseb., Rankin & Arias 72057 (K), , FJ372754, *Llagunoa mollis*, Jaramillollejia et al. 3199 (K), , FJ372748, *Llagunoa nitida* Ruiz & Pav., Acevedo 11143 (US), FJ375198, FJ372747, *Loxodiscus coriaceus,* Biffin EAB03/142 (CANB), FJ375193, FJ372738, *Majidea fosteri* Radlk., lc19695474 (NBGB), FJ375199, FJ372749, *Sinoradlkofera minor* (Hemsley) F.G.Meyer, Tang, S. Q. s.n. (JCT), FJ375201, FJ372750, *Xanthoceras sorbifolium* Bunge, Chase 2866 (K), FJ375202, *Zanha africana* (Radlk.) Exell, Kayombo CK4824 (MO), FJ375204, FJ372757.

Burseraceae
Bursera asplenifolia, AF080012, AF445965. *Boswellia sacra,* AF4455880, AF445957.

Ackerly, D. D., and M. J. Donoghue. 1998. Leaf size, sapling allometry, and Corner's rules: a phylogenetic study of correlated evolution in maples (*Acer*). American Naturalist 152:767-791.

APPENDIX 4.2A

Pairwise comparison of compensatory base changes in combined ITS1 and ITS2

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1 D_bursariifolia	-	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	1	1	1	3	2	0	5	2	2	2	4	1	2	5	9
2 D_glandulosa	0	-	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	2	1	1	1	1	2	1	1	1	2	1	1	3	4
3 D_madagasca	0	0	-	0	0	0	0	0	0	0	0	0	0	0	2	1	1	1	0	4	3	0	4	2	2	1	3	1	1	5	9
4 D_petiolaris	0	0	0	-	0	0	0	0	0	0	0	0	1	0	2	1	1	1	1	2	3	0	6	2	2	2	4	1	3	6	9
5 D_physocarpa	0	1	0	0	-	1	0	0	0	0	1	1	0	1	2	1	1	1	1	5	3	1	6	3	2	2	4	1	3	5	7
6 D_polyandra	0	0	0	0	1	-	0	0	0	0	0	0	0	0	2	1	1	1	0	5	3	0	5	2	1	1	3	1	1	5	8
7 Dtriquetr	0	0	0	0	0	0	-	0	0	0	0	0	0	0	2	1	1	1	0	5	3	0	4	2	1	1	3	1	1	5	9
8 Dviscosa	0	0	0	0	0	0	0	-	0	0	0	0	0	0	2	1	1	1	0	5	3	0	4	2	2	1	3	1	1	5	9
9 Disfilame	0	0	0	0	0	0	0	0	-	0	0	0	0	0	1	1	1	1	0	4	3	0	4	2	1	1	3	1	1	5	9
10Dishispid	0	0	0	0	0	0	0	0	0	-	1	1	0	0	1	1	1	1	0	4	3	0	4	2	2	1	3	1	1	5	10
11Diperioca	0	0	0	0	1	0	0	0	0	1	-	0	1	0	2	1	1	1	0	4	2	0	6	2	1	1	3	1	2	5	8
12Dipinterm	0	0	0	0	1	0	0	0	0	1	0	-	1	0	2	1	1	1	0	3	2	0	5	2	1	1	3	1	2	3	6
13Dipstuart	0	0	0	1	0	0	0	0	0	0	1	1	-	1	3	4	3	0	2	3	2	1	7	3	4	3	4	1	3	5	10
14Loxodiscus	0	1	0	0	1	0	0	0	0	0	0	0	1	-	4	3	3	2	1	5	2	2	5	2	2	1	3	0	2	4	5
15Hararbore	2	0	2	2	2	2	2	2	1	1	2	2	3	4	-	0	0	2	0	5	3	3	5	4	3	3	5	1	4	5	9
16Harrhytic	1	0	1	1	1	1	1	1	1	1	1	1	4	3	0	-	0	1	0	4	3	2	5	3	2	2	4	0	3	7	10
17HarNoah_C	1	0	1	1	1	1	1	1	1	1	1	1	3	3	0	0	-	1	0	4	2	1	4	2	1	1	3	0	2	6	9
18Arfeuillea	1	2	1	1	1	1	1	1	1	1	1	1	0	2	2	1	1	-	1	4	1	1	5	2	1	1	3	1	1	2	4
19Cossinia_p	1	1	0	1	1	0	0	0	0	0	0	0	2	1	0	0	0	1	-	4	0	0	3	1	1	1	3	0	0	6	7
20Llagunoa_n	3	1	4	2	5	5	5	5	4	4	4	3	3	5	5	4	4	4	4	-	4	4	7	4	4	3	5	2	5	7	8
21Majideafo	2	1	3	3	3	3	3	3	3	3	2	2	2	2	3	3	2	1	0	4	-	0	3	3	3	1	3	0	0	4	7
22Sinoradlko	0	1	0	0	1	0	0	0	0	0	0	0	1	2	3	2	1	1	0	4	0	-	3	1	1	0	1	0	1	3	7
23Xanthocera	5	2	4	6	6	5	4	4	4	4	6	5	7	5	5	5	4	5	3	7	3	3	-	4	2	4	3	3	4	4	7
24Acersacch	2	1	2	2	3	2	2	2	2	2	2	2	3	2	4	3	2	2	1	4	3	1	4	-	0	1	1	2	1	3	6
25Acerpalmatum	2	1	2	2	2	1	1	2	1	2	1	1	4	2	3	2	1	1	1	4	3	1	2	0	-	0	0	1	0	3	5
26Dipteronia_dye	2	1	1	2	2	1	1	1	1	1	1	1	3	1	3	2	1	1	1	3	1	0	4	1	0	-	1	1	0	5	6
27Dipteronia_sin	4	2	3	4	4	3	3	3	3	3	3	3	4	3	5	4	3	3	3	5	3	1	3	1	0	1	-	1	1	4	5
28Hippobromu	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	0	2	0	0	3	2	1	1	1	-	0	1	1
29Zanha_afri	2	1	1	3	3	1	1	1	1	1	2	2	3	2	4	3	2	1	0	5	0	1	4	1	0	0	1	0	-	3	6
30Burserra_aspl	5	3	5	6	5	5	5	5	5	5	5	3	5	4	5	7	6	2	6	7	4	3	4	3	3	5	4	1	3	-	3
31Boswelliasacr	9	4	9	9	7	8	9	9	9	10	8	6	10	5	9	10	9	4	7	8	7	7	7	6	5	6	5	1	6	3	-

APPENDIX 4.2B

Pairwise comparison of compensatory base changes in combined ITS1 and ITS2

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
1 D_bursariifolia	-	0	0	0	1	0	0	0	0	0	1	1	2	0	3	0	0	3	2	0	1	3	7	5	5	2	1	3	2	1	2	0	0	5	6
2 D_glandulosa	0	-	0	0	0	0	0	0	0	0	1	1	1	0	3	0	0	3	4	1	2	3	8	6	5	2	2	4	3	2	3	1	1	6	8
3 D_madagascari	0	0	-	0	1	0	0	0	0	0	1	1	2	0	3	0	0	2	3	2	3	5	10	8	4	4	2	4	3	2	3	2	2	6	7
4 D_petiolaris	0	0	0	-	0	0	0	0	0	0	1	1	0	0	3	0	0	1	0	0	0	2	5	2	2	1	0	2	1	0	1	0	0	3	4
5 D_physocarpa	1	0	1	0	-	1	0	0	1	1	1	1	1	1	3	1	1	3	4	4	5	5	11	7	5	3	3	5	5	2	6	2	2	8	9
6 D_polyandra	0	0	0	0	1	-	0	1	0	0	2	2	2	2	3	1	1	4	5	4	5	5	11	8	5	4	3	5	3	3	4	2	2	6	7
7 Dtriquetr	0	0	0	0	0	0	-	0	0	0	1	1	1	0	3	0	0	3	3	1	2	4	9	6	5	4	2	4	2	2	3	2	2	7	9
8 Dviscosa	0	0	0	0	0	1	0	-	1	1	1	1	1	0	3	0	0	3	3	1	2	4	9	7	5	2	1	3	3	1	3	0	0	6	7
9 Disfilame	0	0	0	0	1	0	0	1	-	0	2	2	2	1	4	1	1	4	6	4	5	4	9	7	5	4	4	5	4	4	4	3	3	6	9
10Dishispid	0	0	0	0	1	0	0	1	0	-	2	2	2	1	4	1	1	4	6	4	5	4	9	7	5	4	4	5	4	4	4	3	3	6	9
11Diperioca	1	1	1	1	1	2	1	1	2	2	-	0	2	0	4	0	1	4	3	2	3	3	9	6	6	4	2	4	4	2	4	1	1	7	9
12Dipinterm	1	1	1	1	1	2	1	1	2	2	0	-	2	0	4	0	1	4	3	2	3	3	9	6	6	4	2	4	4	2	4	1	1	7	9
13Dipstuart	2	1	2	0	1	2	1	1	2	2	2	2	-	1	3	2	2	5	6	3	4	6	11	8	7	4	4	5	6	3	5	3	3	10	11
14Loxodiscus	0	0	0	0	1	2	0	0	1	1	0	0	1	-	2	1	1	2	3	1	3	1	6	4	4	4	3	5	4	3	4	2	2	8	8
15Hararbore	3	3	3	3	3	3	3	3	4	4	4	4	3	2	-	1	1	3	4	3	3	3	8	7	5	4	3	5	5	3	4	2	2	8	11
16Harrhytic	0	0	0	0	1	1	0	0	1	1	0	0	2	1	1	-	0	3	3	1	1	4	6	6	4	2	2	4	4	3	4	1	1	8	9
17HarNoah_C	0	0	0	0	1	1	0	0	1	1	1	1	2	1	1	0	-	2	3	1	2	2	7	6	4	3	2	4	4	2	3	1	1	8	10
18Arfeuillea	3	3	2	1	3	4	3	3	4	4	4	4	5	2	3	3	2	-	2	3	4	3	6	7	1	5	2	4	4	2	3	4	4	9	10
19Conchopetalum	2	4	3	0	4	5	3	3	6	6	3	3	6	3	4	3	3	2	-	2	3	1	6	6	4	4	1	3	3	1	2	3	3	8	10
20Cossinia_p	0	1	2	0	4	4	1	1	4	4	2	2	3	1	3	1	1	3	2	-	0	2	8	8	5	3	0	2	2	0	2	0	0	7	9
21Cossinia_aust	1	2	3	0	5	5	2	2	5	5	3	3	4	3	3	1	2	4	3	0	-	3	8	9	5	4	1	2	3	1	2	1	1	8	10
22Diplokeleba_flo	3	3	5	2	5	5	4	4	4	4	3	3	6	1	3	4	2	3	1	2	3	-	6	5	4	0	1	0	3	1	2	0	0	5	7
23Llagunoa_n	7	8	10	5	11	11	9	9	9	9	9	9	11	6	8	6	7	6	6	8	8	6	-	1	6	5	6	7	8	5	6	5	5	10	12
24Llagunoa_mol	5	6	8	2	7	8	6	7	7	7	6	6	8	4	7	6	6	7	6	8	9	5	1	-	6	5	6	6	9	5	6	6	6	10	12
25Majideafo	5	5	4	2	5	5	5	5	5	5	6	6	7	4	5	4	4	1	4	5	5	4	6	6	-	6	4	6	4	4	5	6	6	9	11
26Sinoradlko	2	2	4	1	3	4	4	2	4	4	4	4	4	4	4	2	3	5	4	3	4	0	5	5	6	-	3	3	3	2	4	2	2	6	8
27Hippobromu	1	2	2	0	3	3	2	1	4	4	2	2	4	3	3	2	2	2	1	0	1	1	6	6	4	3	-	2	2	0	1	1	1	6	8
28Zanha_afri	3	4	4	2	5	5	4	3	5	5	4	4	5	5	5	4	4	4	3	2	2	0	7	6	6	3	2	-	2	2	3	2	2	6	11
29Filiciumdecipe	2	3	3	1	5	3	2	3	4	4	4	4	6	4	5	4	4	4	3	2	3	3	8	9	4	3	2	2	-	1	0	1	1	4	6
30Ganophyllum	1	2	2	0	2	3	2	1	4	4	2	2	3	3	3	3	2	2	1	0	1	1	5	5	4	2	0	2	1	-	1	0	0	6	8
31Hypelate	2	3	3	1	6	4	3	3	4	4	4	4	5	4	4	4	3	3	2	2	2	2	6	6	5	4	1	3	0	1	-	1	1	5	8
32Exothea_diph	0	1	2	0	2	2	2	0	3	3	1	1	3	2	2	1	1	4	3	0	1	0	5	6	6	2	1	2	1	0	1	-	0	5	7
33Exotpaniculata	0	1	2	0	2	2	2	0	3	3	1	1	3	2	2	1	1	4	3	0	1	0	5	6	6	2	1	2	1	0	1	0	-	5	7
34Burserra_asp	5	6	6	3	8	6	7	6	6	6	7	7	10	8	8	8	8	9	8	7	8	5	10	10	9	6	6	6	4	6	5	5	5	-	0
35Boswelliasacra	6	8	7	4	9	7	9	7	9	9	9	9	11	8	11	9	10	10	10	9	10	7	12	12	11	8	8	11	6	8	8	7	7	0	-

APPENDIX 5.1

Vouchers/source for taxa used in analyses and GenBank accession numbers for the sequences.

Species, Voucher, Genbank accession ITS, ETS.

Cossinia australiana S.Reyn., Guymer, G.P. & Jessup, L.W. 1542 (CANB), , *Cossinia pacifica* A.C.Sm., Lorence 8462 (PTBG), , *Diplopeltis eriocarpa* Hemsl., White, M.R. 3 (CBG), , *Diplopeltis huegelii* subsp. *subintegra* (A.S.George) Kieghery, Smith, B.H. 1505 (CANB), , *Diplopeltis intermedia* A.S.George var. *intermedia*, Bellairs, D. 6158 (CANB), , *Diplopeltis petiolaris* Benth., Harrington, M. 314 (JCT), , *Diplopeltis stuartii* F.Muell. *var. stuartii,* Kendrick s.n. (CANB), , *Distichostemon arnhemicus* S.T.Reynolds, Purdie, R.W.3208 (CBG), , *Distichostemon barklyanus* S.T.Reynolds, Fraser, A. 285 (CANB), , *Distichostemon dodecandrus* Domin, Harrington, M. 291 (JCT), , *Distichostemon filamentosus* S.Moore, Fryxell, P.A.et al. 4924 (CANB), , *Distichostemon hispidulus* var. *aridus* S.T.Reynolds, West, J.G. 5489 (CANB), , *Distichostemon hispidulus* (Endl.)Baill. var. *hispidulus,* Purdie, R.W. 3405 (CBG), , *Distichostemon malvaceus* Domin, Harrington, M. 301 (JCT), , *Dodonaea adenophora* Miq., KP19883352, , *Dodonaea amblyophylla* Diels, Lyne, A.M.1120 (CBG), , *Dodonaea aptera* Miq., West, J.G. 3250 (CANB), , *Dodonaea baueri* Endl., Jackson, I9282 (CBG), , *Dodonaea biloba* J.G.West, Telford & Carroll 875 (CBG), , *Dodonaea boroniifolia* G.Don, Davies 252 (CBG), , *Dodonaea bursariifolia* F.Muell., Lyne, A.M.958 (CBG), , *Dodonaea caespitosa* Diels, Chinnock, R.J.7378 (CANB), , *Dodonaea camfieldii* Maiden & Betche, Orme, A. 194 (NSW), , *Dodonaea ceratocarpa* Endl., KP19860843, , *Dodonaea concinna* Benth., Lyne, A.M.1057 (CBG), , *Dodonaea coriacea* (Ewart & O.B.Davies)McGill, Egan, S.1298 (CANB), , *Dodonaea divaricata* Benth., West, J.G.4540 (CANB), , *Dodonaea ericoides* Miq., West, J.G.3301 (CBG), , *Dodonaea falcata* J.G.West, West, J.G. 5237 (CANB), , *Dodonaea filifolia* Hook., Purdie, R.W.4242 (CBG), , *Dodonaea filiformis* Link, ANBG 2879 (CBG), , *Dodonaea glandulosa* J.G.West, West, J.G. 3151 (NSW), , *Dodonaea hackettiana* W.Fitzg., West, J.G.3263 (CANB), , *Dodonaea heteromorpha* J.G.West, Kennedy, M. 487 (NSW), , *Dodonaea hexandra* F.Muell, West, J.G.5369 (CANB), , *Dodonaea hirsuta* (Maiden & Betche) Maiden & Betche, Telford, I.R. 3113 (CBG), , *Dodonaea humifusa* Miq. , Barnsley, B. 642 (CBG), , *Dodonaea humilis* Endl., I. Jackson ANBG 2984 (CBG), , *Dodonaea inaequifolia* Turcz., Telford, I.R. 8545 (CBG), , *Dodonaea intricata* J.G.West, S.A.N.P.W.S. 9779 (CBG), , *Dodonaea lanceolata var. subsessifolia* J.G.West, Harrington, M. 292 (JCT), , *Dodonaea larreoides* Turcz., Bellairs, D. 5020 (CANB), , *Dodonaea lobulata* F.Muell, Harrington, M. 317 (JCT), , *Dodonaea macrossanii* F.Muell., Coveny, R.G. 14466 (CANB), , *Dodonaea madagascariensis* Radkl., Birkinshaw C & Raharison R 1407 (MO), , *Dodonaea megazyga* (F.Muell) F.Muell. ex Benth, Coveny, R.G. 16587 (CANB), , *Dodonaea microzyga v. acrotobata* J.G.West, West , J.G. 2703 (CANB), , *Dodonaea multijuga* G.Don, Donaldson, S. 556 (CBG), , *Dodonaea oxyptera* F.Muell, Telford 11705 (CBG), , *Dodonaea pachyneura* F.Muell, Wilson, P.G.1180 (CANB), , *Dodonaea peduncularis* Lindl., Purdie, R.W. 4280 (CBG), , *Dodonaea petiolaris* F.Muell, West, J.G. 3335 (CANB), , *Dodonaea physocarpa* F.Muell, Harrington, M. 293 (JCT), , *Dodonaea pinifolia* Miq. , Harrington, M. 332 (JCT), , *Dodonaea pinnata* Sm., West, J.G. 5464 (ANBG), , *Dodonaea platyptera* F.Muell, West, J.G. 5340 (CANB), , *Dodonaea polyandra* Merr. & L.M.Perry, Harrington, M. 324 (JCT), , *Dodonaea polyzyga* F.Muell., Duretto, M.F.; Davies, T.A.1161 (CANB), , *Dodonaea procumbens* F.Muell., West, J.G. 5502(CANB) , , *Dodonaea ptarmicaefolia* Turcz., Harrington, M. 329 (JCT), , *Dodonaea rhombifolia* N.A.Wakef., Davies 679 (CBG), , *Dodonaea rigida* J.G.West , Grace, J.922

(CANB), , *Dodonaea rupicola* C.T.White, Telford, I.R. 3415 (CBG), , *Dodonaea serratifolia* McGill, Telford, I.R. 9861 (CBG), , *Dodonaea sinuolata* J.G.West *ssp. sinuolata* , Hosking, J.R.1476 (CANB), , *Dodonaea stenophylla* F.Muell, Telford, I.R. 11094 (CBG), , *Dodonaea stenozyga* F.Muell, West, J.G. 2683 (CBG), , *Dodonaea subglandulifera* J.G.West, West, J.G. 1642 (AD), , *Dodonaea tenuifolia* Lindl., Bean, A.R.4617 (CANB), , *Dodonaea tepperi* F.Muell ex Tepper, Davies 1366 (CBG), , *Dodonaea triangularis* Lindl., Telford, I.R. 5610 (CBG), , *Dodonaea trifida* F.Muell, Lally, T.R.904 (CANB), , *Dodonaea triquetra* J.C.Wendl., Davies 240 (CBG), , *Dodonaea truncatiales* F.Muell, West, J.G. 2566 (CBG), , *Dodonaea uncinata* J.G.West, Jobson, P.C.3116 (CANB), , *Dodonaea vestita* Hook., Telford, I.R. 11935 (NSW), , *Dodonaea viscosa* Jacq. subsp *viscosa,* Harrington, M. 300 (JCT), , *Harpullia hillii* F.Muell, Harrington, M. 319 (JCT), , *Harpullia cupanioides* Roxb., Harrington, M. 333, , *Harpullia* sp., Harrington 330 (JCT), , *Harpullia rhyticarpa* C.T.White & W.D.Francis, CSIRO378, , *Harpullia arborea* (Blanco) Radlk., Edwards, K. 38 (JCT), , *Harpullia ramiflora* Radlk., CSIRO518, , *Harpullia pendula* Planch. ex F.Muell., Edwards, K. 202 (JCT), , *Loxodiscus coriaceous* Hook.f., EAB 03/134 (JCT), , .

APPENDIX 5.2

Mean posterior estimates of the sampled parameters for each model for combined analyses of ITS loops + ETSf loops + ITS stems + ETSf stems (in that order).

Model name = Mixed model : REV + dG6 + I & REV + dG6 + I & RNA7C + dG6 + I & RNA7C + dG6 + I

Model2/model1 average substitution rate ratio = 0.73454
Model3/model1 average substitution rate ratio = 1.74537
Model4/model1 average substitution rate ratio = 1.43647

MODEL1 Model name = REV + dG6 + I
FREQUENCIES
F(A) = 0.34817 F(C) = 0.24875 F(G) = 0.16623 F(T) = 0.23685
RATESRATIOS
Rate ratio 1 = 0.35279, Rate ratio 2 = 0.31493, Rate ratio 3 = 0.40263, Rate ratio 4 =2.17079, Rate ratio 5 = 0.42330
RATESCATEGORIES
Alpha parameter = 1.32962
Gamma distribution of rates = yes
Invariant category = yes
Number of gamma categories = 6
Proportion of invariant sites = 0.18528

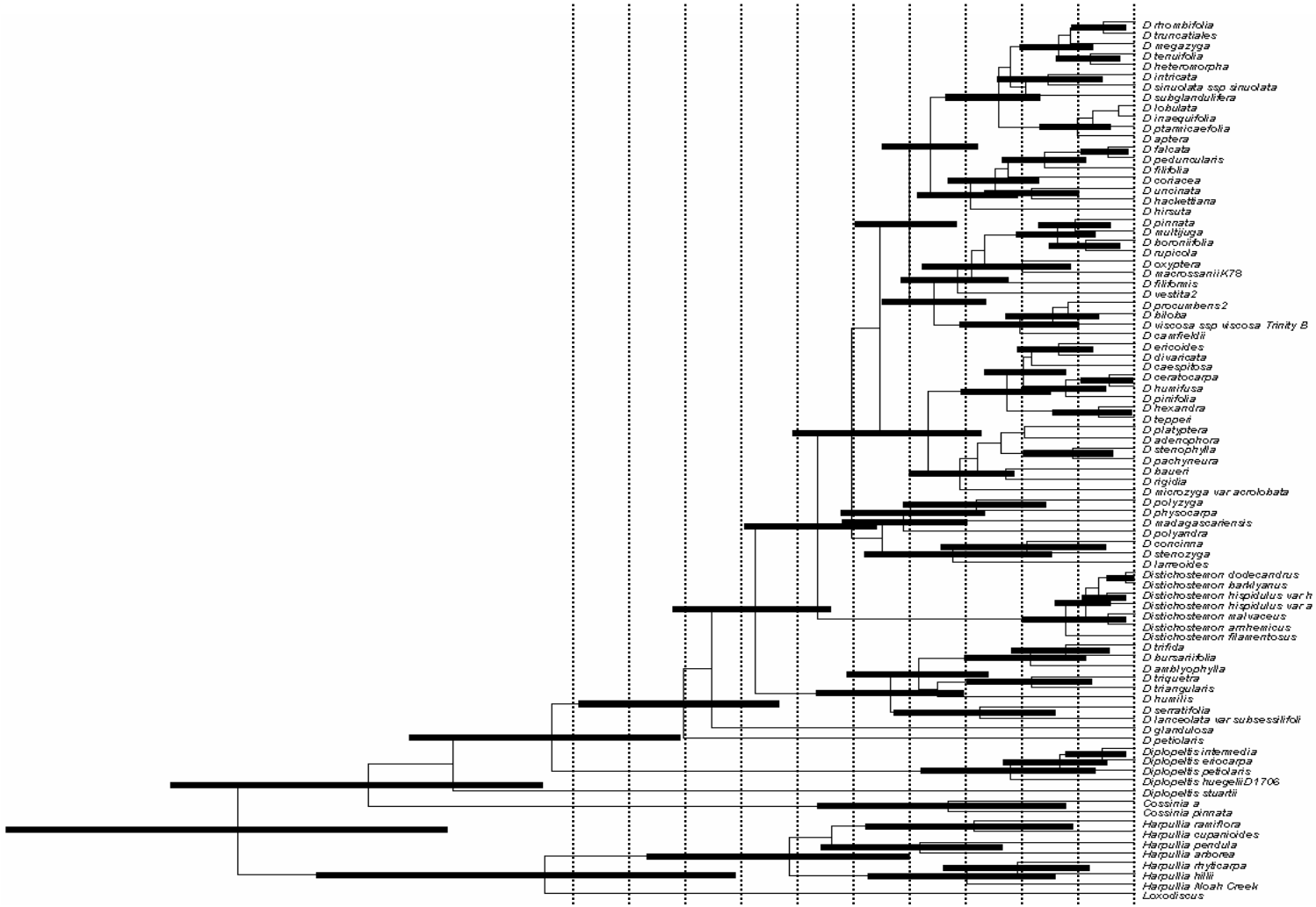
MODEL2 Model name = REV + dG6 + I
FREQUENCIES
F(A) = 0.24398 F(C) = 0.15722 F(G) = 0.19324 F(T) = 0.40556
RATESRATIOS
Rate ratio 1 = 0.43356, Rate ratio 2 = 0.46717, Rate ratio 3 = 0.47193, Rate ratio 4 = 1.79450, Rate ratio 5 = 0.36598
RATESCATEGORIES
Alpha parameter = 1.84557
Gamma distribution of rates = yes
Invariant category = yes
Number of gamma categories = 6
Proportion of invariant sites = 0.07711

MODEL3 Model name = RNA7C + dG6 + I
FREQUENCIES
F(AU) = 0.11152 F(CG) = 0.25022 F(GC) = 0.25535 F(GU) = 0.09189 F(MM) = 0.12651 F(UA) = 0.05868 F(UG) = 0.10583
RATESRATIOS
Rate ratio 1 = 10.10567, Rate ratio 2 = 14.13013, Rate ratio 3 = 6.13205, Rate ratio 4 = 6.69393, Rate ratio 5 = 3.95745, Rate ratio 6 = 4.39128, Rate ratio 7 = 15.76781, Rate ratio 8 = 2.83185, Rate ratio 9 = 4.22270
RATESCATEGORIES
Alpha parameter = 2.16923
Gamma distribution of rates = yes
Invariant category = yes
Number of gamma categories = 6
Proportion of invariant sites = 0.06942

MODEL4 Model name = RNA7C + dG6 + I
FREQUENCIES
F(AU) = 0.20037 F(CG) = 0.17621 F(GC) = 0.18095 F(GU) = 0.13408 F(MM) = 0.11967 F(UA) = 0.10950 F(UG) = 0.07921
RATESRATIOS
Rate ratio 1 = 2.62989, Rate ratio 2 = 2.51197, Rate ratio 3 = 4.46116, Rate ratio 4 = 1.92295, Rate ratio 5 = 1.16306, Rate ratio 6 = 1.68810, Rate ratio 7 = 2.19666, Rate ratio 8 = 1.88825, Rate ratio 9 = 1.45250
RATESCATEGORIES
Alpha parameter = 2.56138
Gamma distribution of rates = yes
Invariant category = yes
Number of gamma categories = 6
Proportion of invariant sites = 0.03337

APPENDIX 5.3 (NEXT PAGE)

The posterior distribution of relative branching times. Bars represent the 95% HPD interval for the relative (mean root height scaled to one) branching times from the Bayesian uncorrelated log normal relaxed clock estimation of phylogeny. Bar appear only on nodes that receive more than 50% PP. The space between vertical dashed lines represents a one million year time interval.


D rhombifolia
D truncatiales
D megazyga
D tenuifolia
D heteromorpha
D intricata
D sinuolata ssp sinuolata
D subglandulifera
D lobulata
D inaequifolia
D ptarmicaefolia
D aptera
D falcata
D peduncularis
D filifolia
D coriacea
D uncinata
D hackettiana
D hirsuta
D pinnata
D multijuga
D boroniifolia
D rupicola
D oxyptera
D macrossaniiK78
D filiformis
D vestita2
D procumbens2
D biloba
D viscosa ssp viscosa Trinity B
D camfieldii
D ericoides
D divaricata
D caespitosa
D ceratocarpa
D humifusa
D pinifolia
D hexandra
D tepperi
D platyptera
D adenophora
D stenophylla
D pachyneura
D baueri
D rigidia
D microzyga var acrolobata
D polyzyga
D physocarpa
D madagascariensis
D polyandra
D concinna
D stenozyga
D lamoides
Distichostemon dodecandrus
Distichostemon barklyanus
Distichostemon hispidulus var h
Distichostemon hispidulus var a
Distichostemon malvaceus
Distichostemon armenicus
Distichostemon filamentosus
D trifida
D bursariifolia
D amblyophylla
D triquetra
D triangularis
D hurelitis
D serratifolia
D lanceolata var subsessilifoli
D glandulosa
D petiolaris
Diplopeltis intermedia
Diplopeltis eriocarpa
Diplopeltis petiolaris
Diplopeltis huegeliiD1706
Diplopeltis stuartii
Cossinia a
Cossinia pinnata
Harpullia ramiflora
Harpullia cupanioides
Harpullia pendula
Harpullia arborea
Harpullia rhyticarpa
Harpullia hillii
Harpullia Moah Creek
Loxodiscus

APPENDIX 6.1

Vouchers/source and where sampled for taxa used in analyses, and GenBank accession numbers for the sequences.

Species, voucher/source, where sampled, Genbank accession ITS, ETS.

Dodonaea biloba J.G.West, Telford & Carroll 875 (CBG), FJ546929, FJ546876.
Dodonaea camfieldii Maiden & Betche, Orme, A. 194 (NSW), FJ546930, FJ546877.
Dodonaea filiformis Link, ANBG 2879 (CBG), FJ546978, FJ546927.
Dodonaea macrossanii F.Muell., Coveny, R.G. 14466 (CANB), FJ546977, FJ546926.
Dodonaea oxyptera F.Muell, Telford 11705 (CBG), FJ546979, FJ546928.
*Dodonaea procumbens*1 F.Muell., West, J.G. 5502 (CANB) , southern tablelands NSW, Australia, 745m, FJ546931, FJ546878.
*Dodonaea procumbens*2 F.Muell., living collection. 7703670 (CBG), Australia, Victoria - Meyer's Nursery as plant 7703670, FJ546932, FJ546879.
Dodonaea vestita Hook., Telford, I.R. 11935 (NSW), FJ546925, FJ546976.
Dodonaea viscosa Arizona1, Yatskievych 05-03 (MO), Tonto National Forest 900m, FJ546933, FJ546880.
Dodonaea viscosa Arizona2, Yatskievych 05-04 (MO), Coronado National Forest 1000m, FJ546934, FJ546881.
Dodonaea viscosa Bolivia, Acevedo, P. 11144 (USI), upland, FJ546935, FJ546882.
Dodonaea viscosa Brazil, Wasum & Wasum182 8(CBG), no details, FJ546936, FJ546883.
Dodonaea viscosa China, Hyland14946 (CANB), open forest 1000m southern central China, FJ546938, FJ546885.
Dodonaea viscosa Colombia, FB/S778 (NPB), Paramo, Chia, FJ546937, FJ546884.
Dodonaea viscosa Hawaii Kauai, Edwards, K. 270 (JCT), upland, FJ546943, FJ546891.
Dodonaea viscosa Hawaii Kona, Morden 1136 (BISH), west coast, FJ546942, FJ546890.
Dodonaea viscosa Hawaii Pohakuloa , Morden 1309 (BISH), upland, FJ546941, FJ546889.
Dodonaea viscosa India, S. Phartyal s.n. (JCT), New Tehri, Uttaranchal (India) central Himalayas, FJ546944, FJ546892.
Dodonaea viscosa Japan, Nakaike, T. s.n. (CANB), fringes of secondary forest Ryukyu Islands 20 metres, FJ546939, FJ546886.
Dodonaea viscosa Maui Poli Poli , Morden 1787 (BISH), upland, FJ546940, FJ546888.
Dodonaea viscosa Maui Ulupalakua , (Howarth et al., 2007) PCMB B80/687878 (BISH), upland area above lava tube cave, AY864896, FJ546887.
Dodonaea viscosa Mexico, V.L.Barradas sn (JCT), upland Mexico City, FJ546945, FJ546893.
Dodonaea viscosa New Caledonia1, Biffin&Craven115 (JCT), 100m on floodplain of Taitoula River, FJ546946, FJ546894.
Dodonaea viscosa New Caledonia2, Brown, E.A. 03/134 (NSW), Mont Dore at 300m on steep slope, FJ546947, FJ546895.
Dodonaea viscosa NZ North Island, 19981090 living collection (RBGE), Lower Hutt 50m, FJ546948, FJ546896.
Dodonaea viscosa NZ South Island1, J.G.West5458 (CANB), South island north coast hill slopes, FJ546949, FJ546897.

Dodonaea viscosa NZ South Island2, J.G.West5283 (CANB), East coast south island coastal hill slopes, FJ546950, FJ546898.
Dodonaea viscosa NZ South Island3, S.J. Wagstaff & M. I. Dawson sn (CANB), Canterbury, FJ546951, FJ546899.
Dodonaea viscosa Oman, J.R.Maconochie3481 (CBG), no details, FJ546952, FJ546900.
Dodonaea viscosa Pagan, Tony Roberts s.n. (JCT), foreshore 1m, FJ546953, FJ546901.
Dodonaea viscosa Papua New Guinea, (Wright et al., 2006), upland, DQ499143, ─ .
Dodonaea viscosa Peru, Merello et al. 1077 (MO), Cajamarca: 2400m, FJ546962, FJ546910.
Dodonaea viscosa South Africa1, KE234 (JCT), Hillcrest, Province of KwaZulu-Natal - coastal, FJ546955, FJ546903.
Dodonaea viscosa South Africa2, Peter Linder s.n. (JCT), Cederberg Western Cape Province 190 klm north of Cape Town South Africa- upland, FJ546954, FJ546902.
Dodonaea viscosa South Africa3, G.Germishuizen4048 (CANB), no details, FJ546956, FJ546904.
Dodonaea viscosa South Africa4, J.G.West5386 (CANB), 100 klm NE of Cape Town in arid shrublands, FJ546957, FJ546905.
Dodonaea viscosa subsp *angustifolia,* M.D.Crisp 4017 (CBG), inland Blue Mountains at 1080m west of Sydney, NSW, Australia, FJ546963, FJ546911.
Dodonaea viscosa subsp *angustissima,* G. Stewart 703 (CBG), slopes of Great Dividing Range west of Canberra 600m, Australia, FJ546964, FJ546912.
Dodonaea viscosa subsp *angustissima*2, MH299 (JCT), inland, Woolshed Falls, Victoria, Australia, FJ546965, FJ546913.
Dodonaea viscosa subsp *angustissima*3, MH321 (JCT), inland, Albury, NSW, Australia, FJ546966, FJ546914.
Dodonaea viscosa subsp *burmaniana*1, I. Telford10625 (CBG), inland, Atherton tablelands 440m, Queensland Australia, FJ546964, FJ546915.
Dodonaea viscosa subsp *burmaniana*2, MH406 (JCT), inland, Bakers Blue Atherton Tablelands, Queensland Australia, FJ546967, FJ546916.
Dodonaea viscosa subsp *cuneata,* West 5467 seed collection (CANB), inland New South Wales, Australia, FJ546969, FJ546918.
Dodonaea viscosa subsp *mucronata,* J.R.Connors977 (CANB), inland, central desert, Australia, FJ546970, FJ546919.
Dodonaea viscosa subsp *spatulata,* Canning5646 (CBG), inland near Horsham, Victoria 600m, Australia, FJ546968, FJ546917.
Dodonaea viscosa subsp *viscosa,* P.R Sharpe, G.Batianoff4029 (CANB), beach front, Airlie Beach, Queensland, Australia, FJ546971, FJ546920.
Dodonaea viscosa subsp *viscosa*, MH325 (JCT), beach front, Yorkey's Knob Beach, Queensland, Australia, FJ546972, FJ546921.
Dodonaea viscosa subsp *viscosa*, Jago, R.L. 4870 (JCT), beach front, Wonga Beach, Queensland, Australia, FJ546974, FJ546923.
Dodonaea viscosa subsp *viscosa,* MH407 (JCT), beach front, Trinity Beach, Queensland, Australia, FJ375187, FJ372732.
Dodonaea viscosa subsp *viscosa*, MH300 (JCT), beach front, Clifton Beach, Queensland, Australia, FJ546973, FJ546922.
Dodonaea viscosa Taiwan1, Chien-I Huang 1897 (HAST), Taipei City 55m, FJ546960, FJ546908.

Dodonaea viscosa Taiwan2, Chien-I Huang 1909 (HAST), Kuanyin Coastal Reserve 100m, FJ546961, FJ546909.
Dodonaea viscosa Tanzania1, Kayombo, C. 4816 (MO), inland Saadani National Park 1190m, FJ546958, FJ546906.
Dodonaea viscosa Tanzania2, Kayombo, C. 4820 (MO), beach front Saadani Village 1m, FJ546959, FJ546907.
Dodonaea viscosa Virgin Islands, Burbidge s.n. (JCT), beach front Sandy Point, St. Croix 2m, FJ546975, FJ546924.

Howarth, F., S. James, W. McDowell, D. Preston, and C. Imada. 2007. Identification of roots in lava tube caves using molecular techniques: implications for conservation of cave arthropod faunas Journal of Insect Conservation 11:251-261.
Wright, S., J. Keeling, and L. Gilman. 2006. The road from Santa Rosalia: a faster tempo of evolution in tropical climates. Proceedings of the National Academy of Sciences USA. 103:7718-7722.

APPENDIX 6.2

Summary statistics for the substitution model parameters from the Bayesian analyses on the concatenated alignment ITS loops(1) + ITS stems(2) + ETSf loops(3) + ETSf stems(4) with a general time reversal model (GTR) plus gamma distribution applied to each of the four partitions. Summaries are based on a total of 12000 samples from 2 runs. Each run produced 8000 samples of which 6000 samples were included.

			95% Cred. Interval	
Parameter	Mean	Variance	Lower	Upper
r(A<->C)(1)	0.170980	0.001654	0.101160	0.256884
r(A<->G)(1)	0.172320	0.002659	0.083409	0.283488
r(A<->T)(1)	0.040931	0.000458	0.010507	0.093078
r(C<->G)(1)	0.155040	0.002751	0.067101	0.269249
r(C<->T)(1)	0.346223	0.003588	0.235983	0.471185
r(G<->T)(1)	0.114505	0.002314	0.037645	0.223396
r(A<->C)(2)	0.165908	0.005839	0.048689	0.336842
r(A<->G)(2)	0.238204	0.009154	0.076221	0.441728
r(A<->T)(2)	0.157120	0.004715	0.049690	0.313003
r(C<->G)(2)	0.071229	0.002275	0.008128	0.187426
r(C<->T)(2)	0.265424	0.006106	0.136212	0.437862
r(G<->T)(2)	0.102115	0.002650	0.026189	0.223735
r(A<->C)(3)	0.063376	0.001405	0.011555	0.154589
r(A<->G)(3)	0.250523	0.005123	0.127645	0.404950
r(A<->T)(3)	0.169101	0.005610	0.048777	0.335152
r(C<->G)(3)	0.025724	0.000140	0.008494	0.054302
r(C<->T)(3)	0.443378	0.006891	0.281221	0.607533
r(G<->T)(3)	0.047898	0.000509	0.015129	0.102227
r(A<->C)(4)	0.099473	0.001847	0.032282	0.198852
r(A<->G)(4)	0.296206	0.003827	0.183783	0.425856
r(A<->T)(4)	0.141712	0.002416	0.059632	0.251908
r(C<->G)(4)	0.044330	0.000487	0.011296	0.097333
r(C<->T)(4)	0.385610	0.004378	0.261685	0.519703
r(G<->T)(4)	0.032671	0.000361	0.006362	0.079595
pi(A)(1)	0.332401	0.001018	0.271771	0.397317
pi(C)(1)	0.268081	0.000796	0.213994	0.324230
pi(G)(1)	0.121764	0.000456	0.083319	0.166167
pi(T)(1)	0.277754	0.000888	0.222655	0.338047
pi(A)(2)	0.174127	0.000805	0.122450	0.232610
pi(C)(2)	0.229522	0.000914	0.172801	0.290711
pi(G)(2)	0.173067	0.000811	0.120526	0.232190
pi(T)(2)	0.423283	0.001315	0.352834	0.495500
pi(A)(3)	0.101606	0.000313	0.069595	0.139475
pi(C)(3)	0.307964	0.000710	0.256231	0.361052

pi(G)(3)	0.404759	0.000856	0.348710	0.462881
pi(T)(3)	0.185672	0.000485	0.144353	0.230162
pi(A)(4)	0.170003	0.000414	0.131562	0.211571
pi(C)(4)	0.234192	0.000524	0.190315	0.280375
pi(G)(4)	0.323751	0.000667	0.274466	0.374764
pi(T)(4)	0.272054	0.000574	0.226903	0.321181
alpha(1)	102.217220	3246.266804	7.363111	195.336220
alpha(2)	3.585303	453.611406	0.047348	61.033633
alpha(3)	98.119067	3445.114631	3.995981	195.158700
alpha(4)	97.611908	3455.130579	0.696568	194.636616