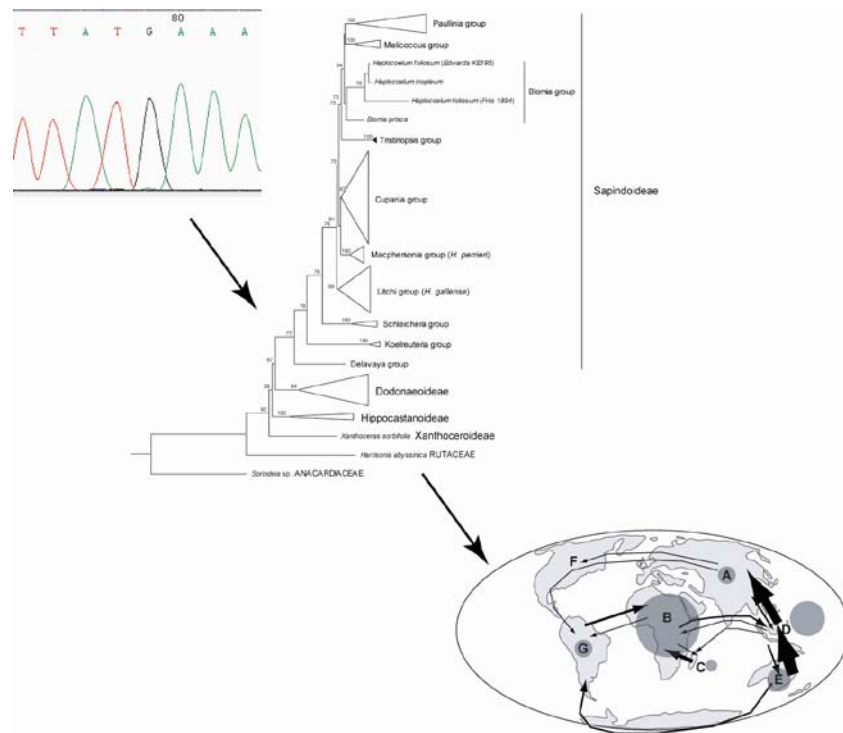


Worldwide biogeography and systematics of Sapindaceae: a molecular and taxonomic survey combining large data sets and novel methodological approaches

par

Sven BUERKI



Thèse acceptée sur proposition du jury :

Prof. Philippe Küpfer (University of Neuchâtel, Switzerland), directeur de thèse
Dr. Nadir Alvarez (University of Neuchâtel, Switzerland), co-directeur de thèse
Dr. Martin W. Callmander (Missouri Botanical Garden, Saint-Louis, U.S.A), rapporteur
Dr. Félix Forest (Royal Botanic Gardens, Kew, U.K.), rapporteur
Dr. Porter P. Lowry II (Missouri Botanical Garden, Saint-Louis, U.S.A), rapporteur
Dr. Sylvain G. Razafimandimbison (Royal Swedish Academy of Science,
Stockholm, Sweden), rapporteur
Dr. Nicolas Salamin (University of Lausanne, Lausanne, Switzerland), rapporteur
Prof. Giorgina Bernasconi (University of Neuchâtel, Neuchâtel, Switzerland), rapporteur

Soutenu le 12 Juin 2009

Université de Neuchâtel
2009

IMPRIMATUR POUR LA THESE

Worldwide biogeography and systematics of Sapinda-
ceae : a molecular and taxonomic survey involving large
data sets and novel methodological approaches

Sven BUERKI

UNIVERSITE DE NEUCHATEL

FACULTE DES SCIENCES

La Faculté des sciences de l'Université de Neuchâtel,
sur le rapport des membres du jury

Mme G. Bernasconi Fusi, MM. P. Küpfer (directeur de thèse),
N. Alvarez (co-directeur de thèse), M.W. Callmander (St Louis, USA),
N. Salamin (Lausanne), F. Forest (Richmond UK),
S.G. Razafimandimbison (Stockholm) et P.P. Lowry (St Louis USA et Paris)

autorise l'impression de la présente thèse.

Neuchâtel, le 29 juin 2009

Le doyen :
F. Kessler

UNIVERSITE DE NEUCHATEL
FACULTE DES SCIENCES
Secrétariat - décanat de la faculté
Rue Emile-Argand 11 - CP 158
CH-2009 Neuchâtel
Felix Kessler

*« Je dédie ce travail à ma maman,
mes deux papas et mon amie »*

TABLE OF CONTENT

Abstract	i
Résumé	iii
Remerciements	v
Context of this study	ix
Introduction	1
Chapter one: <i>Plastid and nuclear DNA markers reveal intricate relationships at subfamilial and tribal levels in the soapberry family (Sapindaceae)</i>	21
Chapter two: <i>New insights on parametric biogeography based on the worldwide soapberry family (Sapindaceae)</i>	45
Chapter three: <i>Abiotic factors as driving force to promote speciation in Sapindaceae</i>	115
Chapter four: <i>Molecular phylogenetic and morphological evidence support recognition of Gereaua, a new endemic genus of Sapindaceae from Madagascar</i>	135
Chapter five: <i>A synoptic revision of the genus Lepisanthes Blume (Sapindaceae) in Madagascar</i>	163
Chapter six: <i>Comparative performance of supertree algorithms in large data sets using the soapberry family (Sapindaceae) as a case study</i>	183
Conclusion & Perspectives	241
Appendix I: <i>Overview of the supertree methods used in this study</i>	255
Appendix II: <i>Curriculum vitae</i>	263

Abstract

The economically important soapberry family (Sapindaceae; Sapindales) comprises about 1900 species (distributed into ca. 140 genera; e.g., *Litchi*, *Paullinia*) mainly found in tropical regions, with only a few genera being restricted to temperate areas. For more than a century, the circumscription of the family (especially the potential inclusion of Aceraceae and Hippocastanaceae within the Sapindaceae) as well as the relationships among subfamilial entities have been widely challenged (chapter 1 for a review). In this study, infrafamilial relationships within the Sapindaceae and its relationships to the closely related Aceraceae and Hippocastanaceae are investigated based on eight nuclear and plastid markers and inferred from the complementary supermatrix (chapter 1) and supertree (chapter 6) approaches. Both approaches support the monophyly of Sapindaceae when Aceraceae and Hippocastanaceae are included and highlight a high level of paraphyly and polyphyly at the subfamilial and tribal levels. The monophyletic status of several genera is even contested (e.g., *Cupaniopsis*, *Haplocoelum*, *Matayba*). In order to maintain monophyly, a new informal classification is proposed based on molecular and morphological evidence. The soapberry family is thus subdivided into four subfamilies and 14 groups as follows (sorted according to phylogenetic relationships): Xanthoceroideae (only composed by *Xanthoceras sorbifolium*), Hippocastanoideae (two groups; including the previous Aceraceae and Hippocastanaceae), Dodonaeoideae (two groups) and Sapindoideae (ten groups). In addition, further molecular and morphological investigations allow the recognition of a new Malagasy genus, *Gereaua*, segregated from *Haplocoelum* (chapter 4). A taxonomic revision of *Lepisanthes* in Madagascar is also proposed with the description of a new species, *L. sambiranensis* (chapter 5). Phylogenetic framework, fossils data and taxa distributions are used to infer the evolutionary history of the soapberry family. This is achieved by applying and comparing state-of-the-art biogeographic methods. Moreover, additional contributions to the biogeographic framework are proposed, for instance the implementation of a biogeographic model based on paleogeographic connections (chapter 2). The influence of divergence time uncertainty on biogeographic scenario is also considered (chapter 2). Finally, the impact of abiotic (e.g., intense tectonic activities, orbital forces) and biotic (e.g., co-evolution plants/animals) factors on the diversification of the Sapindaceae is investigated based on biogeographic inference and divergence time estimations (chapter 3). Results strongly suggest an origin of Sapindaceae in temperate Asia sometime in the Early Cretaceous with a subsequent spread all over the tropics since the Late Paleocene (chapter 3). In this study, it is shown, for the first time, that abrupt climatic change in the Eocene-Oligocene boundary triggered the diversification rates of the Sapindaceae. This paradigm-breaking result is mainly due to the geological and climatic properties of South East Asia that favoured multiple contacts between lineages and further speciation across Laurasian and Gondwanian continents. This study highlights the importance of South East Asia in the evolution of the soapberry family (as well as that of additional angiosperms families) and underlines the importance to preserve this highly endangered area.

Key-words: Aceraceae; biogeography; classification; climate change; Eocene-Oligocene boundary; Hippocastanaceae; large data set; molecular; polyphyly; Sapindaceae; systematics.

Résumé

La famille des Sapindaceae est importante d'un point de vue économique et comprend plus de 1900 espèces (distribuées dans approximativement 140 genres; par exemple *Litchi*, *Paullinia*), majoritairement distribuées en zone tropicale. Cependant, certains genres peuvent coloniser les zones tempérées. Depuis plus d'un siècle, la définition de cette famille (plus particulièrement la possible inclusion des Aceraceae et Hippocastanaceae au sein des Sapindaceae) ainsi que les relations entre les sous-familles, ont été largement débattues (voir chapitre 1 pour un résumé). Dans ce travail, les relations au sein des Sapindaceae, ainsi que celles entre les Aceraceae et Hippocastanaceae, sont étudiées sur la base de huit marqueurs moléculaires (nucléaire et chloroplastiques) en utilisant une approche complémentaire de types supermatrice (chapitre 1) et supertree (chapitre 6). Les deux approches supportent la monophylie des Sapindaceae lorsque les Aceraceae et Hippocastanaceae sont incluses, et montrent un haut taux de paraphylie et polyphylie au niveau des sous-familles et tribus. De plus, les résultats contestent la monophylie de plusieurs genres (par exemple, *Cupaniopsis*, *Haplocoelum*, *Matayba*). Afin de maintenir le critère de monophylie, une nouvelle classification informelle des Sapindaceae est proposée sur la base de caractères moléculaire et morphologique. La famille des Sapindaceae est donc subdivisée en quatre sous-familles et dix groupes comme suit (les sous-familles sont triées par ordre phylogénétique): Xanthoceroideae (comprend uniquement *Xanthoceras sorbifolium*), Hippocastanoideae (deux groupes; comprend les Aceraceae et Hippocastanaceae), Dodonaeoideae (deux groupes) et Sapindoideae (dix groupes). De plus, des analyses moléculaires et morphologiques complémentaires ont permis de reconnaître un nouveau genre endémique de Madagascar, *Gereaua*, ségrégué d'*Haplocoelum* (chapitre 4). Une révision taxonomique du genre *Lepsianthes* à Madagascar est également présentée, dans laquelle une espèce nouvelle est décrite, *L. sambiranensis* (chapitre 5). Finalement, l'arbre phylogénétique ainsi que les données sur les fossiles et la distribution des taxa ont été utilisés pour investiguer l'histoire évolutive des Sapindaceae. Cela a été rendu possible par l'application et la comparaison des toutes dernières méthodes développées en biogéographie. Une contribution au développement des analyses biogéographiques est également proposée par la présentation d'un modèle biogéographique basé sur les relations paléogéographiques (chapitre 2). De plus, l'incertitude sur l'estimation de l'âge des clades a été considérée lors de l'interprétation des scénarios biogéographiques (chapitre 2). Finalement, l'impact des facteurs abiotiques (par exemple, les intenses activités tectoniques ou les forces orbitales) et biotiques (par exemple, la co-évolution plantes/animaux) sur la diversification des Sapindaceae ont été étudiés. Ces analyses supportent une origine des Sapindaceae en Asie tempérée au début du Crétacé suivie par une colonisation des zones tropicales depuis la fin du Paléocène (chapitre 3). Cette étude montre, pour la première fois, que le changement climatique abrupt ayant eu lieu entre l'Eocène et l'Oligocène, a accéléré le taux de diversification des Sapindaceae. Ce résultat, qui s'oppose à la majorité des paradigmes (voir chapitre 3 pour plus d'information), est principalement dû aux propriétés géologiques et climatiques rencontrées en Asie du Sud Est. En effet, cette région a favorisé de multiples contacts entre les lignées de Sapindaceae et de successives spéciations ont eu lieu sur les continents Laurasien et Gondwanien. Cette étude montre l'importance jouée par l'Asie du Sud Est dans l'évolution des Sapindaceae (ainsi que probablement d'autres familles d'angiospermes) et souligne l'importance de préserver cette région qui subit de fortes pressions humaines.

Mots-clés: Aceraceae; biogéographie; classification; changement climatique; Eocène-Oligocène; grand jeu de données; Hippocastanaceae; moléculaire; polyphylie; Sapindaceae; systématique.

Remerciements

Je tiens particulièrement à remercier le Prof. Philippe Küpfer pour m'avoir transmis le « petit plus » qui ouvre les portes de la découverte. Nos discussions autour d'un bon café vont définitivement me manquer!!! Merci pour ta générosité et ton constant soutien.

Je me dois également de remercier du fond du cœur mon collègue et ami, Nadir Alvarez, qui a, tout comme Philippe, permis mon envol tant personnel que scientifique. Nadir, merci pour ta disponibilité et ton unique capacité à transmettre tes connaissances. Ton intelligence et constante motivation ont été fondamentales dans l'élaboration et la concrétisation de ce travail. J'espère, maintenant pouvoir être capable de te retourner l'appareil.

Je suis également reconnaissant envers mon ami, Martin Callmander, qui m'a littéralement donné le virus de la botanique tropicale. Martin a permis la réalisation d'un de mes rêves de gosse, découvrir et décrire une nouvelle espèce. Il a même renchérit ce rêve par la description d'un nouveau genre! De plus, il a été un des piliers fondateurs de l'élaboration de ce travail. Pour cela, et pour plus encore, je te suis infiniment reconnaissant. Je n'oublierai jamais nos bon moments partagés à Madagascar.

Je tiens tout particulièrement à remercier Félix Forest de m'avoir accueilli dans son équipe. Cette rencontre restera pour moi un des points culminants de ce doctorat, tant d'un point de vue scientifique que personnel. J'espère une fois pourvoir atteindre une place autant importante que Félix, tout en conservant la même humilité dont il fait preuve. Je suis également reconnaissant envers Eve Lucas et la petite Maryse pour m'avoir accueilli sous leur toit. Cependant, je ne dois pas oublier qu'une telle rencontre n'aurait pu être possible sans son orchestration, de main de maître, organisée par un homme formidable, Guillaume Gigot. Merci beaucoup Guillaume!!!

Je suis tout particulièrement reconnaissant envers Isabel Sanmartín pour m'avoir accueilli et formé sur les méthodes de reconstruction biogéographiques. Ses connaissances uniques dans ce domaine m'ont véritablement permis d'amener ce travail dans des territoires

qui me semblaient inaccessibles. Isabel m'a poussé dans mes retranchements, mais je dois avouer que rétrospectivement cela a vraiment porté ces fruits. Je ne dois pas non plus oublier, qu'elle m'a donné l'opportunité de rencontrer des personnes compétentes et d'une extrême gentillesse, Johan Nylander et Frederik Ronquist. Je tiens particulièrement à vous remercier pour votre soutien dans les moments difficiles.

Je suis également reconnaissant envers Giorgina Bernasconi pour m'avoir donné la possibilité de finaliser ce travail. Je regrette simplement de ne pas avoir eu plus de temps pour faire ta connaissance. Merci pour la confiance que tu m'as témoigné!

Il faut maintenant que je remercie mon pote de toujours, Nils Arrigo, avec qui j'ai partagé un sacré bout de chemin! Je suis vraiment heureux de te compter parmi mes amis et te souhaite une longue carrière! Peut-être que nos chemins vont se recroiser...

Je me dois aussi de remercier Pete Lowry et Pete Phillipson qui m'ont soutenu et cru en moi depuis plus de cinq ans déjà! Je suis vraiment chanceux de vous avoir rencontré, et souhaite du fond du cœur poursuivre notre collaboration. Il sera bientôt temps pour moi de revenir à Paris afin de finir nos projets en cours...

J'ai également une pensée pour toute l'équipe du Missouri Botanical Garden à Madagascar (et plus particulièrement Sylvie) qui ont permis le bon déroulement de chacune de mes visites. Lorsqu'on invoque Madagascar, Sébastien Wohlhauser (et toute la petite famille) nous apparaît directement. Merci beaucoup Séb pour tous ces bons moments partagés autour du feu.

Je tiens également à remercier Michel Omer Laivao pour sa gentillesse, joie de vivre et surtout son adversité. Omer tu es vraiment un exemple pour moi! A bientôt à Madagascar.

Je tiens également à remercier tous mes amis qui sont ou ont passés par le laboratoire de Botanique évolutive: Julien Burnier, Philippe Chassot, Jonathan Kissling, Anouk Béguin (oups, Sarr), Aurélien Labhardt, Sylvain Lappe, Roberto Guadagnolo, François Felber, Louis Zeltner, Yong-Ming Yuan, Philippe Druart, Nicolas Schoenenberger, Mei Lin Cheung, Jason Grant, Anne Burkhardt, Anne-Marie Labouche et Loïc Fauqueur.

Je dois également remercier le personnel des herbiers de Genève (plus particulièrement Laurent Gautier, Louis Nusbaumer et Nicolas Fumeaux), Paris (plus particulièrement Jean-Noël Labat, Marc Pignal et Thierry Deroin) et Kew pour leur accueil durant mes nombreuses visites.

Je tiens tout particulièrement à remercier les membres du jury qui ont pris la peine de lire ce travail et d'y apporter un nouveau regard. Merci pour le temps que vous m'avez consacré!

Finalement, je suis infiniment reconnaissant envers ma famille pour m'avoir constamment soutenu durant ces années. Cela fait maintenant quelques années que je suis étudiant! Merci Maman et Georges-Alain d'avoir cru en moi et surtout d'avoir su réunir les conditions qui ont permis à la graine de germer! J'espère à mon tour être capable de transmettre cela à mes enfants. Il est très important pour moi de leur dédier ce travail ainsi qu'à mon papa qui doit regarder cela d'en haut. Je veux aussi témoigner mon amour, à ma petite femme, Fanny, qui a été une pièce fondamentale dans la réussite de ce travail ainsi que dans mon accomplissement personnel. Ce travail t'appartient autant qu'à moi!

Ce travail a été rendu possible par l'appui financier du laboratoire de Botanique évolutive de l'Université de Neuchâtel et une bourse obtenue par le fond SYNTHESIS pour aller conduire des recherches au Real Jardin Botanico de Madrid (Espagne).

Context of this study

My Master thesis – conducted in collaboration with the University of Neuchâtel (Switzerland) and the Missouri Botanical Garden (U.S.A.) – was devoted to the taxonomic revision of the genus *Gouania* (Rhamnaceae) in the western Indian Ocean (including Madagascar, the Comoros Islands and the Mascareignes Islands; see Buerki et al., in press). After an intense botanic and human being experience in Madagascar, I greatly expressed my desire to pursue with a PhD study focused on this highly diversified but endangered flora. Based on my taxonomic background, I first planned to apply molecular methods as a tool to unravel taxonomic entities (e.g., circumscription of genera and species concepts) in a poorly known and well diversified family of flowering plants in Madagascar. The choice of the soapberry family (Sapindaceae) was realized with the support of researchers involved in the catalogue of vascular plants of Madagascar (mainly Martin Callmander, Pete Phillipson and Pete Lowry; see Phillipson et al., 2006 for more details on the project). They underlined the necessity to provide a new family treatment for Malagasy Sapindaceae. Since the first treatment published by Capuron in 1969 [in which 26 genera (from which nine are endemic) and less than 100 species are recognized], Malagasy Sapindaceae were poorly investigated as indicated by the thousands of not identified specimens gathered in herbariums all around the world. In his treatment, Capuron (1969) mentioned the difficulties to determine flowering specimens at the species and even generic levels. For instance, to date, there are no distinctive floral characters that allow the discrimination of genera such as *Neotina*, *Tina* and *Tinopsis* (Capuron, 1969). To further investigate this family in Madagascar, three fieldworks (each of two months) were organized and molecular analyses were conducted to assess the relationships between genera and use the phylogenetic framework as a guideline to undertake taxonomic revisions. The preliminary results revealed the para-/polyphyly of several Malagasy genera (e.g., *Neotina*, *Tina* and *Tinopsis*) and even contested the status of tribes as defined by Radlkofer (1933). Such kind of unexpected results avoided to investigate further taxonomic treatments at regional levels and required a broader view of the evolution of the soapberry family. This goal was achieved by increasing the sampling and molecular

Context of this study

markers in order to produce a worldwide phylogeny of Sapindaceae. Such a task will never have been possible without extending my collaborations with other researchers such as Nadir Alvarez (University of Neuchâtel, Switzerland), Félix Forest (Royal Botanic Gardens, Kew, U.K), Isabel Sanmartín (Real Jardín Botánico, Spain), Pedro Acevedo-Rodríguez (Smithsonian Institution, U.S.A.), Johan Nylander (Stockholm University, Sweden) and Mark Harrington (James Cook University, Australia). These multiple collaborations were not limited to the exchange of plant material but also allowed to greatly improve my scientific knowledge and opened never expected areas of investigations.

References

- Buerki, S., Phillipson, P.B., Callmander, M.W., A taxonomic revision of *Gouania* Jacq. (Rhamnaceae) in Madagascar and the other islands of the western Indian Ocean. Ann. Missouri Bot. Garden (in press).
- Capuron, R., 1969. Révision des Sapindacées de Madagascar et des Comores. Mém. Mus. Natl. Hist. Nat. Ser. B Bot. 19, 1-189.
- Phillipson, P.B., Schatz, G.E., Lowry II, P.P., Labat, J.-N., 2006. A catalogue of the vascular plants of Madagascar. In: Ghazanfar, S.A., Beentje, H.J. (Eds), Taxonomy and Ecology of African Plants, their Conservation and Sustainable Use. Proceedings of the 17th AETFAT Congress, Addis Ababa, Ethiopia. Royal Botanic Gardens, Kew, pp. 613—627.
- Radlkofer, L., 1933. Sapindaceae. In: Engler, A. (Ed.), Das Pflanzenreich IV, 165 (Heft 98a-h),. Leipzig, Verlag von Wilhelm Engelmann.

Introduction

An overview of the diversity of the angiosperms – Diversification of flowering plants (angiosperms) represents one of the greatest terrestrial radiations in the recent geological time (Davies et al., 2004). Nowadays, approximately 250,000 extant species, growing in all types of ecosystems, have been described (Soltis et al., 2005). However, this high level of diversity is strongly threatened by the contemporaneous human-mediated global change, through deforestation and changes in land use (Achard et al., 2002; Brooks et al., 2002).

Nonetheless, although tropical ecosystems increasingly suffer from this strong human

pressure, biologists have only a partial knowledge of the biodiversity they comprise and new taxa (species and even genera) are regularly published (see Box). As a consequence, some researchers argue that the total number of angiosperms might be twice the estimation (Bramwell, 2002). The high

Box. The flora of Madagascar

Despite less than ten percent of primary tropical forests remain in Madagascar (Myers et al., 2000; Achard et al., 2002), an impressive number of new taxa are discovered each year (some journals are even devoted to the publication of new taxonomic entities, e.g. *Adansonia*, *Candollea*, *Novon*). Since more than a decade, researchers from the university of Neuchâtel (in collaboration with the Missouri Botanical Garden; USA), organize fieldwork missions in Madagascar. As a result of these missions, new species were recognized in various groups of flowering plants: e.g., Balsaminaceae (Fischer and Rahelivololona, 2007), Pandanaceae (Laivao et al., 2006, 2007; Callmander et al., 2008a), Rhamnaceae (Callmander et al., 2008b; Buerki et al., in press).

diversity of the angiosperms and their ability to spread out over all latitudes and altitudes is highly correlated with the development of key innovations (for instance, the flower, closure of the carpel, double fecundation and the fruit; Frohlich and Chase, 2007), some of these innovations being implicated in coevolutionary processes with animals (Labandeira, 2006; Hu et al., 2008; Crepet and Niklas, 2009).

Recent progress in molecular methods (e.g., popularization of PCR and sequencing techniques) and bioinformatics (e.g., development of phylogenetic algorithms able to handle large data sets and specific models of evolution) have given birth to a large range of applications: for more than twenty years, molecular phylogenies have been used to assess evolutionary (e.g., character evolution, biogeographic history, divergence time analyses) and systematic (e.g., circumscription of taxonomic entities) hypotheses, at all scales of evolution

Introduction

[from the within-species level (e.g., Buerki et al., 2009; Burnier et al., in press) to that of major phyla (e.g., Goloboff et al., 2009)].

Since the publication of the first tree of angiosperm families based on molecular data (Chase et al., 1993), a group of scientists (i.e., the angiosperm phylogeny group; hereafter APG) was created with the aim of producing large-scale angiosperm phylogenies based on nuclear and plastid markers (Soltis et al., 2000; APGII, 2003). Investigations by the APG have also focused on infra-familial phylogenetic analyses, which allowed, for instance, to unravel taxonomically problematic families. For example, the Scrophulariaceae family was shown to be highly paraphyletic and taxa traditionally included in this family were redistributed in other families, mainly Plantaginaceae and Orobanchaceae (Olmstead and Reeves, 1995; Olmstead et al., 2001; Oxelman et al., 2005). Beside assessing relationships within angiosperms, the multiple gene data set gathered by the APG became a major source of inspiration to investigate new areas in plant phylogenetics. For example, patterns of plant-insect coevolution were investigated, addressing the major role played by the interaction between pollinator and flowering plants in their respective synchronous diversification processes (Hu et al., 2008). Although paleobotany and phylogenetics first remained disconnected, studies relying on the molecular clock hypothesis have gradually started to include fossil data into molecular phylogenies, in order to infer the timing of divergence among taxonomic groups. This has led to the estimation of divergence times and rates of evolution in angiosperms (e.g., Wikström et al., 2001; Davies et al., 2004; Magallón and Sanderson, 2001; Magallón and Castillo, 2009). Nevertheless, the assignment of fossils to specific clades remained somewhat problematic (see Magallón et al., 1999). By considering one undisputed fossil taxon (used to fix the split between Fagales and Cucurbitales at 84 Ma), Wikström et al. (2001) were the first to assess divergence time estimations for the angiosperms as a whole: their analysis estimated the origin of the extant angiosperms to be Early to Middle Jurassic [179-158 Ma; corresponding to the split between the ANITA clade *sensu* APGII (2003) and the other families], and the origin of eudicots between Late Jurassic to mid Cretaceous (147-131 Ma). Despite a conservative calibration point, both estimates

were older than fossils records (Magallón et al., 1999). However, this ancient origin of eudicots can be paralleled by paleobotanic investigations: when studying the species diversity of vascular plants through time Niklas et al. (1985) noticed an overall increase of diversity in the Cretaceous. They advocated that this phenomenon was caused by a gradual increase in angiosperm species and a decline of other plant lineages. While this pattern was also observed by complementary studies (e.g., Lidgard and Crane, 1988, 1990), there has been only little focus on explaining the divergence of the main angiosperm families in a spatio-temporal framework.

In order to bridge the gap between paleogeography and angiosperm evolution, we focus here on the study of the highly diverse and worldwide family Sapindaceae, to address whether or not the Cretaceous radiation of angiosperms is compatible with the most recent biogeographic paradigms (Ree and Sanmartín, 2009).

Studying Sapindaceae at a large scale of time and space to test the main biogeographic paradigms – For a few decades, the evolutionary history of major lineages within angiosperms has been investigated (Magallón and Castillo, 2009). The respective importance of biotic and abiotic factors in shaping the current distribution of flowering plants has also been widely discussed, although mainly on the basis of conceptual means. It is only in the last few years that hypotheses were empirically tested. The particular complexity of this kind of analyses at large scales of time and space might explain the poor level of investigations of this topic. In this thesis, I focus on the middle sized worldwide soapberry family (Sapindaceae: Sapindales) as a case study to investigate some of the recently discussed biogeographic paradigms (see Ree and Sanmartín, 2009 for a review) such as the continental location and timing of the diversification in major angiosperm families. Sapindaceae family comprises ca. 1900 species (Acevedo-Rodríguez, pers. comm.) predominantly distributed in pantropical regions with the occurrence of some taxa in temperate areas (e.g., *Acer*, *Aesculus*). Since the first worldwide treatment of Sapindaceae *sensu stricto* (Radlkofer, 1890, 1933), the circumscription of the family as well as the

relationships among subfamilial entities has been widely challenged. Several authors kept the predominantly temperate families Aceraceae and Hippocastanaceae separate from the rest of Sapindaceae (e.g., Müller and Leenhouts, 1976; Takhtajan, 1987, Cronquist, 1988; Dahlgren, 1989), whereas others lumped these families with Sapindaceae (e.g., Umadevi and Daniel, 1991; Judd et al., 1994; Gadek et al., 1996; Savolainen et al., 2000; Thorne, 2000, 2007; APGII, 2003). A recent molecular phylogeny of Sapindaceae *sensu lato* (including Aceraceae and Hippocastanaceae) inferred from plastid markers recognized the subdivision of Sapindaceae into four lineages (Harrington et al., 2005): Xanthoceroideae (including only *Xanthoceras sorbifolia*), Hippocastanoideae (including Aceraceae, Hippocastanaceae and *Handeliiodendron*), Dodonaeoideae and Sapindoideae (circumscriptions of the last two subfamilies were slightly modified; see Harrington et al., 2005 and Thorne, 2007 for more details). However, the sampling and the number of markers used were not sufficient to assess the relationships among and within the major lineages of the family with confidence. Moreover, phylogenetic relationships between families within the Sapindales (Eurosid II; APGII, 2003; Soltis et al., 2005) also remained weakly supported and additional analyses involving broader sampling and more markers are required (Gadek et al., 1996; APGII, 2003; Muellner et al., 2007; Fig. 1).

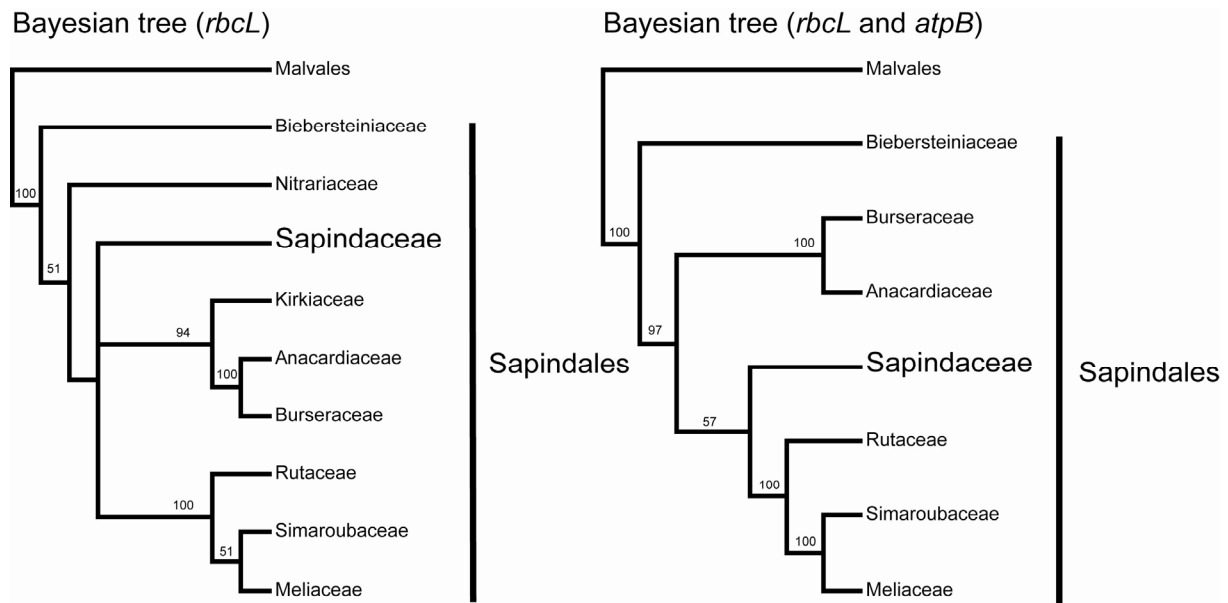


Figure 1. Phylogenetic relationships between families within the Sapindales inferred from plastid markers. Numbers above branches are Bayesian posterior probabilities. Adapted from Muellner et al. (2007).

With the inclusion of Aceraceae and Hippocastanaceae in Sapindaceae, no synapomorphies are currently known for the recognition of the family (Harrington et al., 2005; Thorne, 2007). The span of morphological characters expressed in the soapberry family is somehow representative of the diversity of the angiosperms. Sapindaceae might have a shrub, tree or liana habit; alternate or opposite, caducous or sempervirent, simple or compound (paripinnate or imparipinnate) leaves; hermaphrodite or unisexual, actinomorphic or zygomorphic flowers with more or less specialized nectar discs; and dehiscent or indehiscent fruits; a developed arillode or not (if developed, the arillode might be fleshy or not, coloured or translucent) (see Radlkofer, 1933 for more details). Moreover, a reduction of the number of carpels (from three to one) and ovules (from eight to one) is observed (Radlkofer, 1933). Palynological characters were also investigated by Biesboer (1975) and Müller and Leenhouts (1976). They recognized eight major pollen types (A to H) and several subtypes (e.g., type-A1); however none were restricted to taxonomic entities (with few exceptions

characterizing genera; e.g., type-F restricted to *Diplopeltis*). A survey of the morphological diversity encountered within Sapindaceae is depicted in figure 2.

Angiosperms divergence time analyses estimated the origin of Sapindaceae between the Late Eocene and Early Oligocene (Wikström et al., 2001), whereas an analysis focusing on Sapindales (Muellner et al., 2007) provided an older origin of the family sometime in the Late Cretaceous or Early Paleocene (between 86.3 and 65.7 Myr). The oldest macrofossils of Sapindaceae are fruits and leaves of *Acer* and *Aesculus* discovered in North America from the Late Paleocene (Crane et al., 1990; Manchester, 2001). From the Eocene to Miocene, several pollen and macrofossils were described: e.g., *Allophylus*, *Cardiospermum*, *Dipteronia* and *Pometia* (Adema et al., 1994; Manchester, 1999; Manchester and Tiffney, 2001).

Figure 2. A. *Macphersonia gracilis* (Madagascar): shrub with bi-paripinnate leaves; B. *Cardiospermum halicacabum* (widespread all over the tropics): liana, elliptic capsule and seeds partially covered by a dry white arillode; C. *Tinopsis tamatavensis* (Madagascar): tree (>20 m high), paripinnate leaves, actinomorphic flowers and young indehiscent fruits; D. *Plagioscyphus louvelii* (Madagascar): shrub with cauliflorous inflorescences; E. *Acer pseudoplatanus* (widespread in northern temperate regions, especially in Europe): actinomorphic flowers and samara; F., H. *Xanthoceras sorbifolia* (endemic of temperate forest of China and Korea): large actinomorphic flower with horn-like appendages, schizocarp with 6-8 seeds; G. Zygomorphic flower of *Ungnadia speciosa* (from Florida to Texas); I. 3-locular capsule of *Koelreuteria paniculata* (Asia); J. *Tina striata* (Madagascar): 2-locular capsule, seeds partially covered with a dry orange arillode; K. *Blighia sapida* (Africa): capsule, seeds partially covered by a fleshy toxic, white arillode; L. 3-locular capsule of *Molinaea petiolaris* (Madagascar); M. *Lepisanthes sambiranensis* (Madagascar): indehiscent fruit (notice the aborted ovary), seed without arillode and N. *Tinopsis macrocarpa* (Madagascar): indehiscent fruit and seed fully recovered by a fleshy translucent arillode.

Introduction



Fig. 2. See legend above.

Aims of this study – The aims of this study are divided into three parts. The first part is devoted to the inference of a phylogenetic hypothesis for Sapindaceae based on plastid and nuclear markers with the proposition of an informal classification (chapter 1). Subsequently, this framework was used to compare biogeographic methods and infer the history of Sapindaceae. This was done by comparing the state-of-the-art in biogeographic algorithms [e.g., Bayes-DIVA (Nylander et al., 2008) and Lagrange (Ree and Smith, 2008)] and developing methods that consider divergence time uncertainties and a biogeographic model based on paleogeography (chapter 2). Finally, evidence presented in chapter 2 was used to infer the evolutionary history of the family and assess the impact of biotic and abiotic factors on the worldwide expansion of Sapindaceae (chapter 3). The second part of this study was more taxonomically oriented and provided a contribution to the systematics of Malagasy Sapindaceae. In this part, a new Malagasy genus, *Gereaua*, segregated from *Haplocoelum* is described (chapter 4) and a revision of the Malagasy *Lepisanthes* is proposed (chapter 5). Finally, in the third part of this study, an extended data set of the soapberry family (comprising more than 240 specimens and ca. 70% of the generic diversity) was used to compare the performance of supertree algorithms and provide an update of the informal classification proposed in chapter 1 (chapter 6). See below for a detailed summary of the main results obtained in this study.

Part I: Towards a new classification of Sapindaceae based on molecular and morphological evidence and the biogeographic history of this worldwide family

Chapter one: *Plastid and nuclear DNA markers reveal intricate relationships at the subfamilial and tribal levels in the soapberry family (Sapindaceae)*

In this chapter, the infrafamilial classification of Sapindaceae and its relationships to the closely related Aceraceae and Hippocastanaceae – which have now been included in an expanded definition of Sapindaceae (APGII, 2003; Harrington et al., 2005; Thorne, 2007) – are investigated. This is done by analysing a representative subset of Sapindaceae (85 of the 141 genera) based on eight markers from the plastid and nuclear genomes. Phylogenetic

results highlight a high level of paraphyly and polyphyly at the subfamilial and tribal levels and even contest the monophyletic status of several genera (e.g., *Cupaniopsis*, *Haplocoelum*, *Matayba*). This study confirms that the monotypic Chinese genus *Xanthoceras* is sister to the rest of the family, in which subfamily Hippocastanoideae is sister to a clade comprising subfamilies Dodonaeoideae and Sapindoideae. On the basis of the strong support demonstrated in Sapindoideae, Dodonaeoideae and Hippocastanoideae as well as in 14 subclades, informal groupings are proposed as the basis for a new classification of Sapindaceae.

This study also confirms difficulties to amplify molecular markers in Sapindaceae s.l. caused by mutations occurring in flanking regions of widely used plastid and nuclear regions such as *matK* (Harrington et al., 2005) and the ITS region (Edwards and Gadek, 2001). Those substitutions complicate the compilation of multilocus data sets without missing data. The impact of missing data on phylogenetic analyses was thus investigated by comparing two data sets (both with the same number of taxa, but by different numbers of markers). The data set presenting missing data (18.6%) had a topology fully congruent with the one obtained from a subset without missing data, but including fewer markers. It is therefore shown that the use of additional information led to a consistent result in the relative position of clades and allowed the definition of a new phylogenetic hypothesis. In the last chapter of this thesis, an extended data set of the soapberry family (containing a substantial amount of missing data but many more specimens) is investigated based on supertree and supermatrix approaches.

Chapter two: *New insights on parametric biogeography based on the worldwide soapberry family (Sapindaceae)*

After the establishment of the vicariance paradigm and the rise to prominence of the cladistic biogeographic school, the field of historical biogeography is going through an extraordinary revolution concerning its methods, underlying assumptions, and the kind of questions it aims to answer (Ree and Sanmartín, 2009). These methods are based on fully probabilistic

evolutionary models and are not constrained by the inherent biases of the parsimony (cladistic) approach. They make it possible to incorporate estimates of the evolutionary divergence between lineages (branch lengths) and/or the timing of paleogeographic events into biogeographical inference, thus increasing the accuracy of biogeographic reconstructions. Previous implementations of parametric methods have been limited to small-scales studies in terms of time and of the complexity of the paleogeographic/biogeographic model (Clark et al., 2008; Santos et al., 2009). In this study, the dispersal-extinction-cladogenesis (DEC) likelihood model (Ree et al., 2005; Ree and Smith, 2008) is applied to reconstruct the biogeographic history of the soapberry family. This is done by using a complex worldwide biogeographic model that spans the last 110 Ma and reflects the changing continental configuration through time. Results from this analysis are compared with those resulting from a parsimony-based method, dispersal-vicariance analysis (DIVA; Ronquist, 1997), modified with a Bayesian empirical approach to incorporate phylogenetic uncertainty (Bayes-DIVA; Nylander et al., 2008). Results show that, despite differences in the underlying biogeographic model (i.e., the speciation mode) and the fact that DIVA does not incorporate information on branch lengths, the two methods converge on similar biogeographic histories. The main difference lies in the timing of dispersal events - which in Bayes-DIVA sometimes conflicts with paleogeographic information on the availability of land connections - and in the tendency of this method to push dispersal events to terminal branches to explain widespread terminal ranges. In contrast, Bayes-DIVA showed the highest power (decisiveness) to unequivocally reconstruct ancestral ranges, which may be related to its ability to integrate the uncertainty in the phylogeny through the use of the Bayesian posterior distribution. The biogeographic reconstructions suggest that Sapindaceae originated in Eurasia around the Early Cretaceous, from which they dispersed to North America and proto-SE Asia shortly thereafter. From there, they colonized Africa, Madagascar, and Australia-South America using the Gondwana connection.

Chapter three: *Abiotic factors as driving force to promote speciation in Sapindaceae*

In this chapter, state-of-the-art parametric biogeographic methods were used to assess the influence of biotic and abiotic factors on the diversification of the soapberry family. Most recent research has concluded that abiotic factors – intense volcanic activities, meteoritic impacts or interactions between tectonic and orbital forces – were involved in large-scale extinction processes. For instance, the Cretaceous-Tertiary climate change caused the loss of ca. 60% of plant species as well as a majority of animals, including dinosaurs. In contrast, diversification processes in flowering plants are frequently explained by biotic factors (e.g., diversification of natural enemies or mutualists). In this study, we show, for the first time, that abrupt climatic change in the Eocene-Oligocene boundary triggered the diversification of Sapindaceae. This paradigm-breaking result is mainly due to the geological and climatic properties of South East Asia that favoured multiple contacts between lineages and further speciation across the Laurasian and Gondwanian continents.

Part II: Taxonomic contributions to the systematics of Sapindaceae

Chapter four: *Molecular phylogenetic and morphological evidence support recognition of Gereaua, a new endemic genus of Sapindaceae from Madagascar*

The worldwide phylogeny of Sapindaceae presented in chapter 1 segregated the Malagasy *Haplocoelum perrieri* Capuron from the African *Haplocoelum foliosum* (Hiern) Bullock. In this chapter, additional phylogenetic analyses are conducted by including material of *H. inopleum* Radlk., the type of the genus. This analysis confirms the result obtained in chapter 1 and shows that maintaining a broad circumscription of *Haplocoelum* to include the Malagasy species would render the genus polyphyletic. In order to maintain monophyly, it is therefore necessary to exclude *H. perrieri*, which is transferred to a new, monotypic genus, described as *Gereaua* Buerki & Callm INED. Synapomorphies supporting this taxon and relationships with the most-closely related genera are discussed in light of molecular, morphological and biogeographic evidence.

Chapter five: *A synoptic revision of the genus Lepisanthes Blume (Sapindaceae) in Madagascar*

During recent fieldwork missions organized by the Missouri Botanical Garden in the highly diversified but endangered Galoka and Kalabenono massifs (northwestern Madagascar), a new species of *Lepisanthes* was discovered. In order to provide a generic assessment of the genus *Lepisanthes* for the catalogue of vascular plants of Madagascar (Phillipson et al., 2006), a taxonomic revision of this genus was undertaken. Three species are recognized based on an analysis of morphological characters in combination with eco-geographic parameters. Two infraspecific taxa recognized previously in *Aphania senegalensis* (Juss. ex Poir.) Radlk. [= *Lepisanthes senegalensis* (Juss. ex Poir.) Leenh.] are raised to the species level, viz. *L. chrysotricha* (Capuron) Buerki, Callm. & Lowry INED and *L. perrieri* (Capuron) Buerki, Callm. & Lowry INED. A third species from northwestern Madagascar is described as new, *L. sambiranensis* Buerki, Callm. & Lowry INED. An identification key to the Malagasy species of *Lepisanthes* is presented as well as preliminary assessments of the conservation status for each species.

Part III: An extended data set of the soapberry family to compare supertree methods and provide an update of the familial classification

Chapter six: *Comparative performance of supertree algorithms in large data sets using the soapberry family (Sapindaceae) as a case study*

This chapter focuses on the one hand on methodological aspects and on the other hand, provides an extended phylogenetic hypothesis for Sapindaceae (based on 240 specimens representing more than 70% of the generic diversity) by applying supertree and total evidence algorithms. Supertrees represent comprehensive phylogenetic hypotheses incorporating all taxa present in partially overlapping input trees. With the growing popularity of supertree methods and the overwhelming dominance of the matrix representation with parsimony method (MRP), it becomes necessary to evaluate the performance of the different methods (ca. 14 supertree methods; see Wilkinson et al., 2005) in the context of large data

sets to determine if the supertree algorithms work well and which are the most suitable in the case of large data sets. In this study, the comparative performance of six major methods is investigated for the first time [MRP, MinFlip, Average Consensus (AVCON), Most Similar Supertree (MSS), Split fit and MinCut] by using the soapberry family data set. The performance of each supertree method is assessed according to its compatibility with the input trees (reconstructed with maximum parsimony and maximum likelihood criteria) and with total evidence trees (reconstructed using the same algorithms as the input trees). Results indicate that the behaviour of the supertree methods is not influenced by the algorithm used to reconstruct the input trees. In contrast, results demonstrate a large variation in the performance of the different methods: only three methods – MRP, MinFlip and MinCut – performed properly, whereas the others largely failed in producing accurate topologies. However, the poor performance of AVCON, MSS and Split fit methods could be related to the efficiency of the heuristic searches rather than to the supertree algorithms themselves. Despite the addition of 90 ingroup specimens compared to the data set presented in the chapter 1 (corresponding to an increase of 14 genera) topologies of the maximum likelihood total evidence tree as well as that of the MinFlip supertree were highly congruent with the phylogenetic hypothesis presented in chapter 1. Within Hippocastanoideae, the addition of *Billia* and *Handeliidendron* and of more species of *Aesculus* supported the monophyly of the previously recognized families Aceraceae and Hippocastanaceae (including *Handeliidendron*). Moreover, both analyses recognized the inclusion of *Diplokeleba* (previously included in Cupanieae; Sapindoideae) within the Dodonaeoideae. The ten groups defined in Sapindoideae were confirmed and phylogenetic analyses assigned the monotypic American genus *Ungnadia* to the Delavaya group (which previously included only the monotypic Chinese genus *Delavaya*). Finally, the inclusion of additional genera of tribes Cossinieae and Koelreuterieae revealed the paraphyletic status of both tribes.

References

- Achard, F., Eva, H.D., Stibig, H.-J., Mayaux, P. H., Gallego, J., Richards, T., Malingeau, J.-P., 2002. Determination of deforestation rates of the world's humid forest. *Science* 297, 999—1002.
- Adema, F., Leenhouts, P.W., van Welzen, P.C., 1994. Sapindaceae. *Flora Malesiana Ser. I* 11, 419–768.
- Angiosperm Phylogeny Group (APG II), 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APGII. *Bot. J. Linn. Soc.* 141, 399–436.
- Biesboer, D.D., 1975. Pollen morphology of the Aceraceae. *Grana* 15, 19-27.
- Bramwell, D., 2002. How many plant species are there? *Plant Talk* 28, 32–34.
- Brooks, T. M., Mittermeier, R.A., Mittermeier, C.G., Da Fonseca, G.A.B., Rylands, A.B., Konstant, W.R., Flick, P., Pilgrim, J., Oldfield, S., Magin, G., Hilton-Taylor, C. 2002. Habitat loss and extinction in the hot spots of biodiversity. *Conservation Biology* 16, 909—923.
- Buerki, S., Phillipson, P.B., Callmander, M.W., A taxonomic revision of *Gouania* Jacq. (Rhamnaceae) in Madagascar and the other islands of the western Indian Ocean. *Ann. Missouri Bot. Garden* (in press).
- Buerki, S., Callmander, M.W., Schüpfer, F., Rakotovao, M., Küpfer, P., Alvarez, N., 2009. Malagasy *Dracaena* Vand. ex L. (Ruscaceae): an investigation of discrepancies between morphological features and spatial genetic structure at a small evolutionary scale. *Plant. Syst. Evol.* 280.
- Burnier, J., Buerki, S., Arrigo, N., Küpfer, P., Alvarez, N. Genetic structure and evolution of Alpine polyploid complexes: *Ranunculus kuepferi* (Ranunculaceae) as a case study. *Mol. Ecol.* (in press).
- Callmander, M.W., Buerki, S., Wohlhauser, S., 2008a. A new threatened species of Pandanaceae from Northwestern Madagascar, *Pandanus sermolliana*. *Novon* 18, 421—424.

- Callmander, M.W, Phillipson, P.B., Buerki, S., 2008b. Révision du genre *Bathiorhamnus* Capuron (Rhamnaceae) endémique de Madagascar. *Adansonia*, sér. 3 30, 151—170.
- Capuron, R., 1969. Révision des Sapindacées de Madagascar et des Comores. *Mém. Mus. Natl. Hist. Nat. Ser. B Bot.* 19, 1-189.
- Chase, M.W. (and 41 others) 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Ann. Missouri Bot. Garden* 80, 528—580.
- Clark, J.R., Ree, R.H., Alfaro, M.E., King, M.G., Wagner, W.L., Roalson, E.H., 2008. A comparative study in ancestral range reconstruction methods: retracing the uncertain histories of insular lineages. *Syst. Biol.* 57, 693—707.
- Crane, P.R., Manchester, S.R., Dilcher, D. L., 1990. A preliminary survey of fossil leaves and well-preserved reproductive structures from the Sentinel Butte Formation (Paleocene) near Almont, North Dakota. *Fieldiana, Geol., New Ser.* 20, 1—63.
- Crepet, W. L., Niklas, K. J., 2009. Darwin's second "abominable mystery": Why are there so many angiosperm species? *Am. J. Bot.* 96, 366—381.
- Cronquist, A., 1988. *The evolution and classification of flowering plants*, ed. 2. New York: New York Botanic Gardens.
- Dahlgren, G., 1989. An updated system of classification. *Bot. J. Linn. Soc.* 100, 197—203.
- Davies, T.J., Barraclough, T.G., Chase, M.W., Soltis, P.S., Soltis, D.E., Savolainen, V., 2004. Darwin's abominable mystery: insights from a supertree of the angiosperms. *Proc. Natl. Acad. Sci. USA* 101, 1904—1909.
- Edwards, K.J., Gadek, P.A., 2001. Evolution and biogeography of *Alectryon* (Sapindaceae). *Mol. Phylogenet. Evol.* 20, 14-26.
- Fischer, E., Rahelivololona, M.E., 2007. New taxa of *Impatiens* (Balsaminaceae) from Madagascar. V. New species of *Impatiens* from Masoala Peninsula. *Adansonia*, sér. 3 29, 317—332.
- Frohlich, M.W., Chase, M.W., 2007. After a dozen years of progress the origin of angiosperms is still a great mystery. *Nature* 450, 1184—1189.

- Gadek, P.A., Fernando, E.S., Quinn, C.J., Hoot, S.B., Terrazas, T., Sheahan, M.C., Chase, M.W., 1996. Sapindales: molecular delimitation and infraordinal groups. *Am. J. Bot.* 83, 802–811.
- Goloboff, P.A., Catalano, S.A., Mirande, J.M., Szumik, C.A., Arias, J.S., Källersjö, M., Farris, J.S., 2009. Phylogenetic analysis of 73 060 taxa corroborates major eukaryotic groups. *Cladistics*. (in press).
- Harrington, M.G., Edwards, K.J., Johnson, S.A., Chase, M.W., Gadek, P.A., 2005. Phylogenetic inference in Sapindaceae sensu lato using plastid *matK* and *rbcL* DNA sequences. *Syst. Bot.* 30, 366-382.
- Hu, S., Dilcher, D.L., Jarzen, D.M., Taylor, D.W., 2008. Early steps of angiosperm-pollinator coevolution. *Proc. Natl. Acad. Sci. USA* 105, 240—245.
- Judd, W.S., Sanders, R.W., Donoghue, M.J., 1994. Angiosperm family pairs: preliminary phylogenetic analyses. *Harv. Pap. Bot.* 5, 1–51.
- Labandeira, C. C., 2006. The four phases of plant-arthropod associations in deep time. *Geologica Acta* 4, 409—438.
- Laivao, M.O., Callmander, M.W., Buerki, S., 2006. Sur les *Pandanus* (Pandaceae) à stigmates saillants de la côte est de Madagascar. *Adansonia*, sér. 3 28, 267—285.
- Laivao, M.O., Callmander, M.W., Buerki, S., 2007. Révision de *Pandanus* sect. *Foullioya* Warb. (Pandaceae) à Madagascar. *Adansonia*, sér. 3 29, 39—57.
- Lidgard, S., Crane, P.R., 1988. Quantitative analyses of the early angiosperm radiation. *Nature* 331, 344—346.
- Lidgard, S., Crane, P.R., 1990. Angiosperm diversification and Cretaceous floristic trends: A comparison of palynofloras and leaf macrofloras. *Paleobiology* 16, 77—93.
- Magallón, S., Crane, P.R., Herendeen, P.S., 1999. Phylogenetic patterns, diversity, and diversifications of eudicots. *Ann. Missouri Bot. Garden* 86, 297—372.
- Magallón, S., Sanderson, M.J., 2001. Absolute diversification rates in angiosperms. *Evolution* 55, 1762–1780.

- Magallón, S., Castillo, A., 2009. Angiosperm diversification through time. *Am. J. Bot.* 96, 349–365.
- Manchester, S.R., 1999. Biogeographical Relationships of North American Tertiary Floras. *Ann. Missouri Bot. Garden* 86, 472—522.
- Manchester, S.R., 2001. Leaves and Fruits of *Aesculus* (Sapindales) from the Paleocene of North America. *Int. J. Plant Sci.* 162, 985—998.
- Manchester, S.R., Tiffney, B.H., 2001. Integration of paleobotanical and neobotanical data in the assessment of phytogeographic history of holarctic angiosperm clades. *Int. J. Plant Sci.* 162, S19—S27.
- Muellner, A.N., Vassiliades, D.D., Renner, S. S., 2007. Placing Biebersteiniaceae, a herbaceous clade of Sapindales, in a temporal and geographic context. *Pl. Syst. Evol.* 266, 233-252.
- Müller, J., Leenhouts, P. W., 1976. A general survey of pollen types in Sapindaceae in relation to taxonomy. In: Ferguson, I.K., Müller, J. (Eds.), *The evolutionary significance of the exine*. London, Academic Press, pp. 407–445.
- Myers, N., Mittermeier, R.A., Mittermeier, C.G., Da Fonseca, G.A.B., Kent, J., 2000. Biodiversity hot spots for conservation priorities. *Nature* 403, 853—858.
- Niklas, K., Tiffney, B.H., Knoll, A.H., 1985. Patterns in vascular land plant diversification: An analysis at the species level. In: Valentine, J.W. (Ed.), *Phanerozoic diversity patterns*. Princeton, New Jersey, Princeton University Press, USA, pp. 97 – 128.
- Nylander, J.A.A., Olsson, U., Alström, P., Sanmartín, I., 2008. Accounting for Phylogenetic Uncertainty in Biogeography: A Bayesian Approach to Dispersal-Vicariance Analysis of the Thrushes (Aves: *Turdus*). *Syst. Biol.* 57, 257—268.
- Olmstead, R.G., Reeves, P. A., 1995. Evidence for the polyphyly of the Scrophulariaceae based on chloroplast *rbcl* and *ndhF* sequences. *Ann. Missouri Bot. Garden* 82, 176–193.
- Olmstead, R.G., Jansen, R.K., Kim, K.-J., Wagstaff, S.J., 2000. The phylogeny of the Asteridae s.l. based on chloroplast *ndhF* sequences. *Molec. Phylogen. Evol.* 16, 96–112.

- Oxelman, B., Kornhall, P., Olmstead, R.G., Bremer, B., 2005. Further disintegration of Scrophulariaceae. *Taxon* 54, 411—425.
- Phillipson, P.B., Schatz, G.E., Lowry II, P.P., Labat, J.-N., 2006. A catalogue of the vascular plants of Madagascar. In: Ghazanfar, S.A., Beentje, H.J. (Eds), *Taxonomy and Ecology of African Plants, their Conservation and Sustainable Use. Proceedings of the 17th AETFAT Congress, Addis Ababa, Ethiopia*. Royal Botanic Gardens, Kew, pp. 613—627.
- Radlkofer, L., 1890. Ueber die Gliederung der Familie der Sapindaceen. *Sitz. - Ber. Akad. Wiss. Munchen* 20, 105—379.
- Radlkofer, L., 1933. Sapindaceae. In: Engler, A. (Ed.), *Das Pflanzenreich IV*, 165 (Heft 98a-h), Leipzig, Verlag von Wilhelm Engelmann.
- Ree, R.H., Moore, B.R., Webb, C.O., Donoghue, M.J., 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution* 59, 2299—2311.
- Ree, R.H., Smith, S.A., 2008. Maximum Likelihood Inference of Geographic Range Evolution by Dispersal, Local Extinction, and Cladogenesis. *Syst. Biol.* 57, 4—14.
- Ree, R.H., Sanmartín, I., 2009. Prospects and challenges for parametric models in historical biogeographical inference. *J. Biogeogr.* (in press).
- Ronquist, F., 1997. Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Syst. Biol.* 46, 195—203.
- Santos, J.C., Coloma, L.A., Summers, K., Caldwell, K.P., Ree, R., Cannatella, D.C., 2009. Amazonian amphibian diversity is primarily derived from Late Miocene Andean lineages. *Plos Biology* 7, 448—461.
- Savolainen, V., Fay, M.F., Albach, D.C., Backlund, A., van der Bank, M., Cameron, K.M., Johnson, S.A., Lledó, M.D., Pintaud, J.-C., Powell, M., Sheahan, M.C., Soltis, D.E., Soltis, P.S., Weston, P., Whitten, W.M., Wurdack, K.J., Chase, M.W., 2000. Phylogeny of the eudicots: a newly complete familial analysis based on *rbcL* gene sequences. *Kew Bull.* 55, 257—309.

Introduction

Soltis, D.E., Soltis, P.S., Chase, M.W., Mort, M.E., Albach, T.D., Zanis, M., Savolainen, V., Hahn, W.H., Hoot, S.B., Fay, M.F., Axtell, M., Swensen, S.M., Prince, L.M., Kress, W.J., Nixon, K.C., and Farris, J.S., 2000. Angiosperm phylogeny inferred from 18S rDNA, rbcL, and atpB sequences. *Bot. J. Linn. Soc.* 133, 381-461.

Soltis, D.E., Soltis, P.S., Endress P.K., Chase, M.W., 2005. *Phylogeny and Evolution of Angiosperms*. Sinauer, Sunderland, M.A.

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

Thorne, R. F., 2000. The classification and geography of the flowering plants: dicotyledons of the class Angiospermae. *Bot. Rev.* 66, 441–647.

Thorne, R.F., 2007. An update classification of the class Magnoliopsida (“Angiospermae”). *Bot. Rev.* 73, 67-182.

Umadevi, I., Daniel, M., 1991. Chemosystematics of Sapindaceae. *Feddes Repertorium* 102, 607–612.

Wikström, N., Savolainen, S., Chase, M. W., 2001. Evolution of the angiosperms: calibrating the family tree. *Proc. R. Soc. Lond. B.* 268, 2211—2220.

Wilkinson, M., Cotton, J.A., Creevey, C., Eulenstein, O., Harris, S.R., Lapointe, F.J., Levasseur, C., McInerney, J.O., Pisani, D., Thorley, J.L., 2005. The shape of supertrees to come: Tree shape related properties of fourteen supertree methods. *Syst. Biol.* 54, 419—431.

Chapter one: Plastid and nuclear DNA markers reveal intricate relationships at subfamilial and tribal levels in the soapberry family (Sapindaceae)

Sven Buerki, Félix Forest, Pedro Acevedo-Rodríguez, Martin W. Callmander, Johan A. A. Nylander, Mark Harrington, Isabel Sanmartín, Philippe Küpfer and Nadir Alvarez

Published in *Molecular Phylogenetics and Evolution*

Plastid and nuclear DNA markers reveal intricate relationships at subfamilial and tribal levels in the soapberry family (Sapindaceae)

Sven Buerki^{a,*}, Félix Forest^b, Pedro Acevedo-Rodríguez^c, Martin W. Callmänder^{d,e}, Johan A.A. Nylander^f, Mark Harrington^g, Isabel Sanmartín^h, Philippe Küpfer^a, Nadir Alvarez^a

^a Institute of Biology, University of Neuchâtel, Rue Emile-Argand 11, CH-2009 Neuchâtel, Switzerland

^b Molecular Systematics Section, Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, United Kingdom

^c Department of Botany, Smithsonian Institution, National Museum of Natural History, NHB-166, Washington, DC 20560, USA

^d Missouri Botanical Garden, PO Box 299, 63166-0299, St. Louis, MO, USA

^e Conservatoire et Jardin botaniques de la ville de Genève, ch. de l'Impératrice 1, CH-1292 Chambésy, Switzerland

^f Department of Botany, Stockholm University, SE-10691, Stockholm, Sweden

^g School of Marine and Tropical Biology, James Cook University, PO Box 6811, Cairns, Qld 4870, Australia

^h Department of Biodiversity and Conservation, Real Jardín Botánico – CSIC, Plaza de Murillo 2, 28014 Madrid, Spain

ABSTRACT

The economically important soapberry family (Sapindaceae) comprises about 1900 species mainly found in the tropical regions of the world, with only a few genera being restricted to temperate areas. The infrafamilial classification of the Sapindaceae and its relationships to the closely related Aceraceae and Hippocastanaceae – which have now been included in an expanded definition of Sapindaceae (i.e., subfamily Hippocastanoideae) – have been debated for decades. Here we present a phylogenetic analysis of Sapindaceae based on eight DNA sequence regions from the plastid and nuclear genomes and including 85 of the 141 genera defined within the family. Our study comprises 997 new sequences of Sapindaceae from 152 specimens. Despite presenting 18.6% of missing data our complete data set produced a topology fully congruent with the one obtained from a subset without missing data, but including fewer markers. The use of additional information therefore led to a consistent result in the relative position of clades and allowed the definition of a new phylogenetic hypothesis. Our results confirm a high level of paraphyly and polyphyly at the subfamilial and tribal levels and even contest the monophyletic status of several genera. Our study confirms that the Chinese monotypic genus *Xanthoceras* is sister to the rest of the family, in which subfamily Hippocastanoideae is sister to a clade comprising subfamilies Dodonaeoideae and Sapindoideae. On the basis of the strong support demonstrated in Sapindoideae, Dodonaeoideae and Hippocastanoideae as well as in 14 subclades, we propose and discuss informal groupings as basis for a new classification of Sapindaceae.

Keywords: Aceraceae, Classification, Hippocastanaceae, Molecular phylogeny, Paraphyly, Polyphyly, Sapindaceae, *Xanthoceras*

1. Introduction

The soapberry family (Sapindaceae: Sapindales) comprising c. 1900 species (Acevedo-Rodríguez, personal communication), has a predominantly pantropical distribution with the occurrence of some taxa in temperate areas (e.g., *Acer*, *Aesculus*, *Atalaya*, *Diplopetalis*, *Dodonaea*). Sapindaceae include many economically important species used for their fruits [e.g., guarana (*Paullinia cupana*), litchi (*Litchi chinensis*), longan (*Dimocarpus longan*), pitomba (*Talisia esculenta*) and rambutan (*Nephelium lappaceum*)], wood [e.g., buckeyes (*Aesculus*)] or as ornamentals (*Koelreuteria*, *Ungnadia*).

The circumscription of the family as well as the relationships among subfamilial entities have been widely challenged since the

very first worldwide treatment of Sapindaceae sensu stricto (s.s.) (including subfamilies Sapindoideae and Dodonaeoideae) proposed by Radlkofer (1890, 1933; for a review see Harrington et al., 2005). For instance, several genera within the Sapindoideae (e.g., *Tinopsis* and *Plagioscyphus* from Madagascar; Capuron, 1969) were shown to be morphologically transitional between tribes described by Radlkofer (1933), which prevented the recognition of unequivocal tribes. Within Sapindaceae s.s. the higher taxonomic entities (subfamilies and tribes) were originally defined by Radlkofer (1933) based on the number and type of ovules per locule, the fruit morphology, the presence or not of an arillode, the leaf type and the cotyledon shape. On the basis of macromorphological and palynological characters, Müller and Leenhouts (1976) revised the classification of Radlkofer (1933). They recognized eight major pollen types (A–H) and several subtypes (e.g., type-A1), mainly based on their shape and characteristics of the aperture (Fig. 1).

* Corresponding author. Fax: +41 327183001.

E-mail address: sven.buerki@unine.ch (S. Buerki).

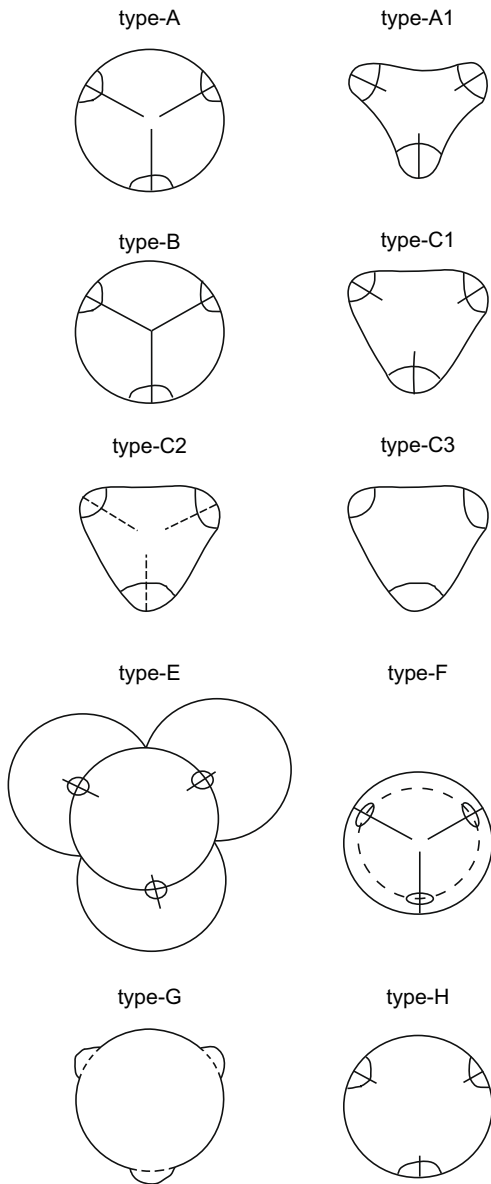


Fig. 1. Schematic representation of pollen types in Sapindaceae following Müller and Leenhouts (1976). See text for explanations regarding the morphological differentiation between pollen types.

The pollen grains in Sapindaceae are triporate [the diporate type-D pollen of *Lophostigma* recognized by Müller and Leenhouts (1976) was wrongly identified; see Acevedo-Rodríguez (1993a)]. Spherical pollen shape occurs in the majority of species (e.g., types A, B and H), whereas a triangular (type-C) or oblate (type-A1) shape is more restricted. The colpi may be absent (e.g., type-G) or parasyncolporate (e.g., type-A) to syncolporate (e.g., type-B) (Fig. 1). Based on those characters Müller and Leenhouts (1976) rearranged the nine tribes of Sapindoideae recognized by Radlkofer (1933) into three taxonomically unranked groups characterized by their distribution, the presence or absence of an arillode surrounding the seed and the pollen types [i.e., group A comprised Sapindeae, Lepisantheae (incl. Aphanieae) and Meliococceae; group B comprised Schleichereae, Nephelieae and Cupanieae; group C comprised Paulinieae and Thouinieae]. They did not, however, modify the classification within the Dodonaeoideae and maintained the five tribes described by Radlkofer (i.e., Cossinieae, Dodonaeae, Doratoxyleae, Harpullieae and Koelreuterieae, 1933). Furthermore, Müller and

Leenhouts (1976) kept the predominantly temperate families Aceraceae and Hippocastanaceae separate from the rest of Sapindaceae. The circumscription of Sapindaceae has been debated ever since. Takhtajan (1987), Cronquist (1988) and Dahlgren (1989) maintained Aceraceae and Hippocastanaceae separate from Sapindaceae, whereas broader concepts of the family have been adopted by several workers (e.g., Umadevi and Daniel, 1991; Judd et al., 1994; Gadek et al., 1996; Savolainen et al., 2000; Thorne, 2000, 2007; APGII, 2003).

Building on a large-scale molecular phylogenetic analysis of Sapindales (Gadek et al., 1996), Harrington et al. (2005) published the first molecular phylogeny of Sapindaceae sensu lato (s.l.) (including Aceraceae and Hippocastanaceae) inferred from the plastid genes *rbcL* and *matK*. Their phylogeny recognized the subdivision of Sapindaceae s.l. into four supported lineages, a monotypic Xanthoceroideae, Hippocastanoideae (including Aceraceae, Hippocastanaceae and *Handeliidendron*), a more narrowly defined Dodonaeoideae and Sapindoideae (including *Koelreuteria* and *Ungnadia*). Relationships between these four lineages remained weakly supported. Confirming previous works based on morphological features, Harrington et al. (2005) highlighted the paraphyletic or polyphyletic nature of several tribes described by Radlkofer (1933).

According to the new assessment of the Sapindaceae s.l. proposed by Thorne (2007; mainly based on Harrington et al., 2005) and a broad review of currently described taxa, it is now widely accepted that the c. 1900 species of this cosmopolitan family are divided into 141 genera (see Table 1; Acevedo-Rodríguez, personal communication). Even if Harrington et al. (2005) covered worldwide representatives of Sapindaceae s.l., the sampling (64 of the 141 genera, i.e., 45.4%) and the number of markers were not sufficient to assess the relationships among and within the major lineages of the family with confidence. In this study we provide a new assessment of the phylogenetic relationships within Sapindaceae s.l. based on 60.3% of the generic diversity (85 of the 141 genera) and including the previously unsampled tribe Cossinieae. The analysis is based on a combination of one nuclear (ITS region; *ITS1*, *5.8S*, *ITS2*) and seven plastid (coding *matK* and *rpoB*; non coding *trnL* intron and intergenic spacers *trnD-trnT*, *trnK-matK*, *trnL-trnF* and *trnS-trnG*) markers. Coding plastid regions have proven to be useful in addressing phylogenetic relationships at higher taxonomic levels (e.g., Clayton et al., 2007; Muellner et al., 2006, 2007; Harrington et al., 2005), whereas noncoding regions (introns and intergenic spacers) were shown to be more useful at lower taxonomic ranks (Baldwin, 1992; Soltis and Soltis, 1998). The combination of several markers from both nuclear and plastid genomes as well as coding and non coding regions are expected to improve the resolution of phylogenetic relationships within the family. In this study, our objectives are (1) to examine the relationships between the traditionally defined Aceraceae and Hippocastanaceae with the rest of Sapindaceae, (2) to evaluate the tribal concepts of Radlkofer (1933) and Müller and Leenhouts (1976), (3) to examine phylogenetic relationships among taxa in light of characters traditionally used to define the higher level groupings in Sapindaceae s.l. (e.g., number of ovules per locule, pollen morphology, leaf type and presence/absence of an arillode) and (4) to propose a new preliminary infrafamilial classification for Sapindaceae s.l.

In addition of being a challenging family at the taxonomic level, the amplification of molecular markers in Sapindaceae s.l. is made difficult by several mutations occurring in flanking regions of widely used plastid and nuclear regions such as *matK* (Harrington et al., 2005) and ITS (Edwards and Gadek, 2001). Those mutations complicate the compilation of multilocus data sets without missing data. Maximizing taxa and markers representation to provide a reliable phylogenetic hypothesis inferred from nuclear and plastid genomes is required to propose a new classification

Table 1
 Intrafamilial classification of Sapindaceae sensu lato (Radlkofer, 1933; Müller and Leenhouts, 1976; Thorne, 2007). Information on number of taxa, habit and distribution of genera were taken from literature (Radlkofer, 1933; Acevedo-Rodríguez, 1993a,b, 2003; Adema et al., 1994; Ferrucci 1991, 1998; Davies, 1997; Davies and Verdcourt, 1998; Klaassen, 1999; Thomas and Harris, 1999; Xia and Gadek, 2007; Mabberley, 2008). Abbreviations are as follows: s, shrub; st, small tree; t, tree; l, liana. Genera sampled for the phylogenetic analysis of Sapindaceae are indicated in bold and genera found to be either paraphyletic or polyphyletic are identified by an asterisk (*).

	Genera	Author	Taxa	Habit	Distribution	
Sapindaceae Jussieu 104/141 genera, 205/1886 species						
Dodonaeoideae Burnett Cossinieae Bl. (Cos) 2/2 genera, 3/7 species	<i>Cossinia</i>	Comm. ex Lam.	4	s-st	Mascarenes, New Caledonia, E Australia, Fiji	
	<i>Lagunoa</i>	Ruiz & Pavón	3	s-st	W tropical South America	
Dodonaeeae Kunth (Dod) 3/5 genera, 5/78 species	<i>Diplopeltis</i>	Endl.	5	s-t	NW Australia	
	<i>Distichostemon</i>	F. Muell.	6	s	Australia	
	<i>Dodonaea</i>	Miller	c. 65	s-st	Mainly in Australia, Malesia, New Guinea, Caribbean and Madagascar	
	<i>Hirania</i>	Thulin	1	s	Somalia	
	<i>Loxodiscus</i>	Hook. f.	1	s	New Caledonia	
Doratoxyleae Radlk. (Dor) 6/9 genera, 8/22 species	<i>Averrhoidium</i>	Baillon	2	t	South America	
	<i>Doratoxylon</i>	Thou. ex Hook. f.	5	st-t	Madagascar and Mascarenes Islands	
	<i>Euchorium</i>	Eckman & Radlk.	1	t	Cuba	
	<i>Exothea</i>	Macfad.	3	t	West Indies, Central America and Florida	
	<i>Filicium</i>	Thw ex Hook. f.	3	s-st	E Africa, Madagascar and SE India	
	<i>Ganophyllum</i>	Blume	2	t	W and C Africa, Andamans and Nicobars to NE Australia and Solomon Islands to Malesia	
	<i>Hippobromus</i>	Ecklon & Zeyher	1	t	South Africa	
	Dodonaeoideae Burnett Harpullieae Radlk. (Har) 6/6 genera, 8/34 species	<i>Hypelate</i>	P. Browne	1	s-st	West Indies and Florida
<i>Zanha</i>		Hiern	4	t	Tropical Africa and Madagascar	
<i>Arfeuillea</i>		Pierre ex Radlk.	1	t	SE Asia	
<i>Conchopetalum</i>		Radlk.	2	st-t	Madagascar	
<i>Eurycorymbus</i>		Handel-Mazzetti	1	t	China	
<i>Harpullia</i>		Roxb.	26	s-st	India, SE China, Malesia to Australia, New Caledonia and Pacific Islands	
<i>Magonia</i>		A. St. Hil.	1	t	South America	
<i>Majidea</i>		J. Kirk ex Oliver	3	t	Tropical Africa and Madagascar	
Hippocastanoideae Burnett 5/5 genera, 18/129 species	<i>Acer</i>	L.	111	s-t	N temperate & tropical mountains	
	<i>Aesculus</i>	L.	13	t	SE Europe, India, E Asia and N America	
	<i>Billia</i>	Peyr.	2	s-t	S Mexico to Tropical South America	
	<i>Dipteronia</i>	Oliver	2	s-st	C&S China	
	<i>Handeliodendron</i>	Rehder	1	s-t	China – deciduous	
Sapindoideae Burnett Cupanieae Reichenb. (Cup) 36/48 genera, 79/462 species	<i>Amesiodendron</i>	Hu	1	t	China, Indo-China and Malesia	
	<i>Aporrhiza</i>	Radlk.	6	t	Tropical Africa	
	<i>Arytera</i>	Blume	c. 28	s-t	Indo-Malesia to E Australia and Pacific	
	<i>Blighia</i>	Koenig	4	t	Tropical Africa	
	<i>Blighiopsis</i>	Van der Vecken	1	t	Tropical Africa	
	<i>Blomia</i>	Miranda	1	t	Mexico	
	<i>Cnesmocarpon</i>	Adema	4	s-st	Australia and Papua New Guinea	
	<i>Cupania</i>	L.	c. 45	s-t	Tropical America	
	<i>Cupaniopsis</i>*	Radlk.	60	s-st	Malesia, New Guinea, N–E Australia, Pacific islands, New Caledonia	
	<i>Dictyoneura</i>	Blume	3	s-st	Malesia	
	<i>Dilodendron</i>	Radlk.	1	t	South America	
	<i>Diploglottis</i>	Hook.f.	12	t	NE Australia and Papua New Guinea	
	<i>Diplokeleba</i>	N.E. Br.	2	st	South America	
	<i>Elattostachys</i>	(Blume) Radlk.	c. 20	s-t	Malesia to Australia, W Pacific	
	<i>Eriocoelum</i>	Hook. f.	c. 10	t	Tropical Africa	
	Sapindoideae Burnett Cupanieae Reichenb. (Cup)	<i>Euphorianthus</i>	Radlk.	1	t	E Malesia
		<i>Gloeocarpus</i>	Radlk.	1	t	Philippines
<i>Gongrodiscus</i>		Radlk.	3	s-t	New Caledonia	
<i>Gongrospermum</i>		Radlk.	1	t	Philippines	
<i>Guioa</i>*		Cav.	65	s-t	SE Asia, Malesia to E Australia; Pacific and New Caledonia	
<i>Haplocoelopsis</i>		F.G. Davies	1	s-t	E Africa	
<i>Jagera</i>		Blume	2	t	New Guinea and Australia	

Table 1 (continued)

	Genera	Author	Taxa	Habit	Distribution
	Laccodiscus	Radlk.	4	s-st	W Africa
	Lepiderema	Radlk.	8	t	Australia and New Guinea
	Lepidopetalum	Blume	7	s-t	India, NE Australia and Solomon Islands
	<i>Lynchodiscus</i>	Radlk.	6	t	W Tropical Africa
	Matayba	Aublet.	c. 56	s-t	Tropical America
	<i>Mischarytera</i>	(Radlk.) H. Turner	3	t	Australia, Papua New Guinea
	Mischocarpus	Blume	15	s-t	SE Asia, Malesia to Australia
	Molinaea	Comm. ex. Juss.	9	s-t	Madagascar, Mascarenes
	Neotina	Capuron	2	t	Madagascar
	Paranephelium	Miq.	4	s-t	SE Asia and W Malesia
	<i>Pavieasia</i>	Pierre	3	t	S China, N Vietnam
	<i>Pentascyphus</i>	Radlk.	1	t	Guyana
	<i>Phyllotrichum</i>	Thorel ex Lecompte	1	t	SE Asia
	Pseudima	Radlk.	3	t	South America
	Rhysotoechia	Radlk.	14	s-t	Australia, New Guinea, Malesia
	Sarcopteryx	Radlk.	12	s-t	Malesia, New Guinea and E Australia
	Sarcotoechia	Radlk.	11	t	NE Australia and New Guinea
	<i>Scyphonychium</i>	Radlk.	1	t	NE Brazil
	<i>Sisyrolepis</i>	Radlk.	1	s-st	Thailand
	Storthocalyx	Radlk.	4	s	New Caledonia
	Synima	Radlk.	2	t	Australia and SE New Guinea
	Tina	Roem. & Schult.	6	s-st	Madagascar
	Toechima	Radlk.	7	t	Australia and New Guinea
	<i>Trigonachras</i>	Radlk.	8	t	Malesia
	<i>Tripterodendron</i>	Radlk.	1	t	Brazil
	Vouarana	Aublet.	1	t	NE South America
Sapindoideae Burnett Koelreuterieae Radlk. (Koe) 2/4 genera, 2/15 species	<i>Erythrophyssa</i>	E. Mey ex Arnott	9	s	Africa and Madagascar
	Koelreuteria	Laxmann	3	t	S China, Japan
	<i>Sinoradlkofera</i>	F.G. Mey	2	st	China and N Vietnam
	<i>Stocksia</i>	Benth.	1	s	E Iran, Afghanistan
Lepisantheae Radlk. (Lep) 4/10 genera, 7/97 species	<i>Chonopetalum</i>	Radlk.	1	t	Tropical W Africa
	Chytranthus	Hook. f.	c. 30	st	Africa
	Glennia	Hook. f.	8	t	Tropical Africa, Madagascar, Sri Lanka, Malesia
	Lepisanthes	Blume	24	s-t	Tropical Africa, Madagascar, S-SE Asia, Malesia and NW Australia
	<i>Namataea</i>	D.W. Thomas & D.J. Harris	1	st	Cameroon
	Pancovia	Willd.	c. 13	st	Tropical Africa
	<i>Placodiscus</i>	Radlk.	c. 15	t	Tropical W Africa
	<i>Pseudopancovia</i>	Pellegrin	1	t	Tropical W Africa
	<i>Radlkofera</i>	Gilg.	1	s-st	Tropical Africa
	<i>Zollingeria</i>	Kurz	3	t	SE Asia and Malesia
Melicocceae Blume (Mel) 5/5 genera, 8/67 species	<i>Castanospora</i>	F. Muell.	1	t	NE Australia
	Melicoccus	P. Browne	10	t	Tropical America
	Talisia	Aublet	52	s-t	Tropical America
	<i>Tristira</i>	Radlk.	1	t	Malesia
	Tristiropsis	Radlk.	3	t	Pacific Ocean, Australia, Solomon Islands and Malesia
Nepheleae Radlk. (Nep) 11/12 genera, 15/77 species	Alectryon	Gaertn.	c. 30	s-st	E Malesia, Australia, New Zealand, New Caledonia, to Hawaii
	Cubilia	Blume	1	t	Malesia
	Dimocarpus	Lour.	6	s-t	S and SE Asia and Australia
	Litchi	Sonn.	1	t	Tropical China to W Malesia
	Nephelium	L.	22	t	SE Asia and Malesia
	<i>Otonephelium</i>	Radlk.	1	t	India
Sapindoideae Burnett Nepheleae Radlk. (Nep)	Pappea	Eckl. & Zeyh.	1	s-t	Tropical E to S Africa
	Podonephelium	Baillon	4	s-t	New Caledonia
	Pometia	Forst. & Forst.	2	t	Malesia and Pacific Islands
	<i>Smelophyllum</i>	Radlk.	1	t	South Africa
	<i>Stadmania</i>	Lam.	6	t	Tropical E Africa, S Africa, Madagascar and Mascarenes Islands
	Xerospermum	Blume	2	s-st	Indochinese Peninsula and Malesia
Paullinieae Kunth (Pau) 4/7 genera, 15/466 species	Cardiospermum	L.	c. 12	l	Tropical and subtropical America; 1 sp. extending to Africa
	<i>Houssayanthus</i>	Hunz.	3	s-l	South America
	<i>Lophostigma</i>	Radlk.	2	l	South America
	Paullinia	L.	c. 200	l	Tropical America and one pantropical sp.
	Serjania	Miller	c. 226	l	Tropical America
	<i>Thinouia</i>	Triana & Planchon	9	l	Tropical America
	Urvillea	Kunth	14	l	Tropical America

(continued on next page)

Table 1 (continued)

	Genera	Author	Taxa	Habit	Distribution
Sapindeae DC (Sap) 3/7 genera, 12/89 species	<i>Atalaya</i>	Blume	12	st	Australia, New Guinea and S Africa
	<i>Deinbollia</i>	Schumach. & Thonn.	c. 40	t	Tropical Africa and Madagascar
	<i>Hornea</i>	Baker	1	s-t	Mauritius
	<i>Porocystis</i>	Radlk.	2	s-t	Tropical South America
	<i>Sapindus</i>	L.	13	t	Tropical to warm temperate regions
	<i>Thouinidium</i>	Radlk.	7	s-t	Mexico and West Indies
	<i>Toulicia</i>	Aublet	14	t	South America
Schleichereae Radlk. (Sch) 8/12 genera, 12/55 species	<i>Beguea</i>	Capuron	1	t	Madagascar
	<i>Bizonula</i>	Pellegrin	1	t	Tropical Africa
	<i>Camptolepis</i>	Radlk.	4	t	E Africa and Madagascar
	<i>Chouxia</i>	Capuron	6	s-st	Madagascar
	<i>Haplocoelum</i> *	Radlk.	c. 6	st-t	Tropical Africa and Madagascar
	<i>Lecaniodiscus</i>	Planch. ex Benth.	3	st	Tropical Africa
	<i>Macphersonia</i>	Blume	8	s-t	Tropical E Africa and Madagascar
	<i>Plagioscyphus</i>	Radlk.	10	st-t	Madagascar
	<i>Pseudopteris</i>	Baill.	3	s	Madagascar
	Sapindoideae Burnett Schleichereae Radlk. (Sch)	<i>Schleichera</i>	Willd.	1	t
<i>Tinopsis</i>		Radlk.	11	t	Madagascar
<i>Tsingya</i>		Capuron	1	t	Madagascar
Thouinieae Bl. (Tho) 6/6 genera, 10/285 species	<i>Allophylus</i>	L.	c. 250	s-st-l	Pantropical
	<i>Athyana</i>	(Griseb.) Radlk.	1	t	South America
	<i>Bridgesia</i>	Bertero ex Cambess.	1	s-st	Chile
	<i>Diatenopteryx</i>	Radlk.	2	t	South America
	<i>Guindilia</i>	Hook & Arn.	3	s	South America
	<i>Thouinia</i>	Poit.	28	l	Mexico and West Indies
Sapindoideae unplaced taxa 2/2 genera, 2/2 species	<i>Delavaya</i>	Franchet	1	s-st	SW China and N Vietnam
	<i>Ungnadia</i>	Endl.	1	s-st	S North America
Xanthoceroideae Thorne & Reveal 1/1 genera, 1/1 species	<i>Xanthoceras</i>	Bunge	1	s-st	N-NE China and Korea

for family Sapindaceae. This was achieved by analysing two data sets based on the same taxa, but including different levels of missing data (i.e., different number of markers). While the inclusion of missing data was widely recognized as a major drawback in phylogenetic analyses during the early 90s (e.g., Huelsenbeck, 1991; Wiens and Reeder, 1995), recent simulations (Wiens, 1998, 2003, 2006) and empirical analyses (Baptiste et al., 2002; Driskell et al., 2004; Phillippe et al., 2004) have shown that taxa comprising high levels of missing data could be accurately placed in phylogenies. Moreover, adding incomplete taxa to a phylogenetic analysis was even shown to improve the accuracy of a given topology, e.g. by subdividing misleading long branches (Wiens, 2005). However, there is a strong heterogeneity in the ability of the different phylogenetic algorithms for managing data sets with substantial levels of missing data (Wiens, 2006), with maximum parsimony performing poorly compared to model-based algorithms such as maximum likelihood and Bayesian inference (Wiens, 2005, 2006).

2. Material and methods

2.1. Taxon sampling

Species names, voucher information, and GenBank accession numbers for all sequences are provided in the Appendix. The sampling strategy was designed to encompass the majority of subfamilies, tribes and genera of the family as recognized by the existing classifications of Radlkofer (1933), Müller and Leenhouts (1976) and Thorne (2007). Ingroup sampling comprised 152 specimens representing 60.3% of the generic diversity (85 of the 141 genera; 28 of the 57 missing genera in this analysis are monospecific; Table 1). The outgroup included Anacardiaceae (*Sorindeia* sp.; defined as outgroup in all analyses; Savolainen et al., 2000; Muellner et al., 2007) and Simaroubaceae (*Harrisonia abyssinica*). Silica-gel dried samples (Chase and Hills, 1991) were collected in the field by the authors and complemented with materials from the DNA banks

of the Missouri Botanical Garden (St. Louis, USA), the Royal Botanic Gardens, Kew (London, UK) and the James Cook University (Cairns, Australia).

2.2. DNA sequencing

Samples from the collections of the Missouri Botanical Garden and field collected samples were extracted in the laboratory of Evolutionary Botany at the University of Neuchâtel (Switzerland) using the QIAGEN DNeasy plant kit (Qiagen, Hilden, Germany) and following the manufacturer's protocol. Samples from the collections of the Royal Botanic Gardens, Kew, were extracted using the 2× cetyltrimethylammonium bromide (CTAB) procedure of Doyle and Doyle (1987) with minor modifications (see Muellner et al., 2005) followed by additional purification using a caesium chloride/ethidium bromide gradient (1.55 g/ml) and a dialysis procedure. The samples from James Cook University (Cairns, Australia) were extracted with the CTAB procedure of Doyle and Doyle (1987).

Seven plastid DNA regions and one nuclear ribosomal DNA region were amplified. Primers for the plastid regions are those described in Edwards and Gadek (2001) for *matK* (specific primer for the Dodonaeoideae were designed by Harrington et al., 2005) and the *trnK-matK* intergenic spacer (IGS), the DNA barcoding project (<http://www.kew.org/barcoding/update.html>) for *rpoB*, Demesure et al. (1995) for the *trnD-trnT* IGS, Taberlet et al. (1991) for *trnL* intron and *trnL-trnF* IGS, and Hamilton (1999) for *trnS-trnG* IGS. Primers for the ITS region are described in White et al. (1990) and additional primers were designed by Edwards and Gadek (2001) for Sapindaceae s.l.

Amplification of selected regions were achieved in a 25 µl reaction mixture containing 5 µl 5× PCR buffer, 1.5 µl 25 mM MgCl₂, 0.5 µl 10 mM dNTPs, 0.5 µl 10 mM primers, 0.2 µl GoTaq polymerase (5 U/µl) (Promega, Madison, WI, USA), and 14.5 µl ddH₂O. The amplification of the *matK* region was improved by the addition of 4% DMSO in the total volume of the PCR mix. PCR was performed

in a Biometra® T3 thermocycler. Initial denaturation was programmed for 2 min at 95 °C, followed by 35 cycles at 95 °C for 45 s, 50 °C for 45 s, 72 °C for 1 min, plus a final extension of 10 min at 72 °C. PCR products were purified using the QIAquick PCR purification kit (Qiagen, Hilden, Germany) and fluorescent sequencing was performed by Macrogen, Inc. (Seoul, South Korea) with the same primers used for PCR amplification.

2.3. Alignment

The program Sequencher version 4.1 (Gene Codes Corp., Ann Arbor, Michigan, USA) was used to assemble complementary strands and verify software base-calling. The eight regions were initially aligned individually with ClustalX (Thompson et al., 1997), and thereafter manually adjusted with the program Bioedit (Hall, 1999) using the similarity criterion (Morrison, 2006). The program Concatenate (Alexis Criscuolo, <http://www.lirmm.fr/~criscuolo/>) was used to construct two combined matrices, differing in the number of markers considered and in the level of missing data (see below).

2.4. Phylogenetic analyses

2.4.1. Single-gene analyses

Individual phylogenetic analyses and their corresponding bootstrap analyses were performed using the maximum likelihood (ML) and maximum parsimony (MP) criteria. Each partition and the combined data sets were analyzed using parsimony ratchet (Nixon, 1999) as implemented in PAUPrat (Sikes and Lewis, 2001). Based on recommendations by Nixon (1999), ten independent searches were performed with 200 iterations and 15% of the parsimony informative characters perturbed. The shortest equally most parsimonious trees were combined to produce a strict consensus tree. To assess the support at each node, non parametric bootstrap analyses (Felsenstein, 1985) were performed using PAUP* version 4.0b10 (Swofford, 2002) with 1000 replicates, SPR branch swapping, simple sequence addition, MULTREES and holding 10 trees per replicate. We used SPR branch swapping because it has been shown to be twice as fast as TBR and results in support percentages that are not significantly different (Salamin et al., 2003).

Model selection for each partition was assessed using Modeltest version 3.7 (Posada and Crandall, 1998) and the Akaike information criterion (Akaike, 1973). ML analyses were performed using RAxML version 7.0.0 (Stamatakis, 2006; Stamatakis et al., 2008) with a 1000 rapid bootstrap analyses followed by the search of the best-scoring ML tree in one single run. This analysis was done using the facilities offered by the CIPRES portal in San-Diego, USA (<http://8ball.sdsc.edu:8888/cipres-web/home>).

In this study, nodes with bootstrap supports (BS) below 50% are considered not supported, 50–74% are considered weakly supported, 75–89% are moderately supported and 90–100% are strongly supported. Topological differences between single-gene phylogenetic trees were compared using TreeJuxtaposer (Munzner et al., 2003), taking into account the level of resolution of each marker and their bootstrap supports. In this study, topological differences having a bootstrap support inferior to 75% were not considered.

2.4.2. Combined analyses

The impact of missing data on combined MP and ML phylogenetic analyses was tested based on two different combined matrices. The first matrix (hereafter named “4 markers” data set) was composed of specimens for which sequence information was available for the nuclear ribosomal ITS region and for three of the seven plastid regions (*rpoB*, *trnL* intron and *trnL-trnF* IGS).

In this combined matrix, the four remaining plastid markers were not included in order to have a complete matrix without missing data. The second combined matrix (hereafter named “4+4 markers” data set) comprised the same set of taxa as the “4 markers” data set, but also included the other four plastid markers (*matK*, *trnD-trnT* IGS, *trnK-matK* IGS and *trnS-trnG* IGS). This data set was designed to evaluate the effect of additional information on the resolution and support of topologies in comparison to the “4 markers” analyses. Taxa for which no sequences were available for a given marker were coded as missing data for the corresponding cells in the combined matrix (sensu Wiens and Reeder, 1995).

Total evidence trees (sensu Kluge, 1989) were determined using both ML and MP criteria on the two data sets using the same settings as in the single-gene analyses. Non parametric bootstrap analyses were performed for the data sets following the same settings as for the single-gene analyses. Before computing total evidence trees, an incongruence length difference (ILD) test (Farris et al., 1994) was performed as implemented in PAUP* version 4.0b10 (Swofford, 2002) with 100 replicates.

2.5. Topological congruence and impact of missing data on combined analyses

Based on analyses of the combined matrices (i.e., “4 markers” and “4+4 markers” data sets), the impact of missing data on MP and ML phylogenetic analyses was investigated (i) by assessing topological distances among trees obtained using different data sets and algorithms and (ii) by comparing taxa groupings (and clade supports) in each topology. The explicitly agree distance (Estabrook et al., 1985; Estabrook, 1992; EA distance) was calculated to evaluate the extent to which total evidence trees were compatible with each other. The EA distance quantifies the differences between trees of the same size (i.e., comprising the same number of terminal taxa). It evaluates the proportion of triplets that are resolved identically in two trees (see Wilkinson et al., 2005). EA distances were calculated using DARWIN 5 (Perrier et al., 2003). The congruence of topological groupings in analyses obtained from different data sets and algorithms was evaluated using TreeJuxtaposer (Munzner et al., 2003) and bootstrap supports of each main clade were compared.

3. Results

3.1. Alignment

The number of sequences included in each single-gene partition varied from 69 in *trnS-trnG* IGS to 154 in *rpoB*, *trnL* intron and *trnL-trnF* IGS (Table 2). For the ITS region, all specimens were sequenced, except the outgroup species *Sorindeia* sp. (i.e., 153 sequences were produced). The alignment length ranged from 363 bp in *rpoB* to 2156 bp in *trnS-trnG* IGS (Table 2). The ITS region had the highest number of variable characters (51.4%), whereas *trnS-trnG* IGS had the lowest (23.8%), even less than the coding regions *matK* and *rpoB* (29.1% and 37.2%, respectively). The same trend was recorded for the percentage of potentially parsimony-informative characters (37.8% for the ITS region and 9.0% for the *trnS-trnG* IGS; Table 2).

The combined data sets consisted respectively of 615 sequences (154 specimens; no missing data in ingroup taxa) for the “4 markers” data set, and 997 sequences (154 specimens; 18.6% missing data) for the “4+4 markers” data set (Table 2). The alignment length of the two data sets was respectively 3031 bp (“4 markers”) and 9657 bp (“4+4 markers”). The “4

markers" data set had a highest percentage of variable characters (44.7%) than the "4+4 markers" data set (37.0%). The same observations were recorded for the percentage of potentially parsimony-informative characters (30.3% for the "4 markers" and 21.2% for the "4+4 markers" data sets; Table 2). However, when considering the total amount of phylogenetic information averaged by the number of taxa, the "4+4 markers" data set showed a value more than twice higher than did the "4 markers" data set (Table 2).

3.2. Phylogenetic analyses

3.2.1. Single-gene analyses

The best-fit model for all partitions was the general time reversible (GTR) with an alpha parameter for the shape of the gamma distribution to account for among-site rate heterogeneity (Yang, 1993). The only exception was for the ITS region for which a proportion of invariable sites was added. Although the MP and ML single-gene analyses provided topologies with different levels of resolution within Sapindaceae s.l. (e.g., the MP trees were usually not resolved in several parts of the tree), no moderately to strongly supported differences (>75%) were observed between single-gene trees. In addition, the ILD test was not significant ($P = 0.9$) and indicated that the eight data sets were congruent. Those results allowed the combination of the partitions in a total evidence approach. Statistics (number of most parsimonious trees; tree length; consistency and retention indices) for each analysis are reported in Table 2.

3.2.2. Combined analyses

The most parsimonious trees for the two combined analyses under the MP criterion were respectively 5889 ("4 markers" data set) and 9843 ("4+4 markers" data set) steps. Under the ML criterion, the best-fit model for the combined matrices was GTR with a proportion of invariable sites and an alpha parameter for the shape of the gamma distribution to account for among-site rate heterogeneity (Yang, 1993). This model was used to perform the ML search (log likelihoods were -34322.2 for the "4 markers" data set and -69253.8 for the "4+4 markers" data set) followed by rapid bootstrap analyses.

3.3. Topological congruence and impact of missing data on combined analyses

The congruence (expressed by $1 - EA$ distance) between total evidence trees compiled under the ML criterion was higher (98% of common triplets between total evidence trees based on "4 markers" and "4+4 markers" data sets) than between total evidence trees obtained under MP criterion (90% of common triplets between total evidence trees based on "4 markers" and "4+4 markers" data sets) (Table 3). The MP "4 markers" total evidence tree exhibits the highest EA distances with the other total evidence trees (Table 3).

Each of the four total evidence analyses showed support for the monophyly of Sapindaceae s.l. as defined by Thorne (2007) including Aceraceae and Hippocastanaceae (Table 4). No matter which data set or algorithm were considered, the family was subdivided into three moderately to strongly-supported lineages and a fourth lineage consisting of *Xanthoceras sorbifolia*, with the following relationships: (*Xanthoceras sorbifolia*, (clade A, (clade B, clade C))) (Table 4, Fig. 2). Despite strong support for each clade, the sister position of the monotypic *Xanthoceras* was not supported in any analyses (see clade A + clade B + clade C in Table 4). This lineage corresponded to subfamily Xanthoche-roideae as described by Thorne (2007). Clade A corresponded to

Table 2 Characteristics of partitions used in the phylogenetic analyses of the Sapindaceae s.l. See text for explanations regarding the compilation of combined data sets (i.e. "4 markers" and "4+4 markers"), IGS, intergenic spacer; the asterisk (*) indicates markers included in the combined "4 markers" phylogenetic analysis; MP, maximum parsimony.

	Single-gene analysis				Combined analyses				
	ITS*	matK	rpoB*	trnD-trnT IGS	trnK-matK IGS	trnL intron*	trnL-trnF IGS*	trnS-trnG IGS	4 markers* 4+4 markers
No. of ingroup sampled species/genera	139/84	110/69	139/84	79/62	100/63	139/84	139/84	67/47	139/84
No. sequences incl. outgroup (in brackets, total number of samples for the combined analyses)	153	119	154	85	109	154	154	69	615 (154)
Sequence length range	650–705	1074–1242	357–363	1086–1425	705–753	510–522	380–430	1311–1365	–
Alignment length	1234	1614	363	1925	931	773	661	2156	9657
Missing data (percentage of ingroup sequences; in brackets percentage of nucleotides for the combined analyses)	0	21.7	0	44.1	28.3	0	0	54.6	18.6 (27.3)
No. constant characters (%)	599 (48.5)	1144 (70.9)	228 (62.8)	1096 (56.9)	530 (56.9)	489 (63.3)	359 (54.3)	1643 (76.2)	6088 (63.0)
No. variable characters (%)	635 (51.4)	470 (29.1)	135 (37.2)	829 (43.1)	401 (43.1)	284 (36.7)	302 (45.7)	513 (23.8)	3569 (37.0)
No. potentially parsimony-informative (PI) characters (%)	467 (37.8)	295 (18.3)	95 (26.2)	405 (21.0)	230 (24.7)	166 (21.5)	190 (28.7)	195 (9.0)	2043 (21.2)
Mean amount of phylogenetic information per sample (averaged by variable sites number/PI sites number)	–	–	–	–	–	–	–	–	8.8/6.0
No. trees retained (MP)	525	1991	1997	2001	1786	1707	1023	1190	1010
Tree length (MP; step)	4365	837	246	1447	790	578	576	798	5889
Consistency Index (MP)	0.282	0.701	0.707	0.731	0.675	0.681	0.680	0.741	0.504
Retention Index (MP)	0.640	0.849	0.899	0.811	0.790	0.879	0.861	0.665	0.698

Table 3

Level of topological agreement (based on EA distances) between total evidence trees inferred from the “4 markers” and “4+4 markers” data sets. See text for explanations regarding the compilation of these data sets. MP, maximum parsimony; ML, maximum likelihood.

	1	2	3	4
1- ML “4 markers”	–			
2- MP “4 markers”	0.177	–		
3- ML “4+4 markers”	0.021	0.189	–	
4- MP “4+4 markers”	0.028	0.173	0.027	–

subfamily Hippocastanoideae (including the previous recognized families Aceraceae and Hippocastanaceae) as described by Harrington et al. (2005) and Thorne (2007). Clade B corresponded to subfamily Dodonaeoideae as described by Harrington et al. (2005) and Thorne (2007) with the addition of *Euphorianthus* (Cupanieae; Sapindoideae). Clade C corresponded to subfamily Sapindoideae (Thorne, 2007; Harrington et al., 2005) plus one representative from Dodonaeoideae, *Conchopetalum*, included in tribe Harpullieae. Clade C was moderately to strongly supported as monophyletic and divided into ten groups, but not in the MP “4 markers” total evidence tree (only one exception: clade V nested in clade VI; Table 4). The bootstrap supports of each clade obtained under the ML algorithm are consistent in both data sets (Table 4), whereas support slightly increases in MP analyses, in parallel to an increase in missing data (Table 4).

The “4 markers” and “4+4 markers” topologies recognized all the classical tribes (except the Paullinieae) as paraphyletic or polyphyletic. However, phylogenetic status of tribes Cossinieae and Koelreuterieae were not tested because only one genus per tribe was considered. In total 5 of the 67 non-monotypic sampled genera (7.5%) are paraphyletic or polyphyletic (*Cupaniopsis*, *Guioa*, *Haplocoelum*, *Matayba*, *Sarcotoechia*). However, the phylogenetic status of some of these genera needs to be treated with caution because of weak bootstrap supports and limited sampling (e.g., *Guioa*).

Table 4

Summary of the bootstrap support for each clade recovered in the four total evidence trees (two data sets and two algorithms). Bootstrap supports for clade C-I are not indicated because this lineage is only composed by *Delavaya yunnanensis*. Note: Although monophyletic, clade C-V is nested into clade C-VI, the latter is not recovered by the MP analysis based on the “4 markers” data set. MP, maximum parsimony; ML, maximum likelihood.

Combined data sets	ML		MP	
	4 markers	4+4 markers	4 markers	4+4 markers
Sapindaceae s.l.	94	91	97	97
Clade A + Clade B + Clade C	65	58	60	57
Clade A	100	100	99	99
Clade B	94	91	99	99
B-I	100	100	100	100
B-II	88	77	86	86
Clade C	92	87	96	98
C-II	100	100	100 100	100
C-III	100	100		100
C-IV	77	98	<50	65
C-IV-a	73	100	73	100
C-IV-b	100	100	<50	83
C-V	100	100	99	100
C-VI	60	75	–	<50
C-VI-a	69	89	<50	65
C-VI-b	65	83	<50	58
C-VII	100	100	100	100
C-VIII	60	61	50	70
C-IX	100	100	99	100
C-X	100	100	93	100

4. Discussion

4.1. Congruence of topologies with and without missing data

Our results indicate a high level of congruence among topologies obtained using data sets with and without missing data and based on different algorithms. Considering the “4 markers” data set (without missing data), MP and ML algorithms however produced slightly different topologies regarding clades C-V and C-VI (i.e., in the MP “4 markers” tree clade C-VI is paraphyletic with the inclusion of the clade C-V, whereas all other topologies considered this clade as monophyletic; Table 4). This could be explained mostly by the small amount of phylogenetic information in the “4 markers” data set that prevent the MP algorithm to find a proper solution (averaged over the number of terminal taxa; Table 2). Although the addition of 4 markers to the data set generated 18.6% of missing data (27.3% of missing nucleotides) in the “4+4 markers” data set, the added information doubled the mean amount of potentially parsimonious-informative characters per terminal taxa and increased the bootstrap support for several nodes in the total evidence trees (Tables 2 and 4). Since our results highlight a high congruence level among topologies obtained with different data sets and algorithms, only the ML total evidence tree inferred from the “4+4 markers” data set will be discussed in order to maximize phylogenetic information (Figs. 2–6).

4.2. Phylogenetic relationships

Our results support (1) the paraphyly of the currently defined Dodonaeoideae and Sapindoideae as defined by Thorne (2007); (2) the polyphyly of all tribes (tribes Cossinieae and Koelreuterieae are not considered because only one genus per tribe was sampled) with the possible exception of Paullinieae – whose monophyletic status shall be evaluated by the inclusion of three missing genera *Houssayanthus*, *Lophostigma* and *Thinouia* in future analyses – and (3) the paraphyly or polyphyly of 5 of the 67 non-monotypic sampled genera (7.5%) included in this study (Table 1).

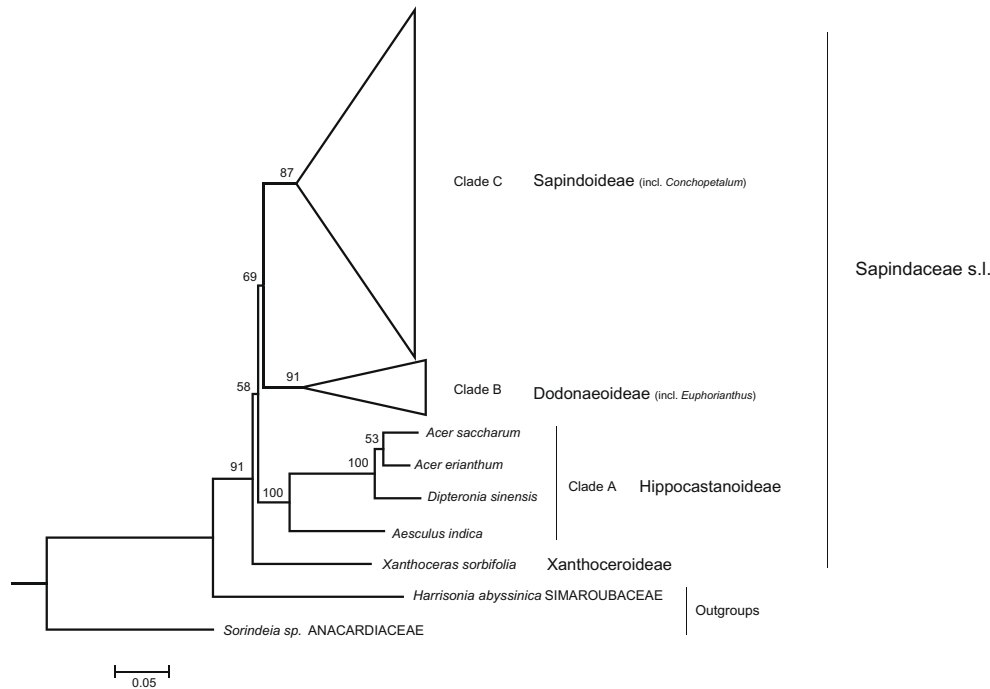


Fig. 2. Best maximum likelihood phylogenetic tree for Sapindaceae s.l. inferred from eight nuclear and plastid nucleotide sequences. Bootstrap supports are indicated above branches. The revised infrafamilial classification based on molecular and morphological characters is indicated in grey. *Abbreviations:* COS, Cossinieae; CUP, Cupanieae; DOD, Dodonaeae; DOR, Doratoxyleae; KOE, Koelreuterieae; HAR, Harpullieae; LEP, Lepisantheae; MEL, Melicocceae; NEP, Nephelieae; PAU, Paullinieae; SAP, Sapindeae; SCH, Schleichereae; THO, Thouinieae.

In light of these results, a new infrafamilial classification for Sapindaceae s.l. is required. However, we recommend caution in formally proposing new tribes until (i) non-molecular synapomorphies supporting putative new tribal delimitations are identified and (ii) the inclusion of missing genera in future phylogenetic analyses. In order to provide efficient guidelines for a new classification of the family, the phylogenetic framework obtained here is discussed according to several key morphological characters such as leaf type (including phyllotaxy), wood anatomy, number of ovules per locule, fruit type and pollen (Fig. 1), as well as geographical distribution. Hereafter, the definition of Dodonaeoideae and Sapindoideae will be expanded to include *Euphorianthus* in the former and *Conchopetalum* in the latter.

Xanthoceroideae and Hippocastanoideae occur mostly in temperate regions [except *Billia* (not included here), which occurs from Mexico to tropical South America], whereas Dodonaeoideae have a temperate (e.g., south of Australia) and tropical pattern of distribution. On the other hand, Sapindoideae have mainly radiated in tropical regions. Within Sapindaceae s.l., a trend towards the reduction of the number of ovule per locule is observed: from six to eight (Xanthoceroideae) to two (Hippocastanoideae and most of the Dodonaeoideae) and finally one (Sapindoideae except *Conchopetalum*). All four subfamilies recognized by Thorne (2007) are discussed separately below.

4.3. Subfamily Xanthoceroideae (Fig. 2)

The phylogenetic position of the monotypic Chinese *Xanthoceras* in relation to the other three main lineages of Sapindaceae remains unsupported (BS < 50) (Fig. 2; Table 4). Nevertheless, this species was moderately supported as the earliest-diverging lineage in Sapindaceae s.l. in earlier studies (*matK*, *rbcl*, Harrington et al., 2005; *rbcl*, Savolainen et al., 2000; 18S rDNA, *atpB*, *rbcl*, Soltis et al., 2000). In the first molecular phylogeny of Sapindaceae

s.l., Harrington et al. (2005) argued that an increased sampling of other monotypic Southeast Asian genera of Harpullieae (e.g., *Arfeuillea*, *Delavaya*, *Eurycorymbus*) and Koelreuterieae (*Sinoradlkofera*) might help break up possible long-branch attraction and stabilize the position of this taxon. However, our study shows that even when considering 60.3% of the generic diversity and including *Arfeuillea*, *Delavaya* and *Eurycorymbus*, the phylogenetic position of this genus remains unchanged. This small shrub is characterized by unusual features in Sapindaceae such as deciduous imparipinnate leaves (vs. deciduous simple leaves or sempervirent imparipinnate or paripinnate leaves in other Sapindaceae), six to eight fertile ovules per locule (generally 1 or 2 ovules per locule in the rest of the family) and the presence of orange horn-like appendages protruding from the disk (absent in other genera). Moreover, this species exhibits a type-A pollen which was expected to be ancestral in Sapindaceae by Müller and Leenhouts (1976) (Fig. 1). However, this pollen type is widespread across the taxa sampled in our phylogeny and is consequently of limited systematic utility.

4.4. Subfamily Hippocastanoideae (Clade A, Fig. 2)

The inclusion of Aceraceae and Hippocastanaceae in Sapindaceae has been debated for decades (e.g., Radlkofler, 1933; Müller and Leenhouts, 1976; Umadevi and Daniel, 1991; Judd et al., 1994) and both are currently included in Sapindaceae by the Angiosperm Phylogeny Group (APGII, 2003). However, the final decision regarding the taxonomic level of this well-supported clade (BS 100, Fig. 2) is somewhat dependant on the placement of *Xanthoceras sorbifolia*. Although *Billia* and *Handeliidendron*, thought to be close relative of *Aesculus* (Xiang et al., 1998; Forest et al., 2001), were not sampled here, the analysis confirms the definition of Hippocastanoideae as previously suggested by Judd et al. (1994) and Harrington et al. (2005). This temperate clade is charac-

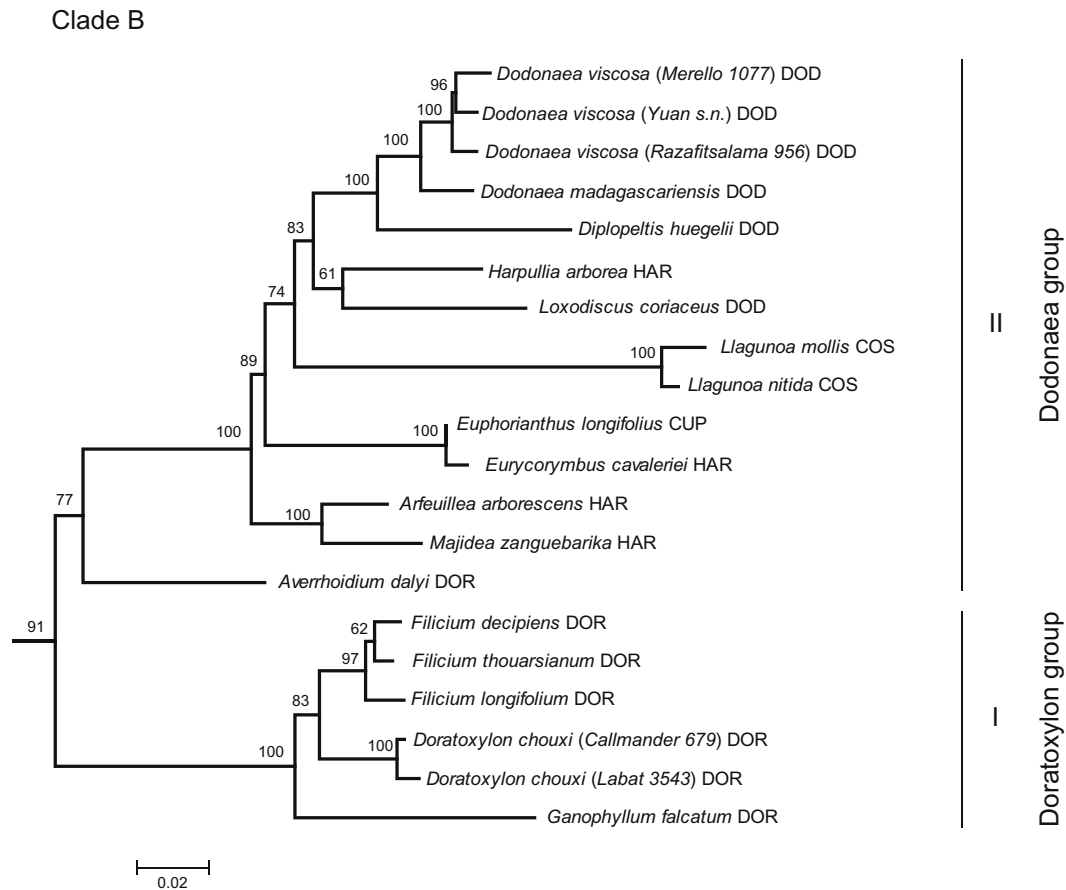


Fig. 3. Relationships within subfamilies Hippocastanoideae (clade A) and Dodonaeoideae (clade B). Bootstrap supports are indicated above branches. The revised infrafamilial classification based on molecular and morphological characters is indicated in grey. See Fig. 2 for abbreviations of tribes.

terized by deciduous opposite simple leaves (generally palmatilobate), two ovules per locule and a type-A pollen (Biesboer, 1975; Müller and Leenhouts, 1976, Fig. 1).

4.5. Subfamily Dodonaeoideae (Clade B, Figs. 2 and 3)

The improved sampling for subfamily Dodonaeoideae (i.e., the addition of genera *Arfeuillea*, *Averrhoidium*, *Doratoxylon*, *Euphorianthus*, *Eurycorymbus*, *Llagunoa* and *Majidea*) allows the recognition of two moderately to well-supported clades (Fig. 3, Table 4). This topology was partially recovered by Harrington et al. (2005), but the addition of new taxa allow their delimitation based on fruit morphology: clade I (Doratoxylon group) occurs from Africa, Madagascar to Australasia and is characterized by indehiscent berry-like fruits, whereas clade II (Dodonaea group) is distributed in South America, Madagascar, Australasia and the Pacific islands (*Dodonaea viscosa* had a worldwide distribution) and comprises species with dehiscent fruits. In addition to the widespread type-A pollen occurring in both clades, specialized pollen types characterizing specific taxa occur in clade II [i.e., type-F (*Diplopeltis huegelii*) and type-H (*Harpullia cupanoides*)] (George and Erdtman, 1969; Müller and Leenhouts, 1976, Fig. 1). Clades I and II have generally two ovules per locule; however a reduction to one ovule per locule occurs independently in the two clades (*Filicium* in clade I and *Euphorianthus* in clade II). Moreover, a few species of *Harpullia* (clade II), such as *H. arborea*, have 1-2 ovules per locule (Adema et al., 1994).

4.6. Subfamily Sapindoideae (clade C, Figs. 2 and 4–6)

4.6.1. Early-diverging lineages (Fig. 4)

Subfamily Sapindoideae is by far the most diverse lineage in terms of species. Based on our analyses, we propose to divide it into ten groups that are discussed in light of their morphological features, geographical distribution and compared to tree topologies obtained by Harrington et al. (2005) (Figs. 4–6). The *Delavaya* group is the first lineage to diverge in Sapindoideae (clade I). Only the Chinese monotypic genus *Delavaya* is included in the present study. Results from Harrington et al. (2005) highlighted the Mexican and Texan genus *Ungnadia* (from which nuclear sequences were unavailable) as the most basal lineage in Sapindoideae. Combined plastid analyses (Buerki, unpublished data) revealed a close-relationship between those two genera as suggested by Judd et al. (1994; based on morphological characters); however this relationship must be further examined using nuclear sequences. The *Delavaya* group is characterized by elongated petal base appendages and glabrous stamens (Judd et al., 1994) and the wood anatomy within the group is identical to the Cupanieae (Klaassen, 1999). The *Koelreuteria* group (clade II, BS 100), here comprising only *Koelreuteria*, is distributed in southern China and western Pacific. The study of Harrington et al. (2005) revealed a close-relationship between this genus and *Smellophyllum* and *Stadmania*, distributed in East-Africa, Madagascar and the Mascarene archipelago. When a broad definition is considered, the *Koelreuteria* group shows both ancestral (type-A pollen; Müller and Leenhouts, 1976, Fig. 1) and

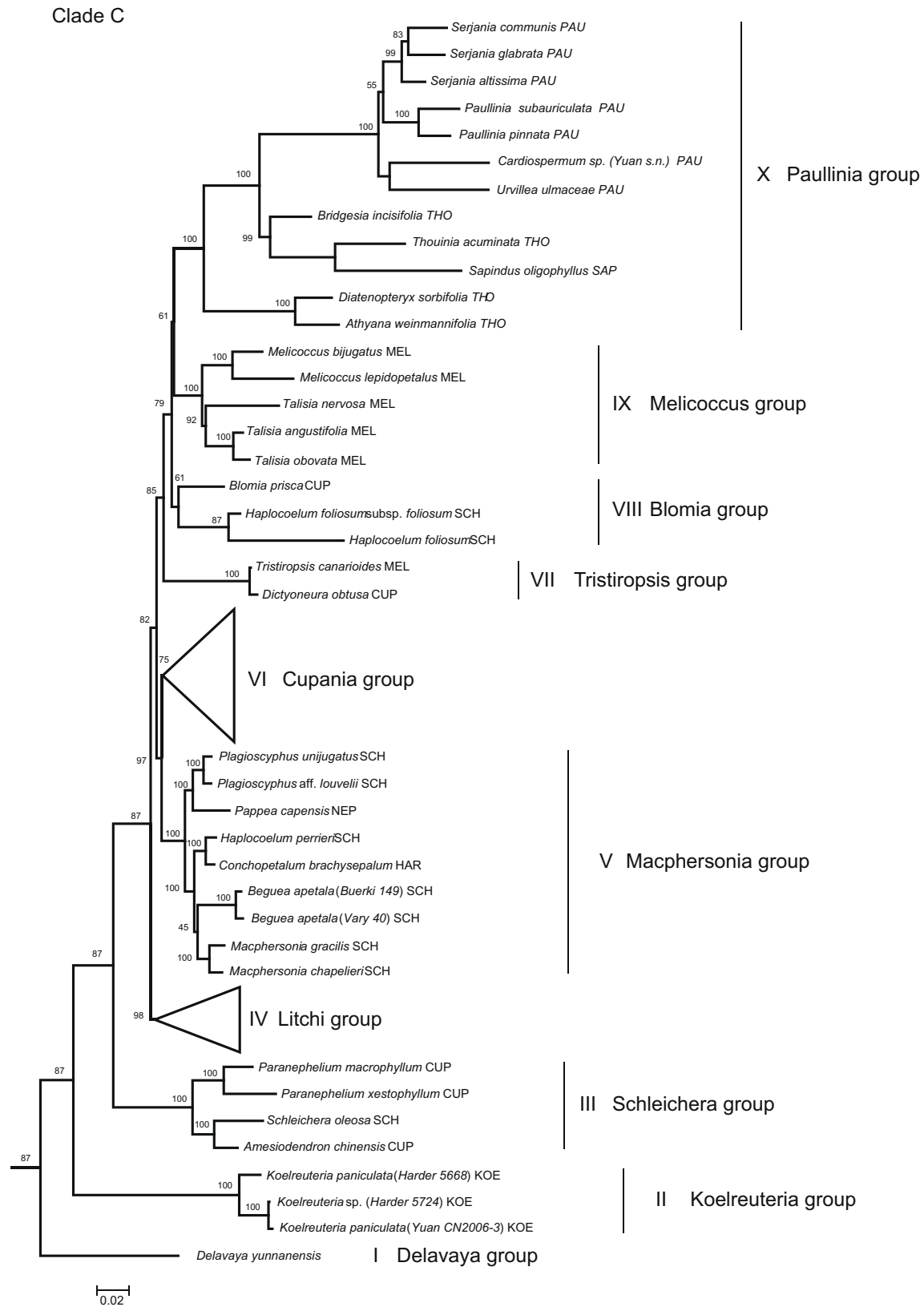


Fig. 4. Relationships within subfamily Sapindoideae (clade C). Bootstrap supports are indicated above branches. The revised infrafamilial classification based on molecular and morphological characters is in grey. See Fig. 2 for abbreviations of tribes.

derived characters (one ovule per locule in *Smellophyllum* and *Stadmaniana*) and is characterized by the presence of trichomes on the anther. Since these two lineages show a disjunct distribution and

transitional character states, they might be relicts of early diversification events in the subfamily (caused by long distance dispersals for example). The *Schleichera* group, which is partially recovered by

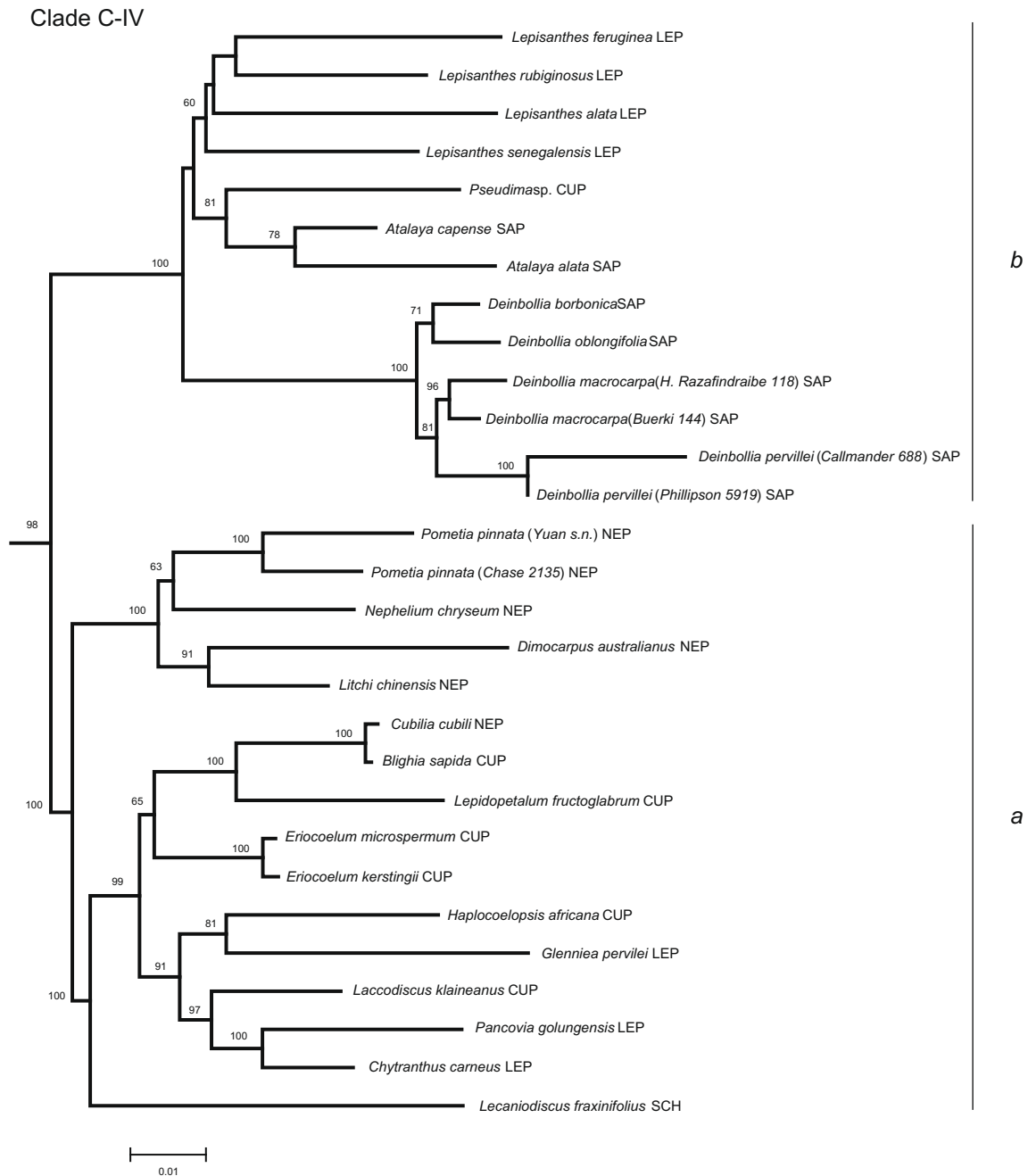


Fig. 5. Phylogenetic relationships within the *Litchi* group (clade C-IV; see Fig. 4). Bootstrap supports are indicated above branches. See Fig. 2 for abbreviations of tribes.

Harrington et al. (2005), here with the inclusion of *Amesiodendron* (Cupanieae), is a well-supported (BS 100) tropical Asian clade (clade III, Fig. 4). This clade is characterized by a Cupanieae-like wood anatomy (Klaassen, 1999) and type-B pollen (Müller and Leenhouts, 1976, Fig. 1).

4.6.2. The *Litchi* group (Figs. 4 and 5)

This clade (clade IV, BS 98, Fig. 4) is divided into two well-supported groups (*a* and *b*; Fig. 5). Clade *a* (BS 100) partially corresponds to the *Dimocarpus* group proposed by Müller and Leenhouts (1976; traditionally comprising *Cubilia*, *Dimocarpus*, *Litchi*, *Nephelium*, *Pometia* and *Xerospermum*) and a heterogeneous group comprising mostly African genera as well as the Indian and Australian *Lepidopetalum*. Our study also confirms the close rela-

tionships of *Pometia* (characterized by type-C1 pollen; Müller and Leenhouts, 1976; van der Ham, 1990, Fig. 1) with the other member of the *Dimocarpus* group as expected by Müller and Leenhouts (1976). The Lepisantheae-type wood anatomy of *Eriocoelum* (Cupanieae; Klaassen, 1999) confirms its relationships with the other genera of Lepisantheae from this clade. A more comprehensive analysis of this clade is currently being undertaken (Buerki, unpublished data).

Clade *b* (BS 100) partially corresponds to group A of Müller and Leenhouts (1976) with the addition of *Pseudima* (Cupanieae). The inclusion of the South American *Pseudima* is supported by type-A pollen (Müller and Leenhouts, 1976, Fig. 1) and similar wood anatomy shared with other Sapindeae (Klaassen, 1999). Our results highlight the close affinities of *Lepisanthes*, *Sapindus* and *Atalaya*,

Clade C-VI

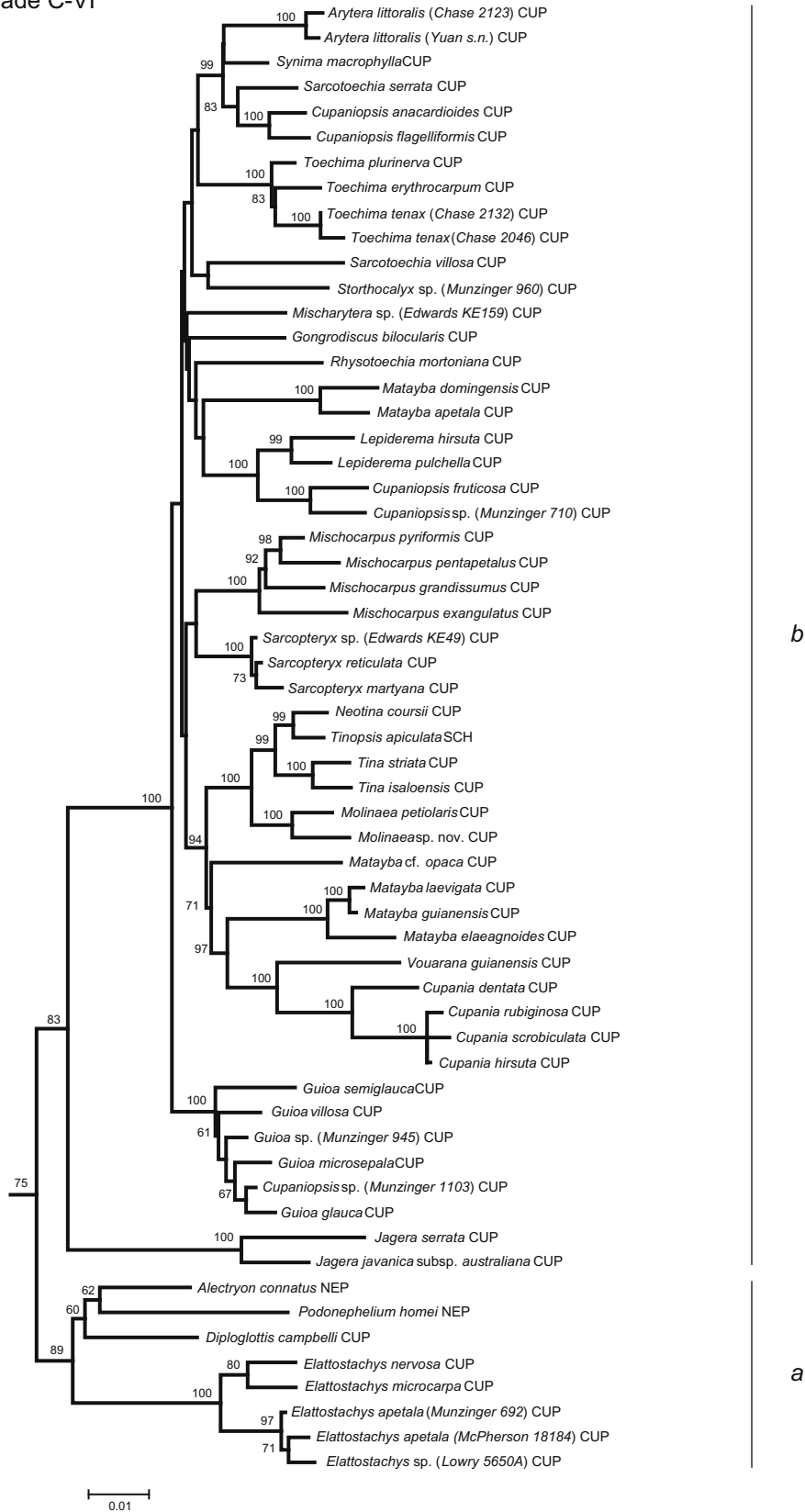


Fig. 6. Phylogenetic relationships within the *Cupania* group (clade C-VI; see Fig. 4). Bootstrap supports are indicated above branches. See Fig. 2 for abbreviations of tribes.

but the understanding of relationships within this group will require additional data. The monophyly of the African-Malagasy *Deinbollia* is supported by molecular analyses and type-A1 pollen (Müller and Leenhouts, 1976, Fig. 1).

4.6.3. The *Macphersonia* group (Fig. 4)

Our study reveals for the first time relationships between southeast African and Malagasy genera (BS 100, Fig. 4). Two strongly supported clades were formed by South African *Pappea capensis* and Malagasy *Plagioscyphus* (BS 100) and Malagasy *Beguea*, *Conchopetalum* and *Haplocoelum perrieri*, as well as east African and Malagasy *Macphersonia* (BS 100). *Pappea* was previously thought to be related to other Nephelieae (*Alectryon*, *Podonephelium*, *Smelopyllum* and *Stadmania*) by Müller and Leenhouts (1976), and placed without support as sister to Paullinieae and Thouinieae by Harrington et al. (2005). The position of *Conchopetalum*, characterized by inflated fruits without arillode, in the traditional core Malagasy Schleichereae, defined by indehiscent fruits and a fleshy arillode surrounding the seed, was an unexpected result (Capuron, 1969). This clade is characterized by actinomorphic flowers, one ovule per locule (except two in *Conchopetalum*) and is distributed throughout Madagascar and southeast Africa.

4.6.4. The *Cupania* group (Figs. 4 and 6)

The Australasian and Malagasy/South American clade VI (BS 75, Fig. 4) encloses the majority of Cupanieae genera (23 of the 32 sampled genera) and is divided into two main groups (Figs. 4 and 6). In the Australasian clade *a* (BS 100), the monophyly of *Elattostachys* is well supported and the expected close relationship between the New Caledonian *Podonephelium* and Australasian and Pacific *Alectryon* is confirmed by this phylogenetic analysis and the shared type-A pollen (Müller and Leenhouts, 1976, Fig. 1). Only one non Cupanieae taxon belongs to clade *b* (BS 100): *Tinopsis apiculata* (Schleichereae). The Malagasy *Tinopsis* was first described as part of the Cupanieae (Radlkofer, 1933) and later transferred to the Schleichereae based on the indehiscence of the fruit and the presence of a fleshy arillode (Capuron, 1969). However, no floral or vegetative characters have been identified to discriminate this genus from the Malagasy Cupanieae genera *Tina* and *Neotina*. This study confirms the close relationships between these genera and supports Radlkofer's (1933) hypothesis. This example and others encountered in clades II and V provide strong arguments supporting the convergent evolution of fruit morphology and consequently its limited systematic utility. The plasticity of fruit types has been demonstrated in several phylogenetic studies performed on a wide range of taxa (e.g., van Welzen, 1990; Adema, 1991; Muellner et al., 2003). The Cupania group is characterized by type-B pollen (except *Alectryon* and *Podonephelium* which have type-A pollen; Müller and Leenhouts, 1976, Fig. 1). In general, taxa within clade *b* present low genetic distances among them while having long terminal branches (especially the Australasian representatives such as *Cupaniopsis*, *Gongrodiscus* and *Toechima*).

4.6.5. The *Paullinia* group and allies (*Tristiropsis*, *Blomia* and *Melicoccus* groups) (Fig. 4)

Although strongly supported in general (except for the *Blomia* group; Table 4), the relationships between these four groups remain unclear (Fig. 4). The monophyly of the Australasian clade VII and the Mexico/East African clade VIII are weakly to well-supported (BS 100 and BS 61, respectively, Fig. 4). To date, no morphological characters have been identified that circumscribe these lineages. The monophyly of the South American clade IX is well supported (BS 100, Fig. 4) and confirms the suggested affinities between *Melicoccus* and

Talisia argued by Acevedo-Rodríguez (2003) based on morphology and pollen characters.

The pantropical clade X (Fig. 4) is strongly supported (BS 100) and corresponds both to the Nomophyllae group defined by Radlkofer (1933) and to the group C proposed by Müller and Leenhouts (1976) containing Paullinieae and Thouinieae. Although no representatives of genus *Allophylus* (Thouinieae) were included here, our study confirms the results of the morphological cladistic analyses of the two tribes conducted by Acevedo-Rodríguez (1993b) and the molecular analyses of Harrington et al. (2005), which show a monophyletic Paullinieae nested in a paraphyletic Thouinieae. Our analysis indicates that the enigmatic species *Sapindus oligophyllus* has affinities with genera in this clade (Fig. 7). The generic position of this taxon has puzzled taxonomists for decades. It was first described as a member of *Aphania* and subsequently transferred in *Sapindopsis*, *Howethoa*, *Sapindus* (see Rauschert, 1982 for review) and recently merged, although informally, in *Lepisanthes* by Xia and Gadek (2007). The increase of sampling and the inclusion of *Allophylus* species might help to circumscribe the position of this taxon. Type-A pollen and the tree life-form are shared by the most basal lineages in this clade (*Athyana weinmannifolia*, *Diatenopteryx sorbifolia* and *Bridgesia incisifolia*; Acevedo-Rodríguez, 1993b, Figs. 1 and 7), whereas the other taxa have a highly specialized pollen type (type-C2-3; Müller and Leenhouts, 1976, Fig. 1) and a tendency towards liana habit. Species with subtype-C pollen do not form a monophyletic group and consequently this character is of limited systematic value (e.g., type-C3 is encountered in *Thouinia* and *Paullinia*; Müller and Leenhouts, 1976; Acevedo-Rodríguez, 1993b, Figs. 1 and 4). Clade X is characterized by zygomorphic flowers, petals with a prominent scale, an unilateral disk and imparipinnate leaves. The liana habit and the development of tendrils and stipules constitute synapomorphies for Paullinieae (Fig. 4).

4.7. Informal tribal groupings within Sapindaceae

The phylogenetic analysis inferred from eight nuclear and plastid regions provides a robust assessment of the relationships within Sapindaceae s.l. (although the relationships between the subfamilies remain weakly supported) (Fig. 2). Nevertheless, the tribal delimitations as currently defined (and based largely on fruit morphology) must be revised because of the plasticity of fruit characters in this group. When Richardson et al. (2000a,b) assessed the tribal classification of Rhamnaceae (also defined by fruit morphology), they encountered the same taxonomic difficulty and proposed a new classification based on molecular data in combination with morphological characters. We follow a similar approach and propose here an informal grouping that could serve as basis for a formal reclassification of Sapindaceae s.l. based on molecular and morphological data. The family is subdivided into four subfamilies (as recognized by Thorne, 2007) and 14 groups: Xanthoceroideae, Hippocastanoideae (two groups); Dodonaeoideae (two groups) and Sapindoideae (10 groups) (Figs. 2–4). The groups within subfamilies might represent circumscriptions for the definition of future tribes.

4.7.1. Subfamily Xanthoceroideae

It includes the monotypic Chinese *Xanthoceras sorbifolia*, this deciduous shrub is characterized by alternate imparipinnate leaves, 6–8 ovules per locule and orange horn-like appendages protruding from the disk (Fig. 2).

4.7.2. Subfamily Hippocastanoideae

Temperate deciduous shrubs and trees (except *Billia* found from Mexico to tropical South America) with simple generally

palmatilobate opposite leaves and 2 ovules per locule (Fig. 2). Although our sampling is limited for this subfamily, results from other studies (Judd et al., 1994; Harrington et al., 2005; Buerki, unpublished data) allows us to suggest two groups within subfamily Hippocastanoideae, corresponding to the formerly recognized families Aceraceae and Hippocastanaceae (Fig. 4):

- *Acer* group (*Acer* and *Dipteronia*): leaves palmately lobed to 3-foliolate or imparipinnate, or entire; actinomorphic unisexual or bisexual flowers and samara;
- *Aesculus* group (*Aesculus*, *Billia* and *Handeliodendron*): leaves palmately divided into 3–5 leaflets; zygomorphic andromonoecious flowers and dehiscent fruit with one seed.

4.7.3. Subfamily Dodonaeoideae

This subfamily is expanded to include *Euphorianthus*, formerly placed in Sapindoideae. The Dodonaeoideae as defined by Radlkofer (1890, 1933) are characterized by the presence of two or rarely more apotropous and upright ovules per locule, or rarely one ovule that is epitropous and pendulous. However, this does not hold anymore because of the inclusion of the above mentioned genus of Sapindoideae showing one campylotropous ovule per locule. This subfamily is divided into two groups (Fig. 3):

- *Doratoxylon* group (*Doratoxyleae*, without *Averrhoidium*): indehiscent berry-like fruits;
- *Dodonaea* group (*Cossinieae*, *Dodonaeae*, *Arfeuillea*, *Averrhoidium*, *Eurycorymbus*, *Euphorianthus*, *Harpullia* and *Majidea*): dehiscent fruits.

4.7.4. Subfamily Sapindoideae

The subfamily Sapindoideae should be expanded to include *Conchopetalum*, formerly placed into the Dodonaeoideae. This subfamily as defined by Radlkofer (1933) is characterized by a single apotropous and upright or ascending ovule per locule; however the inclusion of several genera with two ovules per locule [*Conchopetalum* (this study), *Delavaya*, *Koelreuteria* and *Ungnadia*; Harrington et al., 2005; Thorne, 2007; this study] renders this key-character obsolete. Based on our phylogenetic analysis, ten groups are now recognized (Fig. 4):

- *Delavaya* group (*Delavaya* and *Ungnadia*): two ovules per locule; type-A pollen; elongated basal petals appendages; glabrous stamens and Cupanieae wood anatomy.
- *Koelreuteria* group (*Koelreuteria*, *Smelophyllum* and *Stadmania*): type-A pollen and trichomes on anthers.
- *Schleichera* group (*Amesiodendron*, *Paranephelium* and *Schleichera*): type-B pollen and Cupanieae-type IV wood anatomy.
- *Litchi* group [*Lepisantheae*, *Nepheleae* (without *Alectryon*, *Pappea*, *Podonephelium*, *Stadmania*, *Smelophyllum*), Sapindeae (without *Sapindus oligophyllus*), *Blighia*, *Eriocoelum*, *Haplocoelopsis*, *Laccodiscus*, *Lecaniodiscus*, *Lepidopetalum* and *Pseudima*]: to date, no morphological characters characterizing this group have been identified.
- *Macphersonia* group (*Beguea*, *Conchopetalum*, *Haplocoelum perrieri*, *Macphersonia*, *Pappea*, and *Plagioscyphus*): actinomorphic flowers and one ovule per locule (except two in *Conchopetalum*).

- *Cupania* group (*Cupanieae* [without *Amesiodendron*, *Blighia*, *Blomia*, *Dictyoneura*, *Eriocoelum*, *Haplocoelopsis*, *Laccodiscus*, *Lepidopetalum*, *Pseudima*], *Alectryon*, *Podonephelium* and *Tinopsis*): type-B pollen (except *Alectryon* and *Podonephelium*, which demonstrate type-A pollen).
- *Tristiropsis* group (*Dictyoneura* and *Tristiropsis*): to date, no morphological characters characterizing this group have been identified.
- *Blomia* group (*Blomia* and *Haplocoelum foliosum*): to date, no morphological characters characterizing this group have been identified.
- *Melicoccus* group (*Talisia* and *Melicoccus*): pollen type-A and Melicocceae wood anatomy.
- *Paullinia* group (*Paullinieae*, *Thouinieae* and *Sapindus oligophyllus*): imparipinnate leaves; zygomorphic flowers; petals with a prominent scale and a unilateral disk.

4.8. Conclusions

This study based on eight nuclear and plastid regions and 60.3% of the generic diversity of the Sapindaceae s.l. (152 samples and 139 species) (1) provides strong support for the monophyly of the family when *Xanthoceras sorbifolia*, Aceraceae and Hippocastanaceae are included (although relationships among subfamilies are still weakly supported), (2) highlights a high degree of paraphyly and polyphyly at subfamilial and tribal level, especially in Sapindaceae s.s. (subfamilies Dodonaeoideae and Sapindoideae) and (3) proposes a new informal classification for infrafamilial arrangements. Increased sampling, filled sequence gaps and the compilation of an extensive morphological matrix are now required to establish strong synapomorphies for each phylogenetic clade. A particular attention might be given to inflorescence types (and breeding systems) and floral morphology (e.g., shape and type of petal scale, type of disk, number of carpels, pubescence on the anthers, toxicity of the arilode). This might lead to a new formal infrafamilial classification for Sapindaceae s.l., based on the patterns highlighted in this study.

Acknowledgments

The first author would like to thank Professor Giordina Bernasconi from the University of Neuchâtel (Switzerland) for help and financial support. We are grateful to the following people and institutions for the supply of plant material: the Conservatoire et Jardin botaniques de la ville de Genève (Switzerland); the Malagasy staff of the Missouri Botanical Garden in Antananarivo (Madagascar); Caroline Weckerle (University of Zürich, Switzerland); Laura Vary (University of California-Irvine; USA) and Jean-Noël Labat (Muséum d'Histoire Naturelle, Paris, France). The first author thank Yong-Ming Yuan (University of Neuchâtel, Switzerland) for providing laboratory support and plant material. We thank Nils Arrigo (University of Neuchâtel; Switzerland), Ken Wurdack (Smithsonian Institution, Washington, USA), Jérôme Munzinger (Laboratoire de Botanique, Centre IRD de Nouméa, Nouméa, New Caledonia), Peter Lowry II and Pete Phillipson (Missouri Botanical Garden, Saint-Louis; USA) for valuable comments on the manuscript and herbarium determinations. We thank the Associate Editor and two anonymous reviewers for numerous constructive comments that greatly improved the manuscript.

Appendix

Voucher information and GenBank accession numbers for taxa used in the phylogenetic analysis of family Sapindaceae s.l. (including outgroups). *Abbreviations:* ANH, Andong National University, South Korea; BBG, Bogor Botanic Garden, India, living collections; CSIRO, CSIRO Arboretum, Australia; G, Conservatoire et Jardin Botanique de la ville de Genève, Switzerland; JCT, James Cook University of Northern Queensland, Australia; K, Royal Botanic Gardens, Kew, UK; NEU, Neuchâtel, Switzerland; MO, Missouri Botanical Garden, USA; P, Muséum d'Histoire Naturelle, France; RBG, Royal Botanic Gardens, Kew, UK, living collections; US, Smithsonian Institution, USA; Z, University of Zürich, Switzerland.

Genera	Species	Author	Voucher	Herbarium	Country	GenBank Accession Nos.								
						ITS	matK	rpoB	trnD-trnT	trnK-matK	trnL	trnL-F	trnS-trnG	
Ingroup														
<i>Acer</i>	<i>erianthum</i>	Schwer.	<i>Chase 19983</i>	K	China	EU720501	–	EU720843	EU720980	–	EU721271	EU721459	–	
<i>Acer</i>	<i>saccharum</i>	Marshall	<i>Chase 106</i>	K	Cult. source, Orange Co.	EU720502	–	EU720844	–	–	EU721272	EU721460	–	
<i>Aesculus</i>	<i>indica</i>	(Wall. ex Cambess.) Hook.	<i>Chase 19987</i>	K	India	EU927392	–	EU720845	EU720981	–	EU721273	EU721461	–	
<i>Alectryon</i>	<i>connatus</i>	Radlk.	<i>Chase 2047</i>	K	Australia	EU720415	EU720577	EU720732	EU720928	EU721025	EU721169	EU721357	EU721534	
<i>Amesiodendron</i>	<i>chinensis</i>	(Merr.) Hu	<i>Yuan s.n.</i>	NEU	China	EU720403	–	EU720718	EU720917	–	EU721155	EU721344	EU721525	
<i>Arfeuillea</i>	<i>arborescens</i>	Pierre	<i>Chase 2122</i>	K	Bogor, BG	EU720461	EU720629	EU720793	EU720962	EU721067	EU721229	EU721417	–	
<i>Arytera</i>	<i>littoralis</i>	Blume	<i>Yuan s.n.</i>	NEU	China	EU720405	EU720566	EU720720	EU720919	EU721018	EU721157	EU721346	EU721527	
<i>Arytera</i>	<i>littoralis</i>	Blume	<i>Chase 2123</i>	K	Bogor, BG	EU720462	EU720630	EU720794	EU720963	EU721068	EU721230	EU721418	–	
<i>Atalaya</i>	<i>alata</i>	(Sim) H. Forbes	<i>Edwards KE228</i>	JCT	South Africa	EU720425	EU720593	EU720748	EU720939	EU721036	EU721184	EU721372	EU721543	
<i>Atalaya</i>	<i>capense</i>	R.A. Dyer	<i>Edwards KE 509</i>	JCT	South Africa	EU720429	–	EU720752	–	–	EU721188	EU721376	–	
<i>Athyana</i>	<i>weimmannifolia</i>	(Griseb.) Radlk.	<i>Pennington 17581</i>	MO	Peru	EU720487	EU720649	EU720824	EU720975	EU721086	EU721257	EU721445	EU721576	
<i>Averrhoidium</i>	<i>dalyi</i>	Acev.-Rodr. & Ferrucci	<i>Weckerle 00/03/18-1/1</i>	Z	Peru	EU720495	–	EU720836	–	–	EU721268	EU721456	–	
<i>Beguea</i>	<i>apetala</i>	Capuron	<i>Buerki 149</i>	NEU	Madagascar	EU720491	EU720652	EU720828	EU720978	EU721089	EU721261	EU721449	–	
<i>Beguea</i>	<i>apetala</i>	Capuron	<i>Vary 40</i>	MO	Madagascar	EU720512	EU720663	EU720856	–	EU721100	EU721281	EU721469	–	
<i>Blighia</i>	<i>sapida</i>	K.D. Koenig	<i>Edwards KE86</i>	JCT	West Africa	EU720416	EU720578	EU720733	EU720929	EU721026	EU721170	EU721358	EU721535	
<i>Blomia</i>	<i>prisca</i>	(Standl.) Lundell	<i>Acevedo 12242</i>	US	Mexico, Yucatan	EU720444	EU720611	EU720772	–	EU721050	EU721208	EU721396	–	
<i>Bridgesia</i>	<i>incisifolia</i>	Bertero ex Cambess.	<i>Killip & Pisano 39778</i>	K	Chile	EU720476	EU720645	EU720811	EU720973	EU721082	EU721247	EU721435	–	
<i>Cardiospermum</i>	sp.		<i>Yuan s.n.</i>	NEU	China	EU720399	–	EU720713	EU720912	–	EU721150	EU721339	–	
<i>Chytranthus</i>	<i>carneus</i>	Radlk.	<i>Chase 2868</i>	RBG	–	EU720477	EU720646	EU720812	EU720974	EU721083	EU721248	EU721436	EU721575	
<i>Conchopetalum</i>	<i>brachysepalum</i>	Capuron	<i>Rabarimanarivo 8</i>	MO	Madagascar	EU720530	EU720680	EU720877	–	EU721117	EU721299	EU721487	EU721586	
<i>Cubilia</i>	<i>cubili</i>	(Blanco) Adelb.	<i>Chase 2125</i>	K	Bogor, BG	EU720463	EU720631	EU720795	EU720964	EU721069	EU721231	EU721419	EU721567	
<i>Cupania</i>	<i>dentata</i>	DC.	<i>Acevedo 12241</i>	US	Mexico, Yucatan	EU720523	EU720670	EU720867	EU720988	EU721107	EU721289	EU721477	EU721581	
<i>Cupania</i>	<i>hirsuta</i>	Radlk.	<i>Acevedo 1101</i>	US	French Guiana	EU720521	EU720668	EU720865	–	EU721105	EU721287	EU721475	–	
<i>Cupania</i>	<i>rubiginosa</i>	(Poir.) Radlk.	<i>Mori 8868</i>	MO	French Guiana	EU720481	–	EU720817	–	–	EU721251	EU721439	–	
<i>Cupania</i>	<i>scrobiculata</i>	Rich.	<i>Acevedo 11100</i>	US	French Guiana	EU720524	EU720671	EU720868	EU720989	EU721108	EU721290	EU721478	–	
<i>Cupaniopsis</i>	<i>anacardioides</i>	Radlk.	<i>Chase 217</i>	K	Australia	EU720438	EU720605	EU720763	EU720946	EU721045	EU721199	EU721387	EU721552	
<i>Cupaniopsis</i>	<i>flagelliformis</i>	(Bailey) Radlk.	<i>Edwards KE42</i>	JCT	Australia	EU720432	EU720598	EU720755	EU720942	–	EU721191	EU721379	EU721547	
<i>Cupaniopsis</i>	<i>fruticosa</i>	Radlk.	<i>Munzinger 564</i>	MO	New Caledonia	EU720533	–	EU720881	–	EU721119	EU721302	EU721490	–	
<i>Cupaniopsis</i>	sp.		<i>Munzinger 710</i>	MO	New Caledonia	EU720532	–	EU720880	EU720996	–	EU721301	EU721489	EU721587	
<i>Cupaniopsis</i>	sp.		<i>Munzinger 1103</i>	MO	New Caledonia	EU720507	EU720660	EU720851	–	EU721097	EU721278	EU721466	–	
<i>Deinbollia</i>	<i>borbonica</i>	Scheff.	<i>Edwards KE197</i>	JCT	Tanzania	EU720412	EU720574	EU720729	–	–	EU721166	EU721354	EU721532	
<i>Deinbollia</i>	<i>macrocarpa</i>	Capuron	<i>H. Razafindraibe 118</i>	MO	Madagascar	EU720535	EU720683	EU720883	–	EU721121	EU721304	EU721492	EU721589	
<i>Deinbollia</i>	<i>macrocarpa</i>	Capuron	<i>Buerki 144</i>	NEU	Madagascar	EU720503	EU720656	EU720847	–	EU721093	EU721275	EU721463	–	
<i>Deinbollia</i>	<i>oblongifolia</i>	(E. Mey. ex Arn.) Radlk.	<i>Edwards KE233</i>	JCT	South Africa	EU720427	EU720595	EU720750	–	–	EU721186	EU721374	EU721545	
<i>Deinbollia</i>	<i>pervillei</i>	(Blume) Radlk.	<i>Phillipson 5919</i>	MO	Madagascar	EU720395	EU720560	EU720708	–	EU721012	EU721145	EU721334	–	
<i>Deinbollia</i>	<i>pervillei</i>	(Blume) Radlk.	<i>Callmander 688</i>	MO	Madagascar	EU720514	–	EU720858	–	–	EU721283	EU721471	–	

(continued on next page)

Genera	Species	Author	Voucher	Herbarium	Country	GenBank Accession Nos.								
						ITS	matK	rpoB	trnD-trnT	trnK-matK	trnL	trnL-F	trnS-trnG	
<i>Delavaya</i>	<i>yunnanensis</i>	Franch.	<i>Forrest 20682</i>	MO	China, Yunnan	EU720484	–	EU720821	–	–	EU721254	EU721442	–	
<i>Diatenopteryx</i>	<i>sorbifolia</i>	Radlk.	<i>Zardini 43371</i>	MO	Paraguay	EU720534	EU720682	EU720882	–	EU721120	EU721303	EU721491	EU721588	
<i>Dictyoneura</i>	<i>obtusa</i>	Blume	<i>Edwards KE142</i>	JCT	Australia	EU720428	–	EU720751	–	–	EU721187	EU721375	–	
<i>Dimocarpus</i>	<i>australianus</i>	Leenh.	<i>Edwards KE34</i>	JCT	Australia	EU720433	–	EU720757	–	–	EU721193	EU721381	–	
<i>Diploglottis</i>	<i>campbelli</i>	Cheel	<i>Chase 2048</i>	K	Australian, BG	EU720457	EU720624	EU720788	EU720960	EU721062	EU721224	EU721412	–	
<i>Diplopeltis</i>	<i>huegelii</i>	Endl.	<i>Chase 2192</i>	K	Australia	EU720473	EU720642	EU720807	EU720971	EU721079	EU721243	EU721431	–	
<i>Dipteronia</i>	<i>sinensis</i>	Oliv.	<i>Chase 502</i>	RBG	–	EU720445	EU720612	EU720774	–	–	EU721210	EU721398	–	
<i>Dodonaea</i>	<i>madagascariensis</i>	Radlk.	<i>Bocksberger GB028</i>	NEU	Madagascar	EU720518	–	EU720862	EU720984	–	EU721284	EU721472	–	
<i>Dodonaea</i>	<i>viscosa</i>	Jacq.	<i>Razaftsalama 956</i>	MO	Madagascar	EU720519	EU720666	EU720863	EU720985	EU721103	EU721285	EU721473	–	
<i>Dodonaea</i>	<i>viscosa</i>	Jacq.	<i>Merello 1077</i>	MO	Peru	EU720536	EU720684	EU720884	EU720997	EU721122	EU721305	EU721493	–	
<i>Dodonaea</i>	<i>viscosa</i>	Jacq.	<i>Yuan s.n.</i>	NEU	China	EU720406	EU720567	EU720721	EU720920	EU721019	EU721158	EU721347	–	
<i>Doratoxylon</i>	<i>chouxii</i>	Capuron	<i>Labat JNL3543</i>	P	Madagascar	EU720394	EU720559	EU720707	EU720908	EU721011	EU721144	EU721333	–	
<i>Doratoxylon</i>	<i>chouxii</i>	Capuron	<i>Callmander 679</i>	MO	Madagascar	EU720513	EU720664	EU720857	–	EU721101	EU721282	EU721470	–	
<i>Elattostachys</i>	<i>apetala</i>	Radlk.	<i>Munzinger 692</i>	MO	New Caledonia	EU720537	EU720685	EU720885	EU720998	EU721123	EU721306	EU721494	EU721590	
<i>Elattostachys</i>	<i>apetala</i>	Radlk.	<i>McPherson 18184</i>	MO	New Caledonia	EU720538	EU720686	EU720886	EU720999	EU721124	EU721307	EU721495	EU721591	
<i>Elattostachys</i>	<i>microcarpa</i>	S.T. Reynolds	<i>Edwards KE98</i>	JCT	Australia	EU720409	EU720571	EU720726	–	–	EU721163	EU721351	–	
<i>Elattostachys</i>	<i>nervosa</i>	(F. Muell.) Radlk.	<i>Chase 2022</i>	K	Australian, BG	EU720455	EU720622	EU720786	EU720959	EU721060	EU721222	EU721410	EU721563	
<i>Elattostachys</i>	sp.		<i>Lowry 5650A</i>	MO	New Caledonia	EU720529	EU720679	EU720876	EU720994	EU721116	EU721298	EU721486	EU721585	
<i>Eriocoelum</i>	<i>kerstingii</i>	Gilg ex Engl.	<i>Merello 1586</i>	MO	Ghana	EU720539	EU720687	EU720887	EU721000	EU721125	EU721308	EU721496	EU721592	
<i>Eriocoelum</i>	<i>micropermum</i>	Radlk.	<i>Bradley 1025</i>	MO	Gabon	EU720540	EU720688	EU720888	EU721001	EU721126	EU721309	EU721497	EU721593	
<i>Euphorianthus</i>	<i>longifolius</i>	Radlk.	<i>Chase 2126</i>	K	Bogor, BG	EU720464	–	EU720796	–	–	EU721232	EU721420	–	
<i>Eurycorymbus</i>	<i>cavalerieri</i>	(H. Lévl.) Rehder & Hand.-Mazz.	<i>Yuan s.n.</i>	NEU	China	EU720404	EU720565	EU720719	EU720918	EU721017	EU721156	EU721345	EU721526	
<i>Filicium</i>	<i>decipiens</i>	(Wight & Arn.) Thwaites	<i>Chase 2128</i>	K	Bogor, BG	EU720466	EU720633	EU720798	–	–	EU721234	EU721422	–	
<i>Filicium</i>	<i>longifolium</i>	(H. Perrier) Capuron	<i>Rabenantonandro 1113</i>	MO	Madagascar	EU720541	–	EU720889	–	–	EU721310	EU721498	–	
<i>Filicium</i>	<i>thouarsianum</i>	(A. DC.) Capuron	<i>Antilahimena 5021</i>	MO	Madagascar	EU720493	–	EU720832	–	–	EU721265	EU721453	–	
<i>Ganophyllum</i>	<i>falcatum</i>	Blume	<i>Chase 2129</i>	K	Bogor, BG	EU720467	EU720634	EU720799	–	EU721071	EU721235	EU721423	–	
<i>Glennia</i>	<i>pervillei</i>	(Baill.) Leenh.	<i>Andriamihajarivo 1053</i>	MO	Madagascar	EU720490	EU720651	EU720827	EU720977	EU721088	EU721260	EU721448	–	
<i>Gongrodiscus</i>	<i>bilocularis</i>	H. Turner	<i>Munzinger 749</i>	MO	New Caledonia	EU720542	EU720689	EU720890	–	EU721127	EU721311	EU721499	–	
<i>Guioa</i>	<i>glauca</i>	Radlk.	<i>McPherson 18230</i>	MO	New Caledonia	EU720545	EU720692	EU720893	–	EU721130	EU721315	EU721503	–	
<i>Guioa</i>	<i>microsepala</i>	Radlk.	<i>Munzinger 744</i>	MO	New Caledonia	EU720546	EU720693	EU720894	–	EU721131	EU721316	EU721504	EU721596	
<i>Guioa</i>	<i>semiglauca</i>	(F. Muell.) Radlk.	<i>Chase 2058</i>	K	Australian, BG	EU720458	EU720625	EU720789	–	EU721063	EU721225	EU721413	–	
<i>Guioa</i>	<i>villosa</i>	Radlk.	<i>McPherson 18040</i>	MO	New Caledonia	EU720544	EU720691	EU720892	EU721003	EU721129	EU721314	EU721502	EU721595	
<i>Guioa</i>	sp.		<i>Munzinger 945</i>	MO	New Caledonia	EU720505	EU720658	EU720849	–	EU721095	EU721277	EU721465	–	
<i>Haplocoelopsis</i>	<i>africana</i>	F.G. Davies	<i>Edwards KE276</i>	JCT	Tanzania	EU720441	EU720608	EU720767	EU720949	–	EU721203	EU721391	EU721555	
<i>Haplocoelum</i>	<i>foliosum</i>	(Hiern) Bullock	<i>Friis 1894</i>	MO	Ethiopia	EU720479	–	EU720815	–	–	EU721250	EU721438	–	
<i>Haplocoelum</i>	<i>foliosum</i> subsp. <i>foliosum</i>	(Hiern) Bullock	<i>Edwards KE195</i>	JCT	Tanzania	EU720410	EU720572	EU720727	EU720924	–	EU721164	EU721352	EU721530	
<i>Haplocoelum</i>	<i>perrieri</i>	Capuron	<i>Rakotomalaza 1165</i>	MO	Madagascar	EU720396	–	EU720709	EU720909	–	EU721146	EU721335	EU721519	
<i>Harpullia</i>	<i>arborea</i>	(Blanco) Radlk.	<i>Chase 1353</i>	K	Bogor, BG	EU720448	–	EU720779	–	–	EU721215	EU721403	–	
<i>Jagera</i>	<i>javanica</i>	(Blume) Blume ex Kalkman	<i>Chase 2130</i>	K	Bogor, BG	EU720468	EU720635	EU720800	–	EU721072	EU721236	EU721424	EU721569	
<i>Jagera</i>	<i>javanica</i> subsp. <i>australiana</i>	Leenh.	<i>Edwards KE178</i>	JCT	Australia	EU720442	–	EU720769	–	–	EU721205	EU721393	EU721556	
<i>Koelreuteria</i>	<i>paniculata</i>	Laxm.	<i>Harder 5668</i>	MO	Vietnam	EU720548	EU720695	EU720896	–	EU721133	EU721318	EU721506	–	
<i>Koelreuteria</i>	<i>paniculata</i>	Laxm.	<i>Yuan CN2006-3</i>	NEU	China	EU720397	EU720561	EU720710	–	EU721013	EU721147	EU721336	EU721520	

<i>Koelreuteria</i>	sp.		<i>Harder 5724</i>	MO	Vietnam	EU720547	EU720694	EU720895	EU721004	EU721132	EU721317	EU721505	-
<i>Laccodiscus</i>	<i>klaineanus</i>	Pierre ex Engl.	<i>Walters 1269</i>	MO	Gabon	EU720549	EU720696	EU720897	-	EU721134	EU721319	EU721507	-
<i>Lecaniodiscus</i>	<i>fraxinifolius</i>	Baker	<i>Edwards KE194</i>	JCT	Tanzania	EU720418	EU720580	EU720735	EU720931	EU721028	EU721172	EU721360	EU721536
<i>Lepiderema</i>	<i>hirsuta</i>	S.T. Reynolds	<i>Edwards KE36</i>	JCT	Australia	EU720435	EU720601	EU720759	-	EU721041	EU721195	EU721383	EU721549
<i>Lepiderema</i>	<i>pulchella</i>	Radlk.	<i>Chase 2020</i>	K	Australian, BG	EU720454	-	EU720785	EU720958	-	EU721221	EU721409	-
<i>Lepidopetalum</i>	<i>fructoglabrum</i>	Welzen	<i>Edwards KE139</i>	JCT	Australia	EU720408	-	EU720724	EU720922	-	EU721161	EU721349	EU721528
<i>Lepisanthes</i>	<i>alata</i>	(Blume) Leenh.	<i>Chase 1355</i>	K	Bogor, BG	EU720450	EU720618	EU720781	-	EU721056	EU721217	EU721405	-
<i>Lepisanthes</i>	<i>feruginea</i>	(Radlk.) Leenh.	<i>Chase 1354</i>	K	Bogor, BG	EU720449	EU720617	EU720780	-	EU721055	EU721216	EU721404	-
<i>Lepisanthes</i>	<i>rubiginosa</i>	(Roxb.) Leenh.	<i>Chase 1350</i>	K	Bogor, BG	EU720446	EU720614	EU720776	EU720952	EU721052	EU721212	EU721400	EU721558
<i>Lepisanthes</i>	<i>senegalensis</i>	(Poir.) Leenh.	<i>Callmander 627</i>	MO	Madagascar	EU720492	EU720654	EU720830	EU720979	EU721091	EU721263	EU721451	EU721577
<i>Litchi</i>	<i>chinensis</i>	Sonn.	<i>Yuan s.n.</i>	NEU	China	EU720400	EU720564	EU720715	EU720914	EU721016	EU721152	EU721341	EU721522
<i>Llagunoa</i>	<i>mollis</i>	Kunth	<i>Jaramilloleija 3199</i>	MO	Colombia	EU720482	-	EU720818	-	-	EU721252	EU721440	-
<i>Llagunoa</i>	<i>nitida</i>	Ruiz & Pav.	<i>Pennington 17552</i>	MO	Peru	EU720486	-	EU720823	-	-	EU721256	EU721444	-
<i>Loxodiscus</i>	<i>coriaceus</i>	Hook. f.	<i>Bradford 1136</i>	MO	New Caledonia	EU720488	-	EU720825	-	-	EU721258	EU721446	-
<i>Macphersonia</i>	<i>chapelierii</i>	(Baill.) Capuron	<i>Buerki 138</i>	NEU	Madagascar	EU720459	EU720627	EU720791	EU720961	EU721065	EU721227	EU721415	EU721566
<i>Macphersonia</i>	<i>gracilis</i>	O. Hoffm.	<i>Rabenantoandro 1081</i>	MO	Madagascar	EU720550	EU720697	EU720898	EU721005	EU721135	EU721320	EU721508	EU721597
<i>Majidea</i>	<i>zanguearika</i>	Kirk ex Oliv.	<i>TH275</i>	MO	Madagascar	EU720552	-	EU720900	EU721006	-	EU721322	EU721510	-
<i>Matayba</i>	<i>apetala</i>	Radlk.	<i>Acevedo 11929</i>	US	Jamaica	EU720526	EU720674	EU720871	-	EU721111	EU721293	EU721481	EU721583
<i>Matayba</i>	cf. <i>opaca</i>	Radlk.	<i>Acevedo 11118</i>	US	French Guiana	EU720522	EU720669	EU720866	EU720987	EU721106	EU721288	EU721476	EU721580
<i>Matayba</i>	<i>domingensis</i>	(DC.) Radlk.	<i>Taylor 11819</i>	MO	Caribbean	EU720551	EU720698	EU720899	-	EU721136	EU721321	EU721509	EU721598
<i>Matayba</i>	<i>elaeagnoides</i>	Radlk.	<i>Zardini 43278</i>	MO	Paraguay	EU720553	EU720699	EU720901	-	EU721137	EU721323	EU721511	-
<i>Matayba</i>	<i>guianensis</i>	Aubl.	<i>Acevedo 12342</i>	US	French Guiana	EU720527	EU720675	EU720872	-	EU721112	EU721294	EU721482	-
<i>Matayba</i>	<i>laevigata</i>	Radlk.	<i>Acevedo 12357</i>	US	French Guiana	EU720528	EU720676	EU720873	EU720992	EU721113	EU721295	EU721483	-
<i>Melicoccus</i>	<i>bijugatus</i>	Jacq.	<i>Acevedo s.n.</i>	US	Puerto Rico	EU927391	EU720610	EU720771	-	EU721049	EU721207	EU721395	-
<i>Melicoccus</i>	<i>lepidopetalus</i>	Radlk.	<i>Acevedo 11128</i>	US	Bolivia	EU720443	-	EU720770	-	-	EU721206	EU721394	-
<i>Mischarytera</i>	sp.	-	<i>Edwards KE159</i>	JCT	Australia	EU720417	EU720579	EU720734	EU720930	EU721027	EU721171	EU721359	-
<i>Mischocarpus</i>	<i>exangulatus</i>	(F. Muell.) Radlk.	<i>Edwards KE30</i>	JCT	Australia	EU720434	EU720600	EU720758	EU720943	EU721040	EU721194	EU721382	-
<i>Mischocarpus</i>	<i>grandissimus</i>	Radlk.	<i>Edwards KE37</i>	JCT	Australia	EU720437	EU720604	EU720762	EU720945	EU721044	EU721198	EU721386	EU721551
<i>Mischocarpus</i>	<i>pentapetalus</i>	(Rox.) Radlk.	<i>Chase 2133</i>	K	Bogor, BG	EU720470	EU720637	EU720802	EU720966	EU721074	EU721238	EU721426	EU721571
<i>Mischocarpus</i>	<i>pyriformis</i>	(F. Muell.) Radlk.	<i>Chase 2059</i>	K	Australian, BG	EU720460	EU720628	EU720792	-	EU721066	EU721228	EU721416	-
<i>Molinaea</i>	<i>petiolaris</i>	Radlk.	<i>Rabenantoandro 1448</i>	MO	Madagascar	EU720554	EU720700	EU720902	EU721007	EU721138	EU721324	EU721512	-
<i>Molinaea</i>	sp. nov.		<i>Antilahimena 4301</i>	MO	Madagascar	EU720510	EU720662	EU720854	EU720983	EU721099	EU721280	EU721468	EU721578
<i>Neotina</i>	<i>coursii</i>	Capuron	<i>H. Razafindraibe 119</i>	MO	Madagascar	EU720543	EU720690	EU720891	EU721002	EU721128	EU721313	EU721501	EU721594
<i>Nephelium</i>	<i>lappaceum</i> (=N. <i>chryseum</i>)	L.	<i>Yuan s.n.</i>	NEU	China	EU720401	-	EU720716	EU720915	-	EU721153	EU721342	EU721523
<i>Pancovia</i>	<i>golungensis</i>	(Hiern) Exell & Mendonça	<i>Edwards KE231</i>	JCT	Tanzania	EU720411	EU720573	EU720728	EU720925	EU721022	EU721165	EU721353	EU721531
<i>Pappea</i>	<i>capensis</i>	Eckl. & Zeyh.	<i>Edwards KE232</i>	JCT	South Africa	EU720424	EU720592	EU720747	EU720938	EU721035	EU721183	EU721371	EU721542
<i>Paranephelium</i>	<i>macrophyllum</i>	King	<i>Chase 1356</i>	K	Bogor, BG	EU720451	EU720619	EU720782	EU720955	EU721057	EU721218	EU721406	-
<i>Paranephelium</i>	<i>xestophyllum</i>	Miq.	<i>Edwards KE503</i>	JCT	Asia	EU720420	EU720582	EU720737	-	EU721029	EU721174	EU721362	-
<i>Paullinia</i>	<i>pinnata</i>	L.	<i>Edwards KE199</i>	JCT	Tanzania	EU720413	EU720575	EU720730	EU720926	EU721023	EU721167	EU721355	-
<i>Paullinia</i>	<i>subauriculata</i>	Radlk.	<i>Weckerle 00/03/19-1/1</i>	Z	Peru	EU720494	-	EU720833	-	-	EU721266	EU721454	-
<i>Plagioscyphus</i>	aff. <i>louvelii</i>	Danguy & Choux	<i>Lowry 6034</i>	MO	Madagascar	EU720555	EU720701	EU720903	EU721008	EU721139	EU721325	EU721513	EU721599
<i>Plagioscyphus</i>	<i>unijugatus</i>	Capuron	<i>Buerki 145</i>	NEU	Madagascar	EU720475	EU720644	EU720809	EU721081	EU721245	EU721433	EU721574	-
<i>Podonephelium</i>	<i>homei</i>	Radlk.	<i>Pillon 156</i>	MO	New Caledonia	EU720489	EU720650	EU720826	EU720976	EU721087	EU721259	EU721447	-
<i>Pometia</i>	<i>pinnata</i>	J.R. Forst. & G. Forst.	<i>Chase 2135</i>	K	Bogor, BG	EU720471	EU720638	EU720803	EU720967	EU721075	EU721239	EU721427	EU721572
<i>Pometia</i>	<i>pinnata</i>	J.R. Forst. & G. Forst.	<i>Yuan s.n.</i>	NEU	China	EU720402	-	EU720717	EU720916	-	EU721154	EU721343	EU721524
<i>Pseudima</i>	sp.		<i>McPherson 15867</i>	MO	Panama	EU720556	EU720702	EU720904	EU721009	EU721140	EU721326	EU721514	EU721600
<i>Rhysotoechia</i>	<i>mortoniana</i>	(F. Muell.) Radlk.	<i>Edwards KE117</i>	JCT	Australia	EU720414	EU720576	EU720731	EU720927	EU721024	EU721168	EU721356	EU721533
<i>Sapindus</i>	<i>oligophyllus</i> (=Aphania <i>oligophylla</i>)	Merr. & Chun	<i>Yuan s.n.</i>	NEU	China	EU720407	EU720568	EU720722	EU720921	EU721020	EU721159	EU721159	-
<i>Sarcopteryx</i>	<i>martyana</i>	(F. Muell.) Radlk.	<i>Irvine IRV1810</i>	CSIRO	Australia	EU720426	EU720594	EU720749	EU720940	EU721037	EU721185	EU721373	EU721544
<i>Sarcopteryx</i>	<i>reticulata</i>	S.T. Reynolds	<i>Gray BG1137</i>	CSIRO	Australia	EU720421	EU720587	EU720741	-	EU721033	EU721178	EU721366	EU721539

(continued on next page)

Genera	Species	Author	Voucher	Herbarium	Country	GenBank Accession Nos.							
						ITS	matK	rpoB	trnD-trnT	trnK-matK	trnL	trnL-F	trnS-trnG
<i>Sarcopteryx</i>	sp.	–	Edwards KE49	JCT	Australia	EU720439	EU720607	EU720765	EU720948	EU721047	EU721201	EU721389	EU721554
<i>Sarcotoechia</i>	<i>serrata</i>	S.T. Reynolds	Edwards KE31	JCT	Australia	EU720436	EU720603	EU720761	EU720944	EU721043	EU721197	EU721385	EU721550
<i>Sarcotoechia</i>	<i>villosa</i>	S.T. Reynolds	Edwards KE102	JCT	Australia	EU720419	EU720581	EU720736	–	–	EU721173	EU721361	–
<i>Schleichera</i>	<i>oleosa</i>	(Lour.) Oken	Chase 2137	K	Bogor, BG	EU720423	EU720591	EU720746	EU720937	–	EU721182	EU721370	EU721541
<i>Serjania</i>	<i>altissima</i>	(Poepp.) Radlk.	Weckerle 00/07/02–1/4	Z	Peru	EU720498	–	EU720840	–	–	EU721269	EU721457	–
<i>Serjania</i>	<i>communis</i>	Cambess.	Chase 2138	K	Bogor, BG	EU720472	EU720640	EU720805	EU720969	EU721077	EU721241	EU721429	–
<i>Serjania</i>	<i>glabrata</i>	Kunth	Merello 1058	MO	Peru	EU720557	EU720703	EU720905	EU721010	EU721141	EU721327	EU721515	–
<i>Storthocalyx</i>	sp.	–	Munzinger 960	MO	New Caledonia	EU720504	EU720657	EU720848	–	EU721094	EU721276	EU721464	–
<i>Synima</i>	<i>macrophylla</i>	S.T. Reynolds	Edwards KE19	JCT	Australia	EU720430	EU720596	EU720753	EU720941	–	EU721189	EU721377	EU721546
<i>Talisia</i>	<i>angustifolia</i>	Radlk.	Zardini 43668	MO	Paraguay	EU720558	EU720705	EU720907	–	EU721143	EU721328	EU721516	–
<i>Talisia</i>	<i>nervosa</i>	Radlk.	Pennington 628	MO	–	EU720474	EU720643	EU720808	–	EU721080	EU721244	EU721432	–
<i>Talisia</i>	<i>obovata</i>	A.C. Sm.	R.Lombello 13	MO	Brazil	EU720485	EU720648	EU720822	–	EU721085	EU721255	EU721443	–
<i>Thouinia</i>	<i>acuminata</i>	S. Watson	Liston 633–2	MO	Mexico, Jalisco	EU720478	EU720647	EU720814	–	EU721084	EU721249	EU721437	–
<i>Tina</i>	<i>isaloensis</i>	Drake	Ranirison PR827	G	Madagascar	EU720520	EU720667	EU720864	EU720986	EU721104	EU721286	EU721474	EU721579
<i>Tina</i>	<i>striata</i>	Radlk.	Vary 45	MO	Madagascar	EU720509	EU720661	EU720853	–	EU721098	EU721279	EU721467	–
<i>Tinopsis</i>	<i>apiculata</i>	Radlk.	Buerki 131	NEU	Madagascar	EU720422	EU720589	EU720744	EU720936	EU721034	EU721180	EU721368	EU721540
<i>Toechima</i>	<i>erythrocarpum</i>	(F. Muell.) Radlk.	Edwards KE20	JCT	Australia	EU720431	EU720597	EU720754	–	EU721038	EU721190	EU721378	–
<i>Toechima</i>	<i>plurinerve</i>	Radlk.	Chase 1357	K	Bogor, BG	EU720452	EU720620	EU720783	EU720956	EU721058	EU721219	EU721407	EU721561
<i>Toechima</i>	<i>tenax</i>	(Cunn. ex Benth.) Radlk.	Chase 2046	K	Australian, BG	EU720456	EU720623	EU720787	–	EU721061	EU721223	EU721411	EU721564
<i>Toechima</i>	<i>tenax</i>	(Cunn. ex Benth.) Radlk.	Chase 2132	K	Bogor, BG	EU720469	EU720636	EU720801	EU720965	EU721073	EU721237	EU721425	EU721570
<i>Tristiropsis</i>	<i>acutangula</i>	Radlk.	Chase 1358	K	Bogor, BG	EU720453	EU720621	EU720784	EU720957	EU721059	EU721220	EU721408	EU721562
<i>Urvillea</i>	<i>ulmaceae</i>	Kunth	Weckerle 00/07/05–1/1	Z	Peru	EU720499	EU720655	EU720841	–	EU721092	EU721270	EU721458	–
<i>Vouarana</i>	<i>guianensis</i>	Aubl.	Lucas 109	MO	French Guiana	EU720525	EU720673	EU720870	EU720991	EU721110	EU721292	EU721480	EU721582
<i>Xanthoceras</i>	<i>sorbifolium</i>	Bunge	Yuan CN2006	NEU	China	EU720398	EU720562	EU720711	EU720910	EU721014	EU721148	EU721337	–
Outgroup													
<i>Sorindeia</i>	sp.	–	Buerki 137	NEU	Madagascar	–	–	EU720831	–	–	EU721264	EU721452	–
<i>Harrisonia</i>	<i>abyssinica</i>	Oliv.	Edwards KE510	JCT	Tanzania	EU720440	–	EU720766	–	–	EU721202	EU721390	–

References

- Acevedo-Rodríguez, P., 1993a. A revision of *Lophostigma* (Sapindaceae). *Syst. Bot.* 18, 379–388.
- Acevedo-Rodríguez, P., 1993b. Systematics of *Serjania* (Sapindaceae). I. A Revision of *Serjania* Sect. *Platyococcus*. New York Botanical Garden, New York.
- Acevedo-Rodríguez, P., 2003. Melicocceae (Sapindaceae): *Melicoccus* and *Talisia*. *Flora Neotropica Monogr.* 87, 1–179.
- Adema, F., 1991. *Cupaniopsis* Radlk. (Sapindaceae): a monograph. *Leiden Botanical Series* 15, 1–190.
- Adema, F., Leenhouts, P.W., van Welzen, P.C., 1994. Sapindaceae. *Flora Malesiana Ser. I* 11, 419–768.
- Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Akademiai Kiado (Ed.), *Second International Symposium on Information Theory*, Budapest, pp. 267–281.
- Angiosperm Phylogeny Group (APGII), 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APGII. *Bot. J. Linn. Soc.* 141, 399–436.
- Baldwin, B.G., 1992. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae. *Mol. Phylogenet. Evol.* 1, 3–16.
- Bapteste, E., Brinkmann, H., Lee, J.A., Moore, D.V., Sensen, C.W., Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Müller, M., Philippe, H., 2002. The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium*, *Entamoeba*, and *Mastigamoeba*. *Proc. Natl. Acad. Sci. USA* 99, 1414–1419.
- Biesboer, D.D., 1975. Pollen morphology of the Aceraceae. *Grana* 15, 19–27.
- Capuron, R., 1969. Révision des Sapindacées de Madagascar et des Comores. *Mém. Mus. Natl. Hist. Nat. B Bot.* 19, 1–189.
- Chase, M.W., Hills, H.H., 1991. Silica-gel – an ideal material for field preservation of leaf samples for DNA studies. *Taxon* 40, 215–220.
- Clayton, J.W., Fernando, E.S., Soltis, P.S., Soltis, D.E., 2007. Molecular phylogeny of the tree-of-heaven family (Simaroubaceae) based on chloroplast and nuclear markers. *Int. J. Plant Sci.* 168, 1325–1339.
- Cronquist, A., 1988. *The Evolution and Classification of Flowering Plants*, second ed. New York Botanic Gardens, New York.
- Dahlgren, G., 1989. An updated system of classification. *Bot. J. Linn. Soc.* 100, 197–203.
- Davies, F.G., 1997. A new genus *Haplocoelopsis* (Sapindaceae) from East and Central Africa. *Kew Bull.* 52, 719–723.
- Davies, F.G., Verdcourt, B., 1998. *Flora of Tropical East Africa: Sapindaceae*. A.A. Balkema, Rotterdam.
- Demesure, B., Sodji, N., Petit, R.J., 1995. A set of universal primers for amplification of polymorphic noncoding regions of mitochondrial and chloroplast DNA in plants. *Mol. Ecol.* 4, 129–131.
- Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.
- Driskell, A.C., Ané, C., Burleigh, J.G., McMahon, M.M., O'Meara, B.C., Sanderson, M.J., 2004. Prospects for building the Tree of Life from large sequence databases. *Science* 306, 1172–1174.
- Edwards, K.J., Gadek, P.A., 2001. Evolution and biogeography of *Alectryon* (Sapindaceae). *Mol. Phylogenet. Evol.* 20, 14–26.
- Estabrook, G.F., 1992. Evaluating undirected positional congruence of individual taxa between two estimates of the phylogenetic tree for a group of taxa. *Syst. Biol.* 41, 172–177.
- Estabrook, G.F., McMorris, F.R., Meacham, C.A., 1985. Comparison of undirected phylogenetic trees based on subtrees of four evolutionary units. *Syst. Zool.* 34, 193–200.
- Farris, J.S., Källersjö, M., Kluge, A.G., Bult, C., 1994. Testing significance of incongruence. *Cladistics* 10, 315–319.
- Felsenstein, J., 1985. Confidence-limits on phylogenies – an approach using the bootstrap. *Evolution* 39, 783–791.
- Ferrucci, M.S., 1991. *Flora del Paraguay*. Sapindaceae, vol. 16. *Hélographia SA*, Genève.
- Ferrucci, M.S., 1998. *Flora fanerogamica Argentina*. Sapindaceae 52, 1–165.
- Forest, F., Drouin, J.N., Charest, R., Brouillet, L., Bruneau, A., 2001. A morphological phylogenetic analysis of *Aesculus* L. and *Billia* Peyr. (Sapindaceae). *Can. J. Bot.* 79, 154–169.
- Gadek, P.A., Fernando, E.S., Quinn, C.J., Hoot, S.B., Terrazas, T., Sheahan, M.C., Chase, M.W., 1996. Sapindales: molecular delimitation and infraordinal groups. *Am. J. Bot.* 83, 802–811.
- George, A.S., Erdtman, G., 1969. A revision of the genus *Diplopeltis* Endl. (Sapindaceae). *Grana Palynologica* 9, 92–109.
- Hall, T.A., 1999. Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.
- Hamilton, M.B., 1999. Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Mol. Ecol.* 8, 521–523.
- Harrington, M.G., Edwards, K.J., Johnson, S.A., Chase, M.W., Gadek, P.A., 2005. Phylogenetic inference in Sapindaceae sensu lato using plastid *matK* and *rbcl* DNA sequences. *Syst. Bot.* 30, 366–382.
- Huelsenbeck, J.P., 1991. When are fossils better than extant taxa in phylogenetic analysis? *Syst. Zool.* 40, 458–469.
- Judd, W.S., Sanders, R.W., Donoghue, M.J., 1994. Angiosperm family pairs: preliminary phylogenetic analyses. *Harv. Pap. Bot.* 5, 1–51.
- Klaassen, R., 1999. Wood anatomy of the Sapindaceae. *Int. Assoc. Wood. Anat. J.* 2 (Suppl.), 1–214.
- Kluge, A.G., 1989. A Concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Syst. Zool.* 38, 7–25.
- Mabberley, D.J., 2008. *Mabberley's Plant-Book*. Cambridge University Press, Cambridge.
- Morrison, D.A., 2006. Multiple sequence alignment for phylogenetic purposes. *Aust. Syst. Bot.* 19, 479–539.
- Muellner, A.N., Samuel, R., Johnson, S.A., Cheek, M., Pennington, T.D., Chase, M.W., 2003. Molecular phylogenetics of Meliaceae (Sapindales) based on nuclear and plastid DNA sequences. *Am. J. Bot.* 90, 471–480.
- Muellner, A.N., Samuel, R., Chase, M.W., Pannell, C.M., Greger, H., 2005. *Aglaia* (Meliaceae): an evaluation of taxonomic concepts based on DNA data and secondary metabolites. *Am. J. Bot.* 92, 534–543.
- Muellner, A.N., Savolainen, V., Samuel, R., Chase, M.W., 2006. The mahogany family “out-of-Africa”: divergence time estimation, global biogeographic patterns inferred from plastid *rbcl* DNA sequences, extant, and fossil distribution of diversity. *Mol. Phylogenet. Evol.* 40, 236–250.
- Muellner, A.N., Vassiliades, D.D., Renner, S.S., 2007. Placing Biebersteiniaceae, a herbaceous clade of Sapindales, in a temporal and geographic context. *Pl. Syst. Evol.* 266, 233–252.
- Müller, J., Leenhouts, P.W., 1976. A general survey of pollen types in Sapindaceae in relation to taxonomy. In: Ferguson, I.K., Müller, J. (Eds.), *The Evolutionary Significance of the Exine*. Academic Press, London, pp. 407–445.
- Munzner, T., Guimbretiere, F., Tasiran, S., Zhang, L., Zhou, Y., 2003. TreeJuxtaposer: scalable tree comparison using focus, plus context with guaranteed visibility. *ACM Trans. Graph.* 22, 453–462.
- Nixon, K.C., 1999. The Parsimony Ratchet, a new method for rapid parsimony analysis. *Cladistics* 15, 407–414.
- Perrier, X., Flori, A., Bonnot, F., 2003. Data analysis methods. In: Hamon, P., Seguin, M., Perrier, X., Glaszmann, J.C. (Eds.), *Genetic diversity of cultivated tropical plants*. Science Publishers, Montpellier, Enfield, pp. 43–76.
- Phillipe, H., Snell, E.A., Bapteste, E., Lopez, P., Holland, P.W.H., Casane, D., 2004. Phylogenomics of eukaryotes: impact of missing data on large alignments. *Mol. Biol. Evol.* 21, 1740–1752.
- Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14, 817–818.
- Radlkofer, L., 1890. Ueber die Gliederung der Familie der Sapindaceen. *Sitz. - Ber. Akad. Wiss. München* 20, 105–379.
- Radlkofer, L., 1933. Sapindaceae. In: Engler, A. (Ed.), *Das Pflanzenreich IV*, 165 (Heft 98a-h). Leipzig, Verlag von Wilhelm Engelmann.
- Rauscher, S., 1982. Nomina nova generica et combinationes novae Spermatophytorum et Pteridophytorum. *Taxon* 31, 554–563.
- Richardson, J.E., Fay, M.F., Cronk, Q.C.B., Bowman, D., Chase, M.W., 2000a. A phylogenetic analysis of Rhamnaceae using *rbcl* and *tml-F* plastid DNA sequences. *Am. J. Bot.* 87, 1309–1324.
- Richardson, J.E., Fay, M.F., Cronk, Q.C.B., Chase, M.W., 2000b. A revision of the tribal classification of Rhamnaceae. *Kew Bull.* 55, 311–340.
- Salamin, N., Chase, M.W., Hodkinson, T.R., Savolainen, V., 2003. Assessing internal support with large phylogenetic DNA matrices. *Mol. Phylogenet. Evol.* 27, 528–539.
- Savolainen, V., Fay, M.F., Albach, D.C., Backlund, A., van der Bank, M., Cameron, K.M., Johnson, S.A., Lledó, M.D., Pintaud, J.-C., Powell, M., Sheahan, M.C., Soltis, D.E., Soltis, P.S., Weston, P., Whitten, W.M., Wurdack, K.J., Chase, M.W., 2000. Phylogeny of the eudicots: a newly complete familial analysis based on *rbcl* gene sequences. *Kew Bull.* 55, 257–309.
- Sikes, D.S., Lewis, P.O., 2001. PAUPRAT, University of Connecticut, Storrs.
- Soltis, D.E., Soltis, P.S., 1998. Choosing an approach and appropriate gene for phylogenetic analysis. In: Soltis, D.E., Soltis, P.E., Doyle, J.J. (Eds.), *Molecular Systematics of Plants II: DNA Sequencing*. Kluwer, Boston, Massachusetts, USA, pp. 1–42.
- Soltis, D.E., Soltis, P.S., Chase, M.W., Mort, M.E., Albach, T.D., Zanis, M., Savolainen, V., Hahn, W.H., Hoot, S.B., Fay, M.F., Axtell, M., Swensen, S.M., Prince, L.M., Kress, W.J., Nixon, K.C., Farris, J.S., 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbcl*, and *atpB* sequences. *Bot. J. Linn. Soc.* 133, 381–461.
- Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690.
- Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML web-servers. *Syst. Biol.* 57, 758–771.
- Swofford, D.L., 2002. PAUP*: Phylogenetic Analysis using Parsimony (* and Other Methods), Version 4. Sinauer Associates, Sunderland, Massachusetts.
- Taberlet, P., Gielly, L., Pautou, G., Bouvet, J., 1991. Universal primers for amplification of three noncoding regions of chloroplast DNA. *Plant Mol. Biol.* 17, 1105–1109.
- Takhtajan, A., 1987. *Systema Magnoliophytorum*. Soviet Sciences Press, Leningrad.
- Thomas, D.W., Harris, D.J., 1999. New Sapindaceae from Cameroon and Nigeria. *Kew Bull.* 54, 951–957.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The Clustal X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882.
- Thorne, R.F., 2000. The classification and geography of the flowering plants: dicotyledons of the class Angiospermae. *Bot. Rev.* 66, 441–647.
- Thorne, R.F., 2007. An update classification of the class Magnoliopsida (“Angiospermae”). *Bot. Rev.* 73, 67–182.
- Umadevi, I., Daniel, M., 1991. Chemosystematics of the Sapindaceae. *Feddes Repertorium* 102, 607–612.

- van der Ham, R.W., 1990. Nephelieae pollen (Sapindaceae): form, function, and evolution. *Leiden Botanical Series* 13, 1–255.
- van Welzen, P.C., 1990. *Guioa* Cav. (Sapindaceae): taxonomy, phylogeny, and historical biogeography. *Leiden Botanical Series* 12, 1–315.
- White, T.J., Bruns, T., Lee, S., Taylor, J., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis, M., Gelfand, D., Sninsky, J., White, T. (Eds.), *PCR Protocols: a Guide to Methods and Applications*. Academic Press, San Diego, pp. 315–322.
- Wiens, J.J., 1998. Does adding characters with missing data increase or decrease phylogenetic accuracy? *Syst. Biol.* 47, 625–640.
- Wiens, J.J., 2003. Missing data, incomplete taxa, and phylogenetic accuracy. *Syst. Biol.* 52, 528–538.
- Wiens, J.J., 2005. Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? *Syst. Biol.* 54, 731–742.
- Wiens, J.J., 2006. Missing data and the design of phylogenetic analyses. *J. Biomed. Inform.* 39, 34–42.
- Wiens, J.J., Reeder, T.W., 1995. Combining data sets with different numbers of taxa for phylogenetic analysis. *Syst. Biol.* 44, 548–558.
- Wilkinson, M., Cotton, J.A., Creevey, C., Eulenstein, O., Harris, S.R., Lapointe, F.J., Levasseur, C., McInerney, J.O., Pisani, D., Thorley, J.L., 2005. The shape of supertrees to come: tree shape related properties of fourteen supertree methods. *Syst. Biol.* 54, 419–431.
- Xia, N., Gadek, P.A., 2007. *Flora of China*. 12, 1–6 Available from: <www.efloras.org>.
- Xiang, Q.-Y., Crawford, D.J., Wolfe, A.D., Tang, Y.-C., DePamphilis, C.W., 1998. Origin and biogeography of *Aesculus* L. (Hippocastanaceae): a molecular phylogenetic perspective. *Evolution* 52, 988–997.
- Yang, Z.H., 1993. Maximum-likelihood-estimation of phylogeny from DNA-sequences when substitution rates differ over sites. *Mol. Biol. Evol.* 10, 1396–1401.

**Chapter two: New insights on parametric biogeography
based on the worldwide soapberry family (Sapindaceae)**

**Sven Buerki, Félix Forest, Nadir Alvarez, Johan A. A. Nylander, Nils Arrigo and
Isabel Sanmartín**

**New insights on parametric biogeography based on the worldwide soapberry family
(Sapindaceae)**

**Sven Buerki ^{a,*}, Félix Forest ^b, Nadir Alvarez ^a, Johan A. A. Nylander ^c, Nils Arrigo ^a and
Isabel Sanmartín ^d**

^a Institute of Biology, University of Neuchâtel, Rue Emile-Argand 11, CH-2009 Neuchâtel, Switzerland.

^b Molecular Systematics Section, Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, United Kingdom.

^c Department of Botany, Stockholm University, SE-10691, Stockholm, Sweden.

^d Department of Biodiversity and Conservation, Real Jardin Botánico– CSIC, Plaza de Murillo 2, 28014 Madrid, Spain.

* Corresponding author. Phone: +41327182338; Fax: +41327183001; E-mail address: sven.buerki@unine.ch.

Abstract

In the last years a wide range of methods have been developed in historical biogeography. These methods are based on fully probabilistic evolutionary models and are not constrained by the inherent biases of the parsimony (cladistic) approach. They offer the advantage to allow incorporating into biogeographical inference, estimates of the evolutionary divergence between lineages (branch lengths) and/or the timing of paleogeographic events, thus increasing the accuracy of biogeographic reconstructions. Previous implementations of parametric methods have been limited to small-scales studies in terms of time and of the complexity of the paleogeographic/biogeographic model. Here, we apply one of these methods, the dispersal-extinction-cladogenesis (DEC) likelihood model, to reconstruct the biogeographic history of the plant soapberry family (Sapindaceae), using a complex worldwide biogeographic model that spans the last 110 Ma and reflects the changing continental configuration through time. We compared results from this analysis with those resulting from a parsimony-based method, dispersal-vicariance analysis (DIVA), modified with a Bayesian empirical approach to incorporate phylogenetic uncertainty (Bayes-DIVA). Results show that, despite differences in the underlying biogeographic model (i.e., the speciation mode) and the fact that DIVA does not incorporate information on branch lengths, the two methods converge on similar biogeographic histories. The main difference lies in the timing of dispersal events - which in Bayes-DIVA sometimes conflicts with paleogeographic information on the availability of land connections - and in the tendency of this method to push dispersal events to terminal branches to explain widespread terminal ranges. In contrast, Bayes-DIVA showed the highest power (decisiveness) to unequivocally reconstruct ancestral ranges, which may be related to its ability to integrate the uncertainty in the phylogeny through the use of the Bayesian posterior distribution. Biogeographic reconstructions suggest that the Sapindaceae originated in Eurasia (north China) around the Early Cretaceous, from which they dispersed to North America and proto-SE Asia shortly thereafter. From there, they colonized Africa, Madagascar, and Australia-South America using the Gondwana connection. From the Oligocene onwards, South East Asia seems to

Chapter 2

have acted as both a centre of diversification and a source area of dispersal events over all other tropical landmasses.

Key-words: biogeographic model; dispersal-extinction-cladogenesis; dispersal-vicariance; paleogeography; Sapindaceae

Introduction

After a long period of “stagnation” following the establishment of the vicariance paradigm and the rise to prominence of the cladistic biogeographic school in the mid XX century (Croizat, 1952; Nelson & Platnick, 1981), the field of historical biogeography is going through an extraordinary revolution concerning its methods, underlying assumptions, and the kind of questions it aims to answer (Ree & Sanmartín, 2009). Cladistic biogeographic methods (also termed *pattern-based*; Ronquist, 1997, 1998a) were designed to find patterns without making any assumptions about evolutionary processes (Brooks & MacLennan, 2003; Ebach et al., 2003). Inference about processes was made *a posteriori* from the interpretation of results, by equating congruence in observed patterns to vicariance and interpreting incongruence as the result of lineage-specific events such as dispersal and extinction (Brooks, 2005; Parenti, 2007). By uncoupling the inference of area patterns from the underlying evolutionary processes, cladistic methods made it difficult to compare alternative biogeographic scenarios (Sanmartín, 2007).

“Event-based methods” (Ronquist, 2003) represented an important leap forward over cladistic methods in that they were derived from explicit models of biogeographic processes. These methods use a “deterministic” cost model approach in which each relevant process receives a fixed cost according to a parsimony optimality criterion and the analysis consists of finding the minimum-cost (most parsimonious) reconstruction for the observed distribution ranges (Sanmartín, 2007). Event-based methods provided many advantages over cladistic methods, such as the direct identification of events of interest to the biologist (e.g., dispersal, vicariance, and extinction events), and some of them, such as dispersal-vicariance analysis (Ronquist, 1996, 1997) or parsimony-based tree fitting (Page, 2003; Ronquist, 2003; Sanmartín & Ronquist, 2004), have become increasingly popular both in historical biogeography and coevolution.

Nevertheless, like cladistic methods, event-based methods were limited by their reliance on the “principle of parsimony” for biogeographical inference (Sanmartín, 2009, in press). This has several implications. For example, the cost of the events (i.e., the likelihood

of an event to occur) cannot be estimated from the data, but must be defined in advance using *ad-hoc* optimality criteria, such as maximizing the conservation of biogeographical patterns along a phylogeny (see Sanmartín et al., 2007). Because the “most parsimonious” explanation is the one that implies the minimum number of changes in the geographic range of a lineage, parsimony-based reconstructions tend to underestimate the frequency of lineage-specific events such as dispersal and extinction. For example, in event-based reconstructions dispersal and extinction are penalized by giving them a higher cost than other processes, since these events generate geographic ranges in the descendants that are different from those of their ancestor (i.e., “not phylogenetically conserved”, Sanmartín et al., 2007). Furthermore, parsimony-based reconstructions generally ignore two sources of error (stochastic variance) associated to the biogeographic inference process (Ronquist, 2004; Sanmartín et al., 2008; Ree & Sanmartín, 2009; Sanmartín, in press, 2009): uncertainty associated with reconstructing ancestral states onto a phylogeny (“mapping uncertainty”) – since only minimum-change reconstructions are evaluated – and “phylogenetic uncertainty”, since ancestral areas are reconstructed over a single tree topology, the most parsimonious tree, assuming the phylogeny is known without error. The use of parsimony also limits the type of data that can be used in biogeographical analysis to the topology of the phylogeny and the distribution of the terminal species. Other relevant evidence such as the times of divergence between lineages cannot be easily incorporated within the parsimony framework (Donoghue & Moore, 2003).

In the last few years, new methods based on probabilistic process models have been developed that are not constrained by the inherent biases of the parsimony approach (Ree & Sanmartín, 2009; Sanmartín, in press, 2009). By analogy with the evolution of a character along a phylogenetic tree, these methods – termed “model-based” or “parametric” because they are based on models with parameters – model range evolution, i.e., the change in geographic range from ancestor to descendants, as a stochastic process with discrete states (geographic ranges) that evolves along the branches of the phylogeny according to a probabilistic Markov model. Transitions between states are assumed to occur stochastically

according to a matrix of transition probabilities, whose parameters are biogeographical processes (dispersal, extinction, range expansion) determining the probability of range evolution – by jump, contraction, or expansion – from ancestor to descendant as a function of time.

Relative to event-based and cladistic approaches, model-based methods offer the advantage that they allow incorporating into biogeographical reconstructions estimates of the evolutionary divergence between lineages or their time since cladogenesis – represented by the length of branches in the phylogeny, that is, the probability of biogeographical change is higher along long branches than along shorter ones (Ree & Sanmartín, 2009). Thus, they overcome the known bias of parsimony methods for underestimating the number of changes along branches. Moreover, they can account for the uncertainty in ancestral range reconstruction since all possible biogeographical scenarios are evaluated in estimating the relative probabilities of ancestral areas. In addition, the Bayesian framework allows integration over topological and branch length uncertainty using samples from the posterior distribution of phylogenies (Sanmartín, in press, 2009). Finally, as with event-based methods, assumptions about processes are made explicit and integral to the inference framework but, because the underlying stochastic models are based on well-known probability distributions, parametric approaches provide a more rigorous statistical framework for the testing of alternative biogeographic hypotheses than event-based methods (Sanmartín, in press, 2009).

Recently, two new methods have been proposed that are based on explicit parametric models of biogeographical processes: the Bayesian approach to island biogeography developed by Sanmartín et al. (2008) and the dispersal-extinction-cladogenesis (DEC) likelihood model described by Ree et al. (2005) and further developed by Ree & Smith (2008). The latter is a parametric, extended version of dispersal-vicariance analysis that estimates by maximum likelihood (ML) ancestral ranges, transition rates between ranges, and biogeographic scenarios of range inheritance for a single individual group (Ree & Smith, 2008).

Despite their potential advantages, model-based approaches have also their own limitations (Ree & Sanmartín, 2009; Sanmartín, in press, 2009). The biggest of them is computational feasibility and how to balance this with inferential power and increasing realism of biogeographic scenarios. For example, for methods allowing widespread states such as the DEC model, the size of the matrix increases exponentially with the number of areas. Therefore, it becomes increasingly important to define the areas of analysis carefully, with regard to the kind of question one wants to answer and considering alternative sources of information such as paleogeographic scenarios or the ecological tolerances of the species of interest (Ree & Sanmartín, 2009; see below).

In comparison with parametric methods, an event-based method such as dispersal-
vicariance analysis (DIVA) offers the advantage that it does not require any prior knowledge of the geological history of the areas studied (i.e., the timing of geographical barriers and connection routes) or the divergence times between lineages (Nylander et al., 2008a), and can therefore be applied to phylogenies where branch lengths are meaningless or difficult to interpret, such as in morphologically-based cladograms. Thus, this method has remained widely popular, as evidenced by the large number of papers published that have used it. Moreover, comparisons between DIVA and its ML counterpart, Lagrange, show that for many biogeographic scenarios the two methods provide very similar solutions (Ree et al., 2005; Inda et al., 2008; Xiang & Thomas, 2008; Moore et al., in press). Probably the most important limitation of DIVA, more than the optimization algorithm itself, is the fact that ancestral areas and biogeographic events must be reconstructed onto a fixed, fully resolved tree topology (i.e., a binary tree), assuming that phylogenetic relationships are known without error. This is problematic since unresolved relationships (i.e., polytomies) and weak nodal support are common features in many phylogenies due to low phylogenetic signal or incongruity between character sets. Moreover, some nodes are likely to be more strongly supported than others and this degree of uncertainty should be reflected in the biogeographic inference. To address this problem Nylander et al. (2008a) proposed an empirical Bayesian approach to dispersal-
vicariance analysis (Bayes-DIVA) that accounts for phylogenetic uncertainty in biogeographic

inference by integrating DIVA ancestral area reconstructions over a Bayesian Markov chain Monte Carlo (MCMC) sample of trees representing the posterior distribution of the tree topology. Integrating over the posterior distribution of trees often may help to reduce the uncertainty in the biogeographic reconstruction (Nylander et al., 2008a). Bayes-DIVA has to date been used in several empirical studies, including plants (Antonelli et al., 2009; Roquet et al., in press) and animals (Nylander et al., 2008a).

In a recent comparative study in ancestral range reconstruction methods, Clark et al. (2008) compared parametric versus parsimony approaches to biogeographical inference in the context of island biogeographic scenarios and found that methods that incorporate branch lengths and/or timing of events gave more plausible area range histories. They encouraged the use of alternative biogeographic systems – such as continental lineages – to illustrate further the comparative performance of the methods. Here, we compare a parsimony-based method, dispersal-vicariance analysis developed by Ronquist (1997) and later modified by Nylander et al. (2008a) to incorporate phylogenetic uncertainty with a Bayesian empirical approach (Bayes-DIVA), against a parametric method, the dispersal-extinction-cladogenesis likelihood model developed by Ree & Smith (2008) and implemented in their program Lagrange. These two biogeographic inference methods are appropriate for the analysis of continental biogeographic scenarios because they model dispersal – geographic movement – as the result of allopatric speciation following range expansion, which seems suitable for scenarios in which areas are spatially contiguous. Here, we analyse the benefits and limitations arising from each method's underlying assumptions using the plant family Sapindaceae (soapberry family) as a case study. The distribution of this family spans all continents and its fossil record indicates that its biogeographic history traces back to the Early Cretaceous (see below). Previous implementations of parametric methods such as Lagrange have been limited to small-scale studies (Clark et al., 2008; Santos et al., 2009), both in terms of the temporal scale (the last 20 Ma) and the complexity of the paleogeographic scenario implemented in the model (e.g., Santos et al., 2009). Here, we propose a complex worldwide biogeographic model that spans the last 110 Ma and follows a

stratified paleogeographic scenario reflecting the changing continental configuration and the availability of area connections through time. This is the first time such a complex model has been used in parametric biogeography and we hope it will help to demonstrate the potential of these methods in reconstructing the biogeographic history of lineages and biotas.

The study group

Several lines of evidence support the choice of the Sapindaceae family as an ideal case study to investigate the performance of biogeographic methods. Recent phylogenetic studies have supported the monophyly of this mid-sized (ca. 140 genera, ca. 1900 species) cosmopolitan family (Harrington et al., 2005; Buerki et al., 2009). Based on molecular and morphological characters, Buerki et al. (2009) subdivided Sapindaceae into four subfamilies: Xanthoceroideae, Hippocastanoideae, Dodonaeoideae and Sapindoideae. The first two subfamilies occur in temperate regions, whereas the two remaining are widely distributed in the tropical regions (with the exception of a few genera – e.g. *Dodonaea* – mainly occurring in temperate regions of the Southern Hemisphere; Buerki et al., 2009). At the distribution level, Sapindaceae taxa range from being restricted to narrow areas (e.g., the newly described genus *Gereaua* is endemic to Madagascar; Buerki et al., submitted), occurring across two continents (e.g., *Cupaniopsis* is shared between Australia and southern Asia), to being widespread all over the tropics (e.g., *Allophylus*). Moreover, several fossils associated with monophyletic genera are available, providing accurate points of calibrations for divergence time estimations. Thus, the soapberry family provides a highly suitable subject with which to study the influence of paleogeography and ecological tolerances on range evolution.

Material and Methods

Data set

The data set used to estimate lineage divergence times and ancestral ranges in Sapindaceae is based on Buerki et al. (2009), with the addition of several taxa required to calibrate the divergence analyses (e.g., *Allophylus*, *Paullinia*). Ingroup sampling comprised 148 specimens representing more than 60% of the generic diversity and two outgroups including one species of Anacardiaceae (*Sorindeia* sp.; defined as outgroup in all analyses based on Savolainen et al., 2000 and Muellner et al., 2007) and one species of Simaroubaceae (*Harrisonia abyssinica*). Species names, voucher information, and GenBank accession numbers for all sequences are provided in Buerki et al. (2009) and in Buerki et al. (submitted). The DNA extraction, amplification and sequencing protocols of the nuclear ribosomal region and seven plastid regions are provided in Buerki et al. (2009). The nuclear region is the whole ITS region (*ITS1*, *5.8S* and *ITS2*) and plastid markers include coding (*matK* and *rpoB*) and non-coding regions (the *trnL* intron and the intergenic spacers *trnD-trnT*, *trnK-matK*, *trnL-trnF* and *trnS-trnG*).

Phylogenetic analyses

Bayesian Markov chain Monte Carlo (MCMC) was used to approximate the posterior probability distribution of the phylogeny based on the combined plastid-nuclear dataset, using the program MrBayes v.3.1.2 (Ronquist & Huelsenbeck, 2003). To decrease the complexity and increase mixing in the MCMC, the dataset was divided into two partitions, nuclear (including only the ITS region) and plastid (including the seven plastid markers), and each locus was allowed to have partition-specific model parameters (Ronquist & Huelsenbeck, 2003; Nylander et al., 2004). Model selection for the data partitions in the MCMC was carried out with MrModeltest2 v.2.3 (Nylander, 2004) based on the Akaike information criterion (Akaike, 1974). For both partitions, the best-fitting model was the general time reversible (GTR) model with an alpha parameter for the shape of the gamma

distribution to account for among-site rate heterogeneity (Yang, 1993). Three Metropolis-coupled Markov chains with incremental heating temperature of 0.2 were run for 50 million generations and sampled every 1000th generations. The simulation was repeated twice, starting from random trees. Convergence of the MCMC was checked using the effective sample size criterion for each parameter as implemented in TRACER v.1.4 (Rambaut & Drummond, 2007) and by monitoring cumulative posterior split probabilities and among-run variability of split frequencies using the online tool AWTY (Nylander et al., 2008b). Because of the high number of trees and to avoid misleading relationships caused by underestimated burn-in, 80% and respectively 85% of the trees from runs one and two were discarded and the remaining samples from the independent runs were pooled to obtain the final approximation of the posterior probability distribution of the phylogeny (in all 15362 trees).

To yield a single hypothesis of the phylogeny, the posterior distribution was summarized in the “allcompat consensus tree” from MrBayes. The choice of the allcompat consensus over the most commonly used 50% majority rule consensus tree (“halfcompat consensus”) is based on the fact that all phylogenetic dating methods (see below) and some of the biogeographic methods (e.g., Lagrange) used here require that all nodes in the phylogeny are fully resolved prior to the analysis (i.e., a fully bifurcate tree). Instead of arbitrary solving the polytomies, we chose to base our biogeographic reconstruction on the most likely hypothesis based on our data set. Also, the allcompat consensus differed from the halfcompat tree by only two polytomic nodes. The alternative of using a maximum likelihood tree for dating and biogeographic reconstruction (e.g., Clark et al., 2008; Santos et al., 2009) is not appropriate here, since the Bayes-DIVA method estimates ancestral range probabilities by integrating over the Bayesian stationary distribution. Thus, using a Bayesian allcompat consensus as the reference tree seems more appropriate (see below).

Estimation of lineage divergence times

A likelihood ratio (LR) test (Felsenstein, 1988) was used to determine whether sequence data conformed to the expectation of a molecular clock. First, a maximum

likelihood (ML) total evidence tree (sensu Kluge, 1989) was produced using RAxML v7.0.0 (Stamatakis, 2006; Stamatakis et al., 2008), following Buerki et al. (2009), and using the facilities offered by the CIPRES portal in San-Diego, U.S.A.

(<http://8ball.sdsc.edu:8888/cipres-web/home>). The ML model with and without the enforcement of a molecular clock were calculated using PAUP* version 4.0b10 (Swofford, 2002) on the ML topology. Modeltest 3.7 (Posada & Crandall, 1998) selected the general time reversible (GTR) model with an alpha shape parameter for the gamma distribution with four discrete categories to account for among-site rate heterogeneity (Yang, 1993). Base frequencies, the transition/transversion ratio and the gamma distribution shape were estimated while running the ML analyses. The comparison of the difference between the likelihood scores of the tree with and without an enforced molecular clock, multiplied by two, was compared with a chi square distribution. When the tree is fully resolved, the number of degrees of freedom (df) is $N-2$, where N is the number of taxa, which is equal to the number of internal branches of the tree (Sanderson & Doyle, 2001). The likelihood ratio test strongly rejected the molecular clock: $D_2=2[69029.95-68302.98] = 726.97$, 148 df, $P < 0.001$).

Given the lack of a molecular clock, relative branching times were estimated based on the allcompat consensus using two different relaxed molecular clock approaches: nonparametric rate smoothing (Sanderson, 1997; hereafter NPRS) and penalized-likelihood (Sanderson, 2002; hereafter PL), both implemented in the program r8s v.1.71 (Sanderson, 2004). NPRS is an autocorrelated relaxed-clock method that assumes that the evolutionary rate can evolve over time but it is to some extent inherited from ancestor to descendants. This method “smooths” the rate over the phylogeny by minimizing change in rates between mother-daughter lineages across the whole tree. PL is a semi-parametric method that combines a parametric model (i.e., each branch in the phylogeny may be assigned its own substitution rate) with a non-parametric roughness penalty that attributes higher probability costs to optimizations where rate changes too quickly from branch to branch (Sanderson, 2002).

Commands for the NPRS analysis were as follows: Divtime method=NPRS

algorithm=POWELL. Commands used for the PL analysis were as follows: Set

smoothing=320; Divtime method=PL algorithm=Truncated Newton method. The smoothing value was established using the cross-validation routines implemented in the program r8s. A smoothing value of 320 was applied to the n randomly selected trees (see below). The most external outgroup, *Sorindeia* sp. (Anacardiaceae) was pruned for the estimation of the divergence time as required by the program (see Sanderson, 2004). To account for phylogenetic uncertainty in the estimation of lineage divergence times, NPRS and PL analyses were also performed on a random sample of trees ($n=1000$) from the Bayesian MCMC stationary distribution, after discarding those trees that did not respect the dating priors [$< 2\%$ (236) of the trees]. The software TreeAnnotator (Drummond & Rambaut, 2007) was used to summarize the results and obtain mean values and 95% confidence intervals for node heights for the NRPS and PL methods. Finally, Minimum, Median and Maximum ages for NPRS and PL were extracted using scripts written for the R package APE (Paradis et al., 2004) (available from S.B. on request). These values were used to plot the accumulation of lineages through time (LTT plots, using APE) and to estimate the impact of divergence time uncertainty on biogeographic reconstructions (see below).

Fossil calibration: To estimate absolute ages from the resulting chronograms, we used six fossil calibration points within the Sapindaceae:

- A. The root node (i.e. the most recent common ancestor of the Sapindaceae and Simaroubaceae) was constrained to a maximum age of 125 million years (my) based on the oldest known pollen fossil of eudicotyledones (Doyle & Hotton, 1991; Magallón & Sanderson, 2001);
- B. The stem group of *Acer*, *Aesculus*, and *Dipteronia* was constrained to a minimal age of 55.8 my, based on leaves fossils of *Acer* from the Upper Palaeocene to the Early Miocene (Brown, 1935, 1937; Piel, 1971; Taylor & Taylor, 1993; Manchester, 1999; Pfosser et al., 2002); fruits fossils of *Dipteronia* from the Palaeocene to the

Oligocene (Manchester, 1999; McClain & Manchester, 2001) and fossils of leaves and fruits of *Aesculus* from the Palaeocene (Manchester, 2001);

- C. The stem group of *Dodonaea* and *Diplopeltis* was constrained to a minimal age of 37.2 my based on pollen, leaves and fruits fossils of *Dodonaea* from the Middle Eocene to Pliocene (Berry, 1916, 1930; Song, 1988; Zhi-Chen et al., 2004) and pollen fossil of *Diplopeltis* from the the Middle Eocene (Kemp, 1976). The genus *Diplopeltis* has type-F pollen which is unique within Sapindaceae (see Müller & Leenhouts, 1976);
- D. The stem group of *Koelreuteria* was constrained to a minimal age of 37.2 my based on leaves and fruits fossils of *Koelreuteria* from the Middle Eocene to Oligocene (Brown, 1934; Manchester, 1999; Teodoris, 2003);
- E. The stem group of *Pometia* was constrained to a minimal age of 5.33 my based on pollen fossil of *Pometia* from the Miocene to Pliocene (Müller, 1981). The genus *Pometia* has type-C1 pollen which is unique within Sapindaceae (see Müller & Leenhouts, 1976);
- F. The stem group of *Cardiospermum*, *Paullinia* and *Serjania* was constrained to a minimal age of 37.2 my based on pollen fossils of *Cardiospermum* (Leopold & McGinite, 1972), *Paullinia* (Müller, 1981) and *Serjania* (Graham, 1976) from the Lower to Upper Miocene. The pollen types C2 and C3 within the Sapindaceae are restricted to the genera included in the stem of this group (see Müller & Leenhouts, 1976).

For each calibration point the oldest fossil record was selected and the geological time scale of Gradstein et al. (2004) was used to set the calibration points.

PL divergence time with uncertainty as a substitute of Bayesian methods

In this study, NPRS and PL divergence time algorithms were favoured over fully parametric dating methods such as Bayesian relaxed clock methods, e.g., MultiDivtime (Thorne & Kishino, 2002) and BEAST (Drummond & Rambaut, 2006) because of their

greater ability to handle large and complex data sets. Our dataset was based on eight nuclear and plastid markers and more than 150 specimens. For currently unknown reasons MultiDivtime was unable to provide any results (i.e., the analysis never finished), whereas BEAST analyses showed mixing and convergence problems. Our empirical results suggest that the MCMC research algorithm implemented in BEAST was less efficient than the one implemented in MrBayes, especially with large data sets such as ours. We have now undertaken additional investigations to confirm these empirical results and identify solutions to this matter (Buerki et al., in prep.).

Estimates of lineage divergence times were used to assess the impact of temporal information on biogeographic reconstructions, both indirectly by using time-calibrated branch lengths, and more directly by constraining our biogeographic model according to a temporally stratified paleogeographic model (see below). One limitation of non-Bayesian relaxed clock methods such as PL and NPRS is that divergence times are estimated onto a fixed tree topology with set branch lengths. To approximate at least partially the ability of Bayesian methods to account for phylogenetic uncertainty, we followed a Bayesian empirical approach similar to that of Nylander et al. (2008a). We inferred PL and NPRS divergence times over a representative sample of trees from the MCMC Bayesian stationary distribution and summarized results as mean age values and 95% confidence intervals for each nodal age. These Median, Maximum, and Minimum PL chronograms were later used to infer biogeographic scenarios in Lagrange (see below). We believe that this indirect approach may be useful when large data sets preclude the use of Bayesian relaxed clock methods to estimate phylogenetic and dating uncertainty simultaneously.

Biogeographic analyses

Areas of distribution. – To minimize the effect of taxon sampling bias in our analysis, terminal distributions were coded to follow generic distributions (following Adema et al., 1994; Mabberley, 2008; Acevedo-Rodriguez et al., unpublished data), except when the genus was recognized as para-/polyphyletic (see Buerki et al., 2009 and submitted for more details), in

which case terminals were coded according to the distribution of the species (e.g., *Cupaniopsis*).

Seven geographic areas were defined based on paleogeological history and current distribution patterns in the Sapindaceae (Fig. 1): (A) Eurasia: from Western Europe to Indochina; (B) Africa; (C) Madagascar, including the Comoro Islands and the Mascarene Islands; (D) South East (SE) Asia, including India, the Malaysian Peninsula, Philippines, Sumatra, Borneo, and the Inner Banda Arc, as well as the Pacific Islands (e.g., Hawaii); (E) Australia: including New Guinea, New Caledonia and New Zealand; (F) North America, and (G) South America including Central America and the West Indies (Fig. 1).

Area D ("SE Asia") has a complex paleogeographic (and biogeographic) history, involving numerous small terranes that rifted away from Gondwana during the Paleozoic-Mesozoic and were progressively accreted to the southern part of the Eurasian Plate (North China) at different times during the Mesozoic and Cenozoic (Metcalf, 1998; Sanmartín & Ronquist, 2004). South China and Indochina were the first to be accreted in the Late Devonian-Early Carboniferous, followed by the Quiantang and Sibumasu blocks (the eastern half of the Malaysian Peninsula) in the Permian-Triassic, whereas SW Borneo and the Semitau terranes were derived from the margin of Cathaysia land (South China and Indochina) by the opening of a marginal basin in the Cretaceous-Tertiary (Metcalf, 1998). The rest of insular SE Asian terranes (Sumatra, the rest of Borneo, Celebes, the Inner Banda Arc, etc) were formed as a result of the collision of the Australian Plate with the Eurasian Plate during the Cenozoic. This composite paleogeographic history has two consequences in our biogeographic model. Firstly, area A (Eurasia) is here defined as including South China and the entire Indochina peninsula (Fig. 1), since these terranes have already accreted to the Eurasian Plate by the time of the estimated origin of the Sapindaceae (Early Cretaceous, see also Sanmartín et al., 2001). Secondly: area D (SE Asia) did not exist as we know it today until the Cenozoic, so we need to distinguish between a "proto-SE Asia", the older parts of the Malaysian Peninsula and SW Borneo, which were already in place by the Cretaceous-Tertiary boundary, and insular SE Asia, which did not appear until the Middle-Late Cenozoic.

Moreover, area D – as defined here – did not exist at the time of origin of the Sapindaceae family in the Early Cretaceous. This changing configuration of area D plays a very important role in the definition of our biogeographic model (see below).

Inference methods. – Ancestral area distributions and biogeographic events explaining current distribution patterns in Sapindaceae were inferred by using two different methods: the parsimony-based method of dispersal-vicariance analysis implemented in DIVA v. 1.2 (Ronquist, 2001) and the dispersal-extinction-cladogenesis (DEC) likelihood method implemented in Lagrange (Ree et al., 2005; Ree & Smith, 2008).

Bayes-DIVA analysis: Dispersal-vicariance analysis is a method for inferring the most parsimonious reconstruction of ancestral ranges on a given phylogeny by minimizing the number of dispersal and extinction events that are needed to explain the current terminal distributions (Ronquist, 1997). Unlike other parsimony-based methods, it offers the advantage that area relationships are not constrained to follow a hierarchical, branching scenario so the method can be used to reconstruct biogeographic scenarios with changing continental configurations (Ronquist, 1997; Sanmartín, 2007). DIVA uses a three-dimensional cost matrix to estimate the cost of moving from the ancestor to each of the descendants (see Table 1). It allows two different scenarios for range inheritance at speciation nodes: a) duplication or within-area speciation when the ancestor is distributed in a single area and the two descendants each inherit the entire ancestral range (A to A, **S** in Table 1); b) vicariance when the ancestor occurs in two or more areas and each descendant inherits a non-overlapping subset of the ancestral range (AB to A and B; **Valb** in Table 1). Inheritance of widespread ancestral ranges is not allowed (AB to AB; this event is not allowed in the matrix, “–“ in Table 1). Only one dispersal event per branch (between two ancestral nodes) is allowed in the model, except for terminal branches leading to widespread taxa, for which DIVA postulates multiple dispersal events (see below).

DIVA analyses were run with no constraint in the maximum number of areas allowed at ancestral nodes and with the maximum number of areas constrained to two. Except for higher uncertainty in the assignment of ancestral ranges (i.e., lower marginal probabilities),

results from the unconstrained Bayes-DIVA analysis were highly congruent with the more restricted analysis, with the largest differences found at the deeper nodes. Since it is highly improbable that the ancestor of the Sapindaceae was widespread over all continents – most species of Sapindaceae are restricted to one or two areas, with a few species extending to three areas and rarely to four or five areas (Fig. 5) – only results from the maximum range size of two areas analysis are shown here.

Uncertainty in phylogenetic relationships was accounted for in DIVA by using a novel approach proposed by Nylander et al. (2008a), which integrates DIVA parsimony-based reconstructions over a Bayesian MCMC sample of trees representing the posterior probability of the tree topology. Specifically, we sampled all the trees (15362 in total) from the MCMC post-burnin sample and used scripts to summarize/average ancestral area reconstructions over all sampled trees for each node in a reference tree, in this case the allcompat tree, which was used as the reference. Only those trees containing the node of interest were summarized in estimating the probabilities for that node. This approach allows an estimation of marginal probabilities of ancestral ranges for a given node while integrating over the uncertainty in the rest of the tree topology (Nylander et al., 2008a).

Lagrange analyses. – In addition to the parsimony-based Bayes-DIVA method presented above, a second method based on the likelihood optimality criterion was used to reconstruct ancestral range inheritance scenarios at internal nodes on the phylogeny of the Sapindaceae. The DEC model implemented in the program Lagrange (Ree et al., 2005; Ree & Smith, 2008) is a parametric, extended version of dispersal-vicariance analysis in which dispersal and extinction are modelled as anagenetic processes evolving along the branches of the tree according to a stochastic Markov process, whereas vicariance and within-area speciation are cladogenetic events that determine the inheritance of geographic ranges at speciation nodes (Ree et al., 2005). Range evolution along a phylogenetic branch is governed by a Q matrix of transition rates between ranges, whose parameters are biogeographic processes such as dispersal and local extinction describing the probability of

geographic change from ancestor to descendant by range expansion and range contraction, respectively (Ree & Smith, 2008). Given a tree with estimates of branch lengths, the distributions of terminal species, the Q matrix defining the model of range evolution, and prior probabilities for range inheritance scenarios, this method allows for estimation of the probability of ancestral ranges and the rates of transition between geographical ranges using standard ML inference algorithms (Sanmartín, in press, 2009).

One of the advantages of the DEC model is its flexibility: parameters in the transition matrix can be scaled or modified to reflect the changing paleogeography, the availability of area connections (e.g., land bridges) through time, or the dispersal capabilities of the study group of interest. One way to do this is by constraining the geographic ranges that are valid states in the model, thus reducing the size of the Q transition matrix. For example, one could exclude those ranges that span more unit areas than the current geographic range of extant species (Ree & Sanmartín, 2009). In order to compare with the Bayes-DIVA results, Lagrange analyses were first run unconstrained (M0, all combinations of the seven areas allowed) and then constrained to include only ancestral ranges that span a maximum of two areas (M1). As with Bayes-DIVA, the Lagrange unconstrained (M0) analysis provided very similar results to the constrained (two-area) analysis (M1), but with higher levels of biogeographic uncertainty (i.e., lower probability values, data not shown). Since the M0 analysis required at least five times more computation power than the M1 model, in order to minimize model complexity and computational time, and to provide comparable results with the Bayes-DIVA analysis, only results from the constrained Lagrange M1 analyses are discussed here (see below).

Table 2 shows the Q matrix for the M1 constrained model. The Q matrix defines the rate of instantaneous change (i.e., in an infinitesimal amount of time, dt). The DEC model assumes that only one event, a single dispersal or local extinction event, can occur in an instant of time. Therefore, those transitions that imply more than one event are given a rate of 0 in the Q matrix. For example moving from D to G would imply a range expansion from D to G followed by extinction in D (**Dag + Eg** in Table 2). Like DIVA, Lagrange assumes that

movement between single areas goes through a widespread ancestor. However, unlike DIVA (Ronquist, 1997), which enforces vicariance of widespread ancestors, the DEC model is more flexible by allowing a “peripatric” mode of speciation. For widespread ancestors, lineage divergence could arise either between a single area and the rest of the range or within a single area (Ree et al., 2005). Therefore, inheritance of widespread ancestral ranges at speciation nodes is allowed in Lagrange (AB to AB, see Table 2), although for one of the descendants only. Another difference is that, unlike DIVA where the maximum number of areas forming ranges at ancestral nodes can be constrained independently from the distribution of the terminals, the Q transition matrix in Lagrange required the inclusion of a few ancestral ranges covering three, four, and five unit areas, which were present in the most widespread terminals (e.g., *Dodonaea viscosa* in abcdeg, see Fig. 5). This is because the ML optimization of geographic ranges at ancestral nodes is conditional on the range distribution of the descendants (Ree et al., 2005), so these widespread ranges are “dragged” back over the tree as conditional likelihoods are being calculated in the optimization upward pass from the tips to the root. As seen below, this considerably increased the uncertainty in nodal ancestral range reconstruction, especially for the deeper nodes.

Finally, in order to account for the changing paleogeography over which Sapindaceae evolved, and to take advantage of the flexibility of the DEC model, we defined a third Lagrange model (Str) in which we stratified the phylogeny into different time slices reflecting the changing continental configuration, as tectonic plates moved apart, broke up, or collided with each other through time. Based on paleogeographic reconstructions and estimates of divergence times between the main Sapindaceae lineages, we constructed a stratified model with four time slices: before 80 million years ago (Ma), between 80-65 Ma, between 65-35 Ma, and between 35 Ma and present day (Fig. 2). For each time slice, we defined a Q matrix in which transition rates were made dependent on the geographic connectivity between areas (i.e., through land bridges, changing sea levels, wind currents, etc.). This was done by scaling the rate of dispersal in the Q matrix according to the availability of area connections. For example, the dispersal rate between two areas that were not tectonically connected at a

given time slice was scaled downwards by a factor of 0.01 to reflect the low probability of movement between those two areas; dispersal through abiotic factors such as wind or ocean currents was scaled down to 0.5 of the global dispersal rate, since it usually involves a more restrictive, stochastic “sweepstakes” dispersal (see Fig. 2 and the section below for a more complete description of our paleogeographic-biogeographic model). In addition, to decrease the uncertainty created by widespread terminal ranges being dragged up towards the root, and to make it comparable with the constrained Bayes-DIVA analysis, we constrained the root node of the Lagrange stratified analysis as occurring in area “A” only (which is also the ancestral area reconstructed by Bayes-DIVA).

Stratified biogeographical model (Fig. 2) – During the Early to Late Cretaceous (the period covered by the first time slice) Pangaea was subdivided into two distinct continents: Laurasia and Gondwana (Smith et al., 1988; Sanmartín, 2005) that were isolated by the Tethyan Seaway. The Laurasian landmass comprised modern Eurasia and North America (in our model areas A and F, respectively), which were at that time connected by land corridors across a narrow Atlantic Ocean (Smith et al., 1988), and later through the so-called trans-Atlantic Thulean land bridge (Tiffney, 1985a; Sanmartín et al., 2001). After the opening of the Atlantic Ocean and the breaking of the Thulean Bridge around the Middle Eocene, the main dispersal route between North America and Eurasia switched to Beringia, which was then situated further south than today (Tiffney, 1985a,b). Dispersal across Beringia was possible until the end of the Neogene, when the Bering Strait opened up (3.5 Ma, Sanmartín et al., 2001). This continuous connection between North America and Europe is reflected in our model by allowing free dispersal between areas A and F across the four time slices (e.g., rates scaled by “1”, Fig. 2). As explained above, area D (SE Asia) did not exist before the Early Paleocene (c. 60 Mya) so this area was not included in the first time slice (see Metcalfe, 1998 for more details). The break up of Gondwana started in the Late Jurassic-Early Cretaceous (165-130 Mya) but dispersal between the Gondwanan landmasses (areas B, C, G and E in our model) across Antarctica was still possible until the Late Cretaceous

(65-60 Ma), when New Zealand and New Caledonia drifted away from West Antarctica, breaking land connections with Australia and South America (see Sanmartín, 2005; Sanmartín & Ronquist; 2004 and references therein). Although Africa (B) broke away from Gondwana in the Early Cretaceous, dispersal between Africa (B), Madagascar (C) and Australia (E) was still possible through the Kerguelen Plateau, a continental block situated off the eastern coast of India that seems to have acted as a land bridge between Africa-India-Madagascar and Australia-Antarctica during the Late Cretaceous (see Sanmartín & Ronquist, 2004 and references therein) and which foundered at the end of the Cretaceous. Thus, dispersal among these Gondwanan landmasses was also allowed during the second time slice (80-65 Ma). The end of land dispersal between South America (area G) and Africa (B) has been dated circa 95 Ma (Morley, 2003), so direct movement between these two areas was constrained to occur only in the first time slice (> 80 Ma). Africa and Madagascar (C) broke away in the Early Cretaceous (121 Ma) but they have always been geographically close (ca. 300 km across the Mozambique channel), and this proximity is reflected in our model by allowing dispersal between these two areas across all four times slices (Fig. 2). In the Late Cretaceous to Early Paleocene (second time slice, Fig. 2), North America (F) and South America (G) became connected through the proto-Greater Antilles land bridge (Briggs, 2004). This connection was disrupted in the Paleocene/Eocene (third time slice; Fig. 2) with the foundering of the Caribbean Plate and reappeared during the Neogene (fourth time slice; Fig. 2) first through the Greater Antilles and the Avies Ridge (Iturralde & McPhee, 1999) and later through the uplift of the Panama Isthmus (3.5 Ma). In the Late Cretaceous-Early Paleocene (second time slice, Fig. 2), the southeastern part of the Malaysian peninsula and southwest Borneo (hereafter referred as “paleo-D”) were already in place (see above and Fig. 2 in Medcalfe, 1998), and dispersal between Eurasia (A) and this part of SE Asia became possible (second time slice; Fig. 2). During the Paleogene (third time slice; Fig. 2), as India rapidly drifted northwards on its way to collide with Eurasia, it formed a possible dispersal route between proto-SE Asia (D), Madagascar (C) and Africa (B), all of which were by then tectonically isolated (i.e., the so-called Indo-Madagascar connection). By the Late

Eocene (third time slice; Fig. 2), all areas were in place, except for the remaining of area D (insular SE Asia and Pacific islands), which appeared later in the Early Oligocene (fourth time slice; Fig. 2). At this time, most Gondwanan landmasses had broken up from Antarctica, except for South America (G) and Australia (E), which were still connected across the Antarctic Peninsula (Sanmartín, 2005; Fig. 2). In the Paleogene the northeastward movement of the African plate (B) ended with its collision with Eurasia, closing the Tethyan Seaway and ending the isolation between Europe and Africa. Therefore, dispersal between these two areas (A and B) was allowed from this time slice onwards (Fig. 2). By the Early Oligocene (fourth time slice; Fig. 2), the Australian Plate (E), which had started to drift northwards towards Eurasia (third time slice; Fig. 2), collided with the Eurasian Plate; this resulted in the formation of many of the SE Asian islands (D) and the starting of biotic exchange (dispersal events) with the neighbouring regions (Fig. 2; notice that dispersal between D and E was already allowed in the third time slice but with a lower rate, 0.5). The final opening of the Drake Passage between South America and Antarctica broke the connection with Australia. This and the northward drift of Australia led to the establishment of the equatorial and west wind drift currents (Sanmartín et al., 2007), which make dispersals between the Southern Hemisphere landmasses once again possible, albeit with a lower probability than land dispersal through direct tectonic connections.

Dating and biogeographic uncertainties – Impact of divergence time uncertainty on biogeography was assessed by performing Lagrange (M1 and Str) analyses on minimum (Min PL), median (Med PL) and maximum (Max PL) estimations of node ages in the PL chronogram. We were especially interested in those nodes for which the interval between minimum and maximum node age (i.e., confidence intervals) spread across two different time slices. To test the congruence between biogeographic scenarios among different inference methods, we recorded the ancestral range with the maximum probability for each node in the phylogeny, and compared these values pairwise among the three methods: Bayes-DIVA, Lagrange M1, and Lagrange Str. We defined two types of incongruence affecting nodal

ancestral area assignment: a) "Hard incongruence": the analyses inferred different ancestral range/s, and b) "Soft incongruence": the analyses inferred the same ancestral range/s but with different probabilities. We also computed the mean of the maximum probabilities for area assignments across all nodes in the Med PL chronogram for each of the three methods and compared the means statistically by using a Wilcoxon rank test (Bauer, 1972). This provided a first evaluation of the level of biogeographic uncertainty in each method: the higher the level of uncertainty, the lower the mean of maximum probabilities assigned to the ancestral areas.

Finally, we plotted ancestral ranges inferred by each of the three methods on the Med PL chronogram and compared them pairwise to identify differences in the biogeographic scenarios. Relative probabilities for ancestral range at internal nodes were represented by pie charts and cases of hard incongruence between methods were also pinpointed (see Figs. 7-9). We used these data together with the transition matrix describing biogeographic movement between states (Tables 1, 2) to estimate the frequency of dispersal events for Bayes-DIVA, Lagrange M1 and Lagrange Str, as well as the frequency of local extinction events for Lagrange M1 and Lagrange Str (Bayes-DIVA does not estimate extinction, see below). To compute these values: (i) we used the statistical language R (R Development Core Team 2009) and the APE package (Paradis et al., 2004) to extract for each node the most likely ancestral area, the node age, and the branch length uniting this node with the descendant node. These data were used to build dispersal/extinction contingency tables showing the frequency and type of transition events (changes in geographic range) between ancestral and descendant nodes in the phylogeny. These contingency tables were used to estimate the frequency and direction of dispersal events and the frequency of extinction events (observed and predicted by the model) according to the cost matrix or transition model implemented in DIVA and Lagrange. Finally, the number of dispersal events optimized at internal and terminal branches were compared across methods using a paired Wilcoxon rank test.

Results

Divergence time estimations

Divergence times were estimated based on the allcompat tree from MrBayes. Inter-relationships among genera and supra-specific groups in this tree were in agreement with Buerki et al. (2009), and nearly 95% of all tree nodes were strongly supported [Bayesian posterior probability (BPP) > 0.95]. For example, of the six calibration points assigned to nodes or lineage divergence splits, only one node (E) received a BPP value lower than 1 (BPP = 0.98) (see Fig. 3 for the location of calibration points).

The assumption of a molecular clock was strongly violated in the LR test ($p < 0.0001$), reinforcing the need to use a relaxed-clock method such as PL or NPRS to estimate divergence times. Figure 3 shows mean values and 95% confidence intervals of node ages for the Bayesian consensus as estimated by NRPS and PL methods. Uncertainty in age estimation in NPRS (as evidenced by the length of the confidence interval) showed no correlation to nodal age (Fig. 4A), whereas PL showed a positive correlation between nodal age and dating uncertainty, with the length of confidence intervals generally decreasing from older (basal) to younger (more distal) nodes (Fig. 4B). Moreover, in general confidence intervals were larger for NPRS than for PL ($\chi^2=98.92$; $P \text{ value} < 10^{-4}$; Fig. 4C) and estimated mean age values were consistently older in NPRS than in PL, regardless of the level of dating uncertainty and nodal age (Fig. 4D; paired Wilcoxon rank test: $V=392$; $P \text{ value} < 10^{-4}$). These results suggest that PL estimates are more robust (less sensitive) to dating uncertainties than NPRS estimates. Previous studies (Linder et al., 2005; Bell & Donoghue, 2005) have shown that PL is also less sensitive to taxon sampling than NPRS. Therefore, biogeographic analyses were only performed on the chronograms derived by PL and discussion of biogeographic events and paleogeographic scenarios will be done considering only the PL divergence time estimates since they show higher robustness to dating uncertainty than NPRS.

Biogeographic analyses

Combining uncertainties in divergence time and biogeography – In general, Bayes-DIVA showed significantly higher maximum probabilities assigned to ancestral area(s) per node than any of the two Lagrange methods (Wilcoxon rank test: $W=18189$; P value $< 10^{-4}$; Fig. 5A). Maximum probabilities for ancestral ranges were also higher in Lagrange Str than in Lagrange M1 but the difference was not significant ($W = 9639$; P value = 0.110; Fig. 5A). Also, Bayes-DIVA showed a significantly higher number of inferred dispersal events at terminal branches than at internal branches (Wilcoxon Rank test: $W=9078$; P value=0.01; Fig. 5B), whereas Lagrange M1 and Str showed no significant difference between these two types of dispersal events (P value of 0.507 and 0.413, respectively; Fig. 5B). The dispersal rate was 30 times higher in Lagrange M1 ($d = 0.025$) than in Lagrange Str ($d = 0.009$) (Fig. 5C).

Numbers of soft and hard incongruities within and between methods are indicated in Table 3 and in Figs. 6-9 (information about node numbers are provided in Fig. S1). Soft incongruities were mainly restricted to basal nodes; for example, for nodes 151 and 152 corresponding to the most recent common ancestors (MRCA) of Sapindaceae (Fig. S1), all three analyses (Bayes-DIVA, Lagrange M1 and Str) recovered area “A” as the ancestral range but with different probability values (Figs. 6-9; Table 3). Hard incongruence cases were common both at terminal nodes connecting with widespread taxa – for example, node 167 (Fig S1) leading to *Dodonaea viscosa*, a widespread species distributed in six of the seven areas, abcdefg (Figs. 6-9 and Tables 3, 4), and at nodes that are shared between two time slices, for example node 294 (Figs. 6-9; Tables 3, 4). The first type of hard incongruence was more frequent between Bayes-DIVA and the Lagrange methods, whereas the second type was more likely to occur between Lagrange Str and the other two methods. Figure 6 shows that age incongruence was to some extent correlated with biogeographic uncertainty, since the number of soft and hard incongruent nodes between the three methods was found to increase around the point where there is a transition between time slices: for example, there is an increase in dating uncertainty at node 154 between time

slices one and two; this agrees with an increase in biogeographic uncertainty – both in the area assigned (hard incongruence) and its level of support (soft incongruence) – among the three methods for this node and surrounding nodes (nodes 294, 154, 160 and 174; Fig. 6). The same pattern can be found for node 202 and surrounding nodes 175-265 between time slices two and three. Moreover, this pattern can also be found *within* time slices: for example there is an increase in biogeographic uncertainty for nodes 161-164 within the third time slice that is coincident with higher dating uncertainty in these two nodes. Curiously, the largest number of hard incongruent nodes was found between time slices two and three, coincident with a dramatic change in continental configuration (see Fig. 2). No clear pattern was observed linking soft and/or hard incongruence to node age, although most hard incongruities were likely to occur in basal nodes (Fig. 6).

Biogeographic scenarios (Figs. 7-9; Tables 5-6; Supplementary Table 1-3)

For each method, the main biogeographic movements through time as inferred from contingency tables (see and Supplementary Tables 1-3) are depicted on paleogeographic maps representing the four time slices (see Figs. 7-9). Comparison of the frequency of dispersal and extinction events among the three methods (Table 5) indicates that Bayes-DIVA inferred more dispersal events than Lagrange M1 and Str (70, 41 and 57 dispersal events, respectively, see Table 5). The difference is highest for the third time slice, with 13 dispersals in Bayes-DIVA and only 3 and 7 dispersals in Lagrange M1 and Lagrange Str (Figs. 7-9; Table 5). This suggests that dispersal events in Lagrange are delayed in time in comparison to Bayes-DIVA (Figs. 7-8). Although Lagrange M1 incorporates time into biogeographic inference through the use of branch lengths, it does not incorporate prior information on the changing paleogeography into the biogeographic model. This method is therefore important for understanding the differences between the three methods caused by (i) the different assumptions of the biogeographic model (i.e., DIVA versus Lagrange) and/or (ii) the integration of paleogeographic information into biogeographic inference (Lagrange M1 vs. Lagrange Str). When considering paleogeographic connections, Lagrange M1 was in

better agreement with the paleogeographic scenario (and therefore with Lagrange Str) than Bayes-DIVA. The latter reconstructed dispersals that were not allowed by the model; for example, the dispersal event from A to D in the first time slice (area D did not exist at this point), followed by a second dispersal from D to E in the second time slice (these two areas were separated by an ocean gulf at this time, Fig. 7; Table 5; see Fig. 2 for more details). In general, for the same type of dispersal event, Bayes-DIVA tended to infer them slightly earlier in time (“accelerated dispersal”) than either of the two Lagrange methods, although the difference was larger with Lagrange Str, as may be expected. Moreover, some unexpected relationships were observed involving jump dispersals in Bayes-DIVA, such as the dispersal event from Eurasia (A) to South America (G) in the third time slice (no direct connection between these two areas at this time, see Fig. 2), or the dispersal from G to C in the fourth time slice (Fig. 7), which also contradicts our paleogeographic scenario (Fig. 2). Jump dispersal from A to G was also inferred by Lagrange M1 (Fig. 7) albeit with a low probability. Regarding extinction, Lagrange M1 and Lagrange Str analyses inferred respectively 33 and 37 extinction events distributed over the third and fourth time slices (Table 6). Among areas, areas B and D had the highest number of inferred extinctions, whereas only one extinction event was observed in area F in both analyses (Table 6; Fig. 9). Regardless of the biogeographic model and assumptions of each model, all three methods agree on inferring that the Sapindaceae originated in Eurasia (area “A”; most likely in temperate regions of Asia) around the Early Cretaceous, with subsequent dispersal to SE Asia between the Late Cretaceous to Early Paleocene (Figs. 7-8). Biogeographic scenarios also suggest that the SE Asia region has been a centre of diversification and dispersal over the tropics for Sapindaceae (Figs. 7-8). This is particularly true during the fourth time slice (from the Early Oligocene to present) with dispersals from SE Asia (D) to Australia (E), Africa (B) and Madagascar (C), among others (Fig. 7-8). Reverse migrations are also observed in our biogeographic reconstruction, with several dispersal events from Africa (C) and Australia (E) to SE Asia (D), and even from Eurasia (A; most likely Indochina) to SE Asia, sometime in the Late Oligocene (Figs. 7-8). South American lineages of Sapindaceae probably originated

from Australian ancestors, with at least three dispersals inferred during the Paleogene (probably through Antarctica) and the Miocene (probably mediated by the West Wind Drift current) (Figs. 2, 9). Inferred exchanges between South America and Africa during this time slice (Fig. 9) were probably mediated by the South Equatorial Current (Fig. 2).

Discussion

Definition of areas in parametric biogeography

Ree & Sanmartín (2009) pointed out that the definition of operational areas in a biogeographic analysis constitutes a critical step in parametric biogeography – especially for models considering widespread states (Lagrange) – because, the size of the transition probability matrix (Q) increases exponentially with the number of areas. Actually, careful area definition is also important for parsimony-based methods (DIVA), since a greater number of areas implies less amount of repeatable biogeographic data across taxa and this increases the difficulty of extracting general patterns (e.g., dispersal frequencies). In plants, the definition of areas for biogeographic analysis, especially at worldwide scale, remains controversial (e.g., Takhtajan, 1978, 1986; Cox, 2001). In general, most researchers circumscribe areas according to the ecological tolerance and current distribution patterns of their group of interest (“criterion of sympatry”, Sanmartín, 2009, in press). This has the problem that different groups may exhibit different distribution patterns and area definition may not be comparable across groups. In this study, we used the geological criterion of Sanmartín & Ronquist (2004) and defined areas of study according to paleogeographic history (i.e., plate tectonics). This approach has the advantage that it decreases the subjectivity in area definition and can be applied to different plant groups, but it also requires an accurate knowledge of the geological history of the areas. Another difficulty in area definition is that the area circumscription may change through time, as different terranes break up and collide (Ree & Sanmartín, 2009). For instance, area D (SE Asia) has a complex geological history implying different configurations through time (see Material and Methods

and Fig. 3 in Metcalfe, 1998 for more details), and this needs to be considered for defining both area boundaries (Fig. 1) and their connection to other areas (Fig 2).

A stratified biogeographic model: Ree & Sanmartín (2009) argued that when the number of states increases relatively to the amount of data (especially in the case of Lagrange), it becomes increasingly worthwhile to impose geographic structure on model parameters governing the transitions between those states. We have used this approach in our Lagrange Str model, in which dispersal rates were scaled based on the availability of area connections through time. To our knowledge, this is the first time such an approach has been used in Lagrange at a worldwide scale. Ree et al. (2005) proposed a similar biogeographic-paleographic model for the Northern Hemisphere, which has later been used (and extended) by other authors (e.g., Moore & Donoghue, 2007). However, this model was based on the first DEC approach proposed by Ree et al. (2005) – implemented in the program AREA (Smith, 2006) – in which the availability of area connections is modelled by an exponential parameter representing the probability of successful dispersal through time. However, because the Q transition matrix was calculated for each branch length independently using simulations rather than estimated analytically from the data as in Lagrange (Ree & Smith, 2008), this method becomes intractable for large phylogenies and complex biogeographic scenarios such the one analysed here. Instead, we propose for the first time a stratified worldwide biogeographical model that takes into account area histories and connections through time (Fig. 2) and that can be used to estimate ancestral ranges and range inheritance scenarios in large phylogenies using the maximum likelihood framework.

One important point to consider when constructing stratified models is the reducibility of the Markov chain biogeographic model. In the stratified model, range evolution along internodes (branch lengths) within each time slice is determined by the corresponding Q matrix of that particular period. Between time slices, where internal branches cross boundaries between time slices, likelihood for ancestral ranges are conditional on those ranges being a valid outcome of the preceding period, that is, they are allowed by the Q

matrix from the preceding period (Ree & Smith, 2008). This may introduce problems with the reducibility of the Markov chain Q matrices: a Markov chain is said to be irreducible if its state space is a communicating class, that is, if it is possible to get to any state from any state. Direct dispersal between two areas can be disallowed in a biogeographic scenario by giving it a rate of 0 in the Q matrix. However, it should still be possible to move between those two areas or states in the model by using other dispersal pathways, that is, by moving through intermediate areas in the Q matrix. In constructing stratified biogeographic models in which some dispersal pathways are disallowed within time slices, there is a high risk of encountering problems with reducible Markov chains *between* time slices. For example, in our stratified model we encountered the problem with branches that crossed between the second and third time slice because this point marks an important change in continental configuration: in time slice 2, Gondwanan landmasses (B, C, E, G) were still connected but they broke apart and moved northwards to establish new connections with Laurasia at time slice 3 (see Fig. 2). One solution to this problem (applied in our study) is to replace all rate values of "0" in the Q matrix by low but positive values of "0.01". This solution is also biologically sound since dispersal events between disjunct landmasses, even though rare, can still be possible even more so for plants with easily dispersed seeds. In general, disallowing dispersal between areas by assigning them a rate of "0" makes sense in the case of volcanic island systems such as Hawaii, in which islands "pop out" from the ocean in a sequence and they cannot be colonized before their time of emergence (see Ree & Smith, 2008). However, it makes less sense for continental scenarios in which the landmasses were always present, although with different size and area connectivity across time slices. The only exception to this rule was area D, for which dispersal to/from any other area was given a rate of "0" in the first time slice (Fig. 2), as this area as we define it here has not yet been accreted at that time.

Impact of divergence time uncertainty on biogeography (Fig. 6 and Tables 3, 4)

Most biogeographic studies do not incorporate the error (stochastic variance) associated with estimating lineage divergence times on phylogenies. At most, they represent dating uncertainty as confidence intervals for divergence time estimates at nodes, and they interpret the results from the biogeographic reconstruction in view of this uncertainty. Here, we show that dating uncertainty may have some impact on the biogeographic reconstruction itself, both in the assignment of ancestral ranges to nodes (hard incongruities) and in the “decisiveness” with which an area can be assigned to a node, that is, the different maximum probabilities assigned to the same ancestral area (i.e., soft incongruities; Fig. 6 and Table 3). Our study suggests that biogeographic uncertainty increases at nodes where dating uncertainty is also high, i.e., where the difference between maximum and minimum ages (ΔPL) is largest (Fig. 6). This can be observed in both Lagrange M1 and Lagrange Str (Fig. 6) but the effect of uncertainty is more severe in Lagrange Str because its biogeographic model is made directly dependent on a temporally stratified paleogeographic model. This can be observed in the fact that Lagrange Str exhibits the largest number of hard incongruent nodes (Table 3) and that these are mainly located at nodes whose confidence intervals span two different times slices (Table 4). On the other hand, Lagrange M1 exhibits a higher percentage of soft incongruities at nodes than the stratified model (Table 3), which fits with the generally higher uncertainty in Lagrange M1.

Bayes-DIVA vs. Lagrange: a multivariate problem! (Figs. 5-9; Table 3)

Clark et al. (2008) recently reviewed different methods of ancestral range reconstruction using island systems as a case-study biogeographical scenario. They showed that by considering branch lengths and/or the timing of events, parametric methods gives more accurate, nuanced inferences of ancestral ranges and biogeographic history than parsimony-based approaches, but that the former methods could benefit from adopting a Bayesian strategy to incorporate phylogenetic uncertainty into biogeographical inference. They concluded that future comparison of methods using different biogeographic systems

such as continental scenarios, are needed to illustrate further the comparative performance of these (parsimony *versus* parametric) methods (page 705, Clark et al., 2008). Continental scenarios differ from island systems in that areas are contiguous (i.e., they share an edge), so widespread ancestral ranges – ancestral distributions formed by two or more areas – are valid states in the biogeographic model and dispersal is generally modelled as the equivalent to “range expansion” (the ancestor moves into a new area but also retains its original distribution) followed by range division. Both dispersal-vicariance analysis (DIVA) and the DEC likelihood model are based on this model of “vicariance-mediated allopatry”, in which dispersal is modelled as occurring along the branches leading to the widespread node, followed by range division by vicariance (DIVA) or by vicariance and/or peripatric speciation (Lagrange). In these models, dispersal leads to vicariance but it is not directly associated with cladogenesis (Sanmartín, 2007). This contrasts with character evolutionary models, such as Fitch Parsimony, in which dispersal between single areas is optimized onto the branches subtending from speciation events (Sanmartín, 2007), or the Bayesian approach to island biogeography developed by Sanmartín et al. (2008), in which range evolution is modelled as “dispersal-mediated allopatry”, that is, areas are isolated by barriers and dispersal is immediately followed by speciation (“jump dispersal”; Clark et al., 2008; Sanmartín, 2009, in press). These methods are more appropriate for island systems in which species are not expected to retain their widespread distribution for long after dispersal. Speciation takes place instantaneously after dispersal: as soon as the founder population moves into a new area, it becomes a new species (Clark et al., 2008; Sanmartín, 2009, in press; Ree & Sanmartín, 2009).

Despite these similarities, the models implemented by Lagrange and DIVA differ in two aspects that have an effect on biogeographic scenarios. In DIVA, widespread ancestral ranges are always divided at speciation nodes by vicariance: descendants inherit non-overlapping subsets of areas. Lagrange is more flexible than DIVA in allowing alternative modes of speciation (i.e., range inheritance scenarios). For widespread ancestral ranges, lineage divergence could arise either between a single area and the rest of the range

(“vicariance”) or within a single area, in which case one daughter species inherits a single-area range and the other inherits the entire ancestral range (“peripatric speciation”; Ree et al., 2005). Such scenarios lead to non-identical range inheritance. The effect of this distinction is twofold: first, DIVA reconstructions tend to overestimate the frequency of terminal dispersal events compared to Lagrange (Fig. 5B; see also Nylander et al., 2008a). Enforcing vicariance on widespread ancestors means that it is more parsimonious for DIVA to explain widespread terminal ranges by assuming terminal dispersal events than inheritance of a more reduced ancestral range with several subsequent internal dispersals. For example, DIVA would model range evolution from A to ABC as sympatric speciation followed by two terminal dispersals (+C + B), whereas Lagrange would model it as a widespread ancestor (AB or AC) followed by one dispersal event (+B or + C). Therefore, allowing range inheritance by peripatric speciation has the effect in Lagrange of increasing the uncertainty in ancestral geographic ranges (i.e., there are more alternative scenarios; see Fig. 5A) but it also diminishes the frequency of terminal dispersals (Fig. 5B). Another consequence of the different treatment of widespread ranges is that dispersal events in Lagrange are delayed in time when compared to the same events in DIVA (e.g., dispersal from D to E in Figs. 7-8). Since Lagrange allows widespread ancestral ranges to be maintained through speciation events (peripatric speciation), there is no need to postulate “intermediate” dispersals to explain why two consecutive ancestral nodes have the same widespread range (AB to AB). For example, the widespread area DE is maintained in Lagrange M1 and Str from node 264 to the terminals, whereas DIVA inferred the ancestor in area E and subsequent dispersal events to D in terminal branches (Figs. 7-9; S1).

Another difference between the DIVA and Lagrange models is the inference of extinction events. As discussed by other authors (Ree et al., 2005; Sanmartín, 2007; Nylander et al., 2008a), DIVA would never infer extinction events unless explicit geographical constraints are placed on the model (Ronquist, 1996). This contributes to an increase in the number of terminal dispersal events, since cases where one descendant is widespread and the other has a more restricted distribution would be inferred by secondary (post-speciation)

dispersal in the most widespread descendant (Nylander et al., 2008a). Lagrange instead could explain this by peripatric speciation (one descendant inherits the whole ancestral range) or by extinction. Lagrange does explicitly incorporate extinction into the model as a parameter in the Q matrix producing range contraction, and in this respect it is more realistic than DIVA. However, by forcing dispersal events between single areas to go through a widespread state (see Table 2), Lagrange may overestimate the frequency of extinction, especially for those areas where there is a high frequency of interchange with other areas, e.g., area D works as a sort of “dispersal highway” to other areas in time slice 4 (Fig. 7-8; Table 8). There are two types of extinction events in Lagrange models: “observed” extinction events in which an ancestral range is reduced by range contraction (e.g., AB to A in Table 2), and unobserved extinction events that are predicted by the model (e.g., “jump dispersal” A to G requires Dag + Ea in Table 2). These constraints may be realistic in cases of contiguous areas or areas that come together after the disappearance of a barrier but they are less so in cases where areas are separated by ocean barriers such as in dispersal mediated by the West Wind Drift (Fig. 2).

Another interesting distinction between the three methods is their different “decisiveness”, that is, the different level of biogeographic uncertainty in assigning ancestral ranges to nodes, “soft incongruence” (Fig. 6; Table 3). Bayes-DIVA is the most “decisive”: maximum probabilities assigned to ancestral ranges at nodes were significantly higher than in the two Lagrange analyses (Figs. 5A, Figs. 6-9). Lagrange Str was more “decisive” than Lagrange M1, although the difference was not significant. However, it is curious to notice that Lagrange Str has levels of decisiveness, measured as the number of soft incongruent nodes, more similar to Bayes-DIVA than to Lagrange M1 (Table 3). One reason for this is that Lagrange M1 is more similar to the parsimony-based DIVA model in that it has a flat prior on dispersal rates: all dispersal pathways are allowed in the Q matrix and this fact translates into more uncertainty in inferring ancestral ranges. In contrast, Lagrange Str, by imposing geographic structure to the model – by assigning lower rates to certain dispersal pathways – results in an

increase in decisiveness and a decrease in uncertainty compared to Lagrange M1. In a similar way, by incorporating uncertainty in the tree topology – the only parameter or variable considered by DIVA when inferring ancestral ranges – the Bayesian empirical approach to DIVA, Bayes-DIVA, helps to increase the decisiveness (confidence) in ancestral range reconstruction. In this sense, Bayes-DIVA and Lagrange Str. are more similar in that they allow more accurate biogeographic reconstructions by incorporating uncertainty in the parameters used in biogeographic inference: topology (Bayes-DIVA) and dispersal rates (Lagrange Str).

Perspectives

There are few points that we should consider in further biogeographic investigations. The first point focuses on the way to deal with hard incongruities between min and max divergence time estimations. One possibility could be to simply combine biogeographic results inferred from min and max divergence time estimations by averaging the probabilities from each. A second point that we should investigate deals with the quantification of large-distance migration in parametric biogeographic inference. Currently, in the Q matrix, all areas that are supposed to show a very low probability of dispersal have a “0.01” probability. However, should not this value be modulated to reflect possible paleowinds or paleo-sea-currents? What do we know about them? Also in the context of large-distance dispersal, would we expect that North-South events are more likely than East-West events involving species with fleshy fruits due to latitudinal bird migrations? In contrast, for anemochorous fruits, East-West winds might be of major importance. Thus, a next step might be to reconstruct ancestral character states for the fruit morphology (e.g., fleshy vs. non-fleshy fruits; winged vs. non-winged fruit) to infer node-by-node the probability of migration through large distances. This would result in a Q matrix that would vary node by node as a function of fruit morphology.

Acknowledgments

The first author is grateful to Philippe K pfer and Giorgina Bernasconi (University of Neuch tel, Switzerland) for financial support and comments on previous version of the manuscript. We are grateful to Frederik Ronquist (Uppsala University, Sweden) for valuable comments regarding biogeographic analyses. We thank Pedro Acevedo-Rodr guez (Smithsonian Institute, U.S.A.), Mark Harrington (James Cook University, Australia), Caroline Weckerle (Institute of Systematic Botany, Z rich, Switzerland), the Royal Botanic Gardens, Kew (U.K.), the Missouri Botanical Garden (U.S.A) and Laurent Gautier, Louis Nusbaumer and Patrick Ranirison from the Conservatory and Botanical Garden of Geneva, Switzerland for plant material. Financial support to the first author was provided by the laboratory of evolutionary Biology (University of Neuch tel, Switzerland) and by the European Community's Programme Structuring the European Research Area under a SYNTHESYS grant to visit the Real Jard n Bot nico (CSIC) in Madrid.

References

- Adema, F., Leenhouts, P.W., van Welzen, P.C., 1994. Sapindaceae. *Flora Malesiana Ser. I* 11, 419–768.
- Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Akademiai, K. (Ed.), *Second International Symposium on Information Theory*. Budapest, pp. 267–281.
- Antonelli, A., Nylander, J.A.A., Persson, C., Sanmart n, I., 2009. Tracing the impact of the Andean uplift on Neotropical plant evolution: evidence from the coffee family. *Proc. Natl. Acad. Sci. USA*. In press.
- Bauer, D.F. 1972. Constructing confidence sets using rank statistics. *J. Amer. Statistical Assoc.* 67, 687–690.
- Bell, C.D., Donoghue, M.J., 2005. Dating the diversification of Dipsacales: comparing models, genes, and evolutionary implications. *Am. J. Bot.* 92, 284–296.

- Berry, E.W., 1916. The lower Eocene floras of southeastern North America. USGS Prof. Pap. 91, 1—481.
- Berry, E.W., 1930. Revision of the lower Eocene Wilcox flora of the southeastern states, with descriptions of new species, chiefly from Tennessee and Kentucky. USGS Prof. Pap. 156, 1—196.
- Briggs, J.C., 1994. The genesis of Central America: Biology versus geophysics. *Glob Ecol Biogeogr* 4,169—172.
- Brooks, D.R., 2005. Historical biogeography in the age of complexity: expansion and integration. *Revista Mexicana de Biodiversidad*, 76, 79—94.
- Brooks, D.R., MacLennan, D.A., 2003. Extending phylogenetic studies of coevolution: secondary brooks parsimony analysis, parasites, and the great apes. *Cladistics* 19, 104—119.
- Brown, R.W., 1934. The recognizable species of the green river flora. USGS Prof. Pap. 185-C, 45—77.
- Brown, R.W., 1935. Miocene leaves, fruits, and seeds from Idaho, Oregon, and Washington. *Journal of Paleontology* 9, 572—587.
- Brown, R.W., 1937. Additions to some fossil floras of the western USGS Prof. Pap. 1865, 163—206.
- Buerki, S., Forest, F., Acevedo-Rodríguez, P., Callmander, M.W., Nylander, J.A.A., Harrington, M., Sanmartín, I., Küpfer, P., Alvarez, N., 2009. Plastid and nuclear DNA markers reveal intricate relationships at subfamilial and tribal levels in the soapberry family (Sapindaceae). *Mol. Phylogenet. Evol.* 51, 238—258.
- Clark, J.R., Ree, R.H., Alfaro, M.E., King, M.G., Wagner, W.L., Roalson, E.H., 2008. A comparative study in ancestral range reconstruction methods: retracing the uncertain histories of insular lineages. *Syst. Biol.* 57, 693—707.
- Cox, C.B., 2001. The biogeographic regions reconsidered. *J. Biogeogr.* 28, 511—523.
- Croizat, L., 1952. *Manual of phytogeography*. The Hague, Dr. W. Junk, Netherlands.

- Doyle, J.A., Hotton, C.L., 1991. Diversification of early angiosperm pollen in a cladistic context. In: Blackmore, S., Barnes, S.H. (Eds.) Pollen and spores: patterns of diversification. Clarendon Press, Oxford, U.K, pp. 169–195.
- Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7, 214.
- Ebach, M.C., Humphries, C.J., Williams, D.M., 2003. Phylogenetic biogeography deconstructed. *J. Biogeogr.* 30, 1285—1296.
- Felsenstein, J., 1988. Phylogenies from molecular sequences: inference and reliability. *Annu. Rev. Genet.* 22, 521–565.
- Gradstein, F.M., Ogg, G.O., 2004. Geologic Time Scale 2004 - why, how, and where next! *Lethaia* 37, 175—181.
- Graham, A., 1976. Studies in Neotropical paleobotany. II. The Miocene communities of Veracruz, Mexico. *Ann. Missouri Bot. Gard.* 63, 787–842.
- Harrington, M.G., Edwards, K.J., Johnson, S.A., Chase, M.W., Gadek, P.A., 2005. Phylogenetic inference in Sapindaceae sensu lato using plastid *matK* and *rbcL* DNA sequences. *Syst. Bot.* 30, 366-382.
- Inda, L.A., Segarra-Moragues, J.A., Müller, J., Peterson, P.A., Catalán, P., 2008. Dated historical biogeography of the temperate Loliinae (Poaceae, Pooideae) grasses in the northern and southern hemispheres. *Mol. Phylogenet. Evol.* 46, 932—957.
- Iturralde-Vinent, M.A., MacPhee, R.D.E., 1999. Paleogeography of the Caribbean region: implications for Cenozoic biogeography. *Bull. Am. Mus. Nat. Hist.* 238, 1—95.
- Kemp, E.M., 1976. 1976. Early Tertiary pollen from Napperby, central Australia. *B.M.R. J. Aust. Geol. Geophys.* 1, 109–114.
- Kluge, A.G., 1989. A Concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Syst. Zool.* 38, 7-25.
- Leopold, E.B., MacGinitie, H.D., 1972. Development and Affinities of Tertiary Floras in the Rocky Mountains, In: Graham, A. (Ed.), *Floristics and Paleofloristics of Asia and Eastern North America*. Elsevier Pub. Co., Amsterdam, pp. 147-200.

- Linder, H.P., Hardy, C.R., Rutschmann, F., 2005. Taxon sampling effects in molecular clock dating: an example from the African Restionaceae. *Mol. Phylogenet. Evol.* 35, 569—582.
- Mabberley, D.J., 2008. *Mabberley's plant-book*. Cambridge University Press, Cambridge.
- Manchester, S.R., 1999. Biogeographical Relationships of North American Tertiary Floras. *Ann. Missouri Bot. Garden* 86, 472—522.
- Manchester, S.R., 2001. Leaves and Fruits of *Aesculus* (Sapindales) from the Paleocene of North America. *Int. J. Plant Sci.* 162, 985—998.
- Manchester, S.R., Tiffney, B.H., 2001. Integration of paleobotanical and neobotanical data in the assessment of phytogeographic history of holarctic angiosperm clades. *Int. J. Plant Sci.* 162, S19—S27.
- Magallón, S., Sanderson, M.J., 2001. Absolute diversification rates in angiosperms. *Evolution* 55, 1762–1780.
- McClain, A.M., Manchester, S.R., 2001. *Dipteronia* (Sapindaceae) from the Tertiary of North America and implications for the phytogeographic history of the Aceroideae. *Am. J. Bot.* 88, 1316—1325.
- Metcalf, I., 1998. Paleozoic and Mesozoic geological evolution of the SE Asia region: multidisciplinary constraints and implications for biogeography. In: Halle, R., Holloway, J.D. (Eds.), *Biogeography and Geological Evolution of SE Asia*. Leiden, Backhuys Publisher, pp. 25—41.
- Moore, B.R., Smith, S.A., Ree, R.H., Donoghue, M.J. Incorporating fossil data in biogeographic inference: a likelihood approach. *Evolution* (in press).
- Moore, B. R., Donoghue, M.J., 2007. Correlates of diversification in the plant clade Dipsacales: Geographic movement and evolutionary innovations. *Amer. Nat.* 170, S29—S55.
- Muellner, A.N., Vassiliades, D.D., Renner, S. S., 2007. Placing Biebersteiniaceae, a herbaceous clade of Sapindales, in a temporal and geographic context. *Pl. Syst. Evol.* 266, 233-252.

- Müller, J., Leenhouts, P. W., 1976. A general survey of pollen types in Sapindaceae in relation to taxonomy. In: Ferguson, I.K., Müller, J. (Eds.), The evolutionary significance of the exine. London, Academic Press, pp. 407–445.
- Müller, J., 1981. Fossil pollen record of extant angiosperms. *Bot. Rev.* 47, 1—140.
- Nelson, G.J., Platnick, N.I., 1981. Systematics and Biogeography: cladistics and vicariance. Columbia University Press, New York.
- Nylander J.A.A., 2002. MrModeltest v.1.0. Program distributed by the author. Department of Systematic Zoology, Uppsala University, Uppsala. Website <http://www.ebs.uu.se/systzoo/staff/nylander.html>
- Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P., Nieves-Aldrey, J.L., 2004. Bayesian phylogenetic analysis of combined data. *Syst. Biol.* 53, 47—67.
- Nylander, J.A.A., Olsson, U., Alström, P., Sanmartín, I., 2008a. Accounting for Phylogenetic Uncertainty in Biogeography: A Bayesian Approach to Dispersal-Vicariance Analysis of the Thrushes (Aves: *Turdus*). *Syst. Biol.* 57, 257—268.
- Nylander, J.A.A., Wilgenbusch, J.C., Warren, D.L., Swofford, D.L., 2008b. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics* 24, 581—583.
- Page, R.D.M., 2003. Tangled Trees: Phylogeny, Cospeciation, and Coevolution. University Of Chicago Press.
- Paradis, E., Claude, J., Strimmer, K., 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20, 289—290.
- Parenti, L., 2006. Common cause and historical biogeography. *Biogeography in a changing world* (ed. by M.C. Ebach and R.S. Tangney), pp. 61–71. CRC Press, London.
- Pfossner, M.F., Guzy-Wróbelska, J., Sun, B.-Y., Stuessy, T.F., Sugawara, T., Fujii, N., 2002. The origin of species of *Acer* (Sapindaceae) endemic to Ullung Island, Korea. *Syst. Bot.* 27, 351—367.
- Piel, K.M., 1971. Palynology of Oligocene sediments from central British Columbia. *Canad. J. Bot.* 49, 1885–1920.

- Posada, D., Crandall, K. A., 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14, 817–818.
- Rambaut, A., Drummond, A.J., 2004. Tracer v1.3 <http://evolve.zoo.ox.ac.uk/software.html>
- Ree, R.H., Moore, B.R., Webb, C.O., Donoghue, M.J., 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution* 59, 2299—2311.
- Ree, R.H., Smith, S.A., 2008. Maximum Likelihood Inference of Geographic Range Evolution by Dispersal, Local Extinction, and Cladogenesis. *Syst. Biol.* 57, 4—14.
- Ree, R.H., Sanmartín, I., 2009. Prospects and challenges for parametric models in historical biogeographical inference. *J. Biogeogr.* (in press).
- Roquet, C., Sanmartín, I., García-Jacas, N., Sáez, L., Wikström, N., Aldasoro, J.J. Reconstructing the history of Campanulaceae s. str. with a Bayesian approach to molecular dating and dispersal-vicariance analysis. *Mol. Phylogenet. Evol.* (in press).
- Ronquist, F., 1996. DIVA version 1.1. Computer program and manual available by anonymous FTP from Uppsala University (<ftp.uu.se> or <ftp.systbot.uu.se>).
- Ronquist, F., 1997. Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Syst. Biol.* 46, 195–203.
- Ronquist, F., 1998. Phylogenetic approaches in coevolution and biogeography. *Zoologica scripta* 26, 313—322.
- Ronquist, F., 2001. DIVA 1.2. Computer programme distributed by the author. Available from <http://www.ebc.uu.se/systzoo/research/diva/diva.html>.
- Ronquist, F., 2003. Parsimony analysis of coevolving associations. *Tangled trees. Phylogeny, cospeciation, and coevolution.* Univ. of Chicago Press, Chicago, IL:22–64.
- Ronquist, F., 2004. Bayesian inference of character evolution. *Trends Ecol. Evol.* 19, 475—481.
- Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572—1574.

Sanderson, M.J., 1997. A non-parametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.* 14, 1218—1231.

Sanderson, M.J., Doyle, J.A., 2001. Sources of error and confidence intervals in estimating the age of angiosperms from *rbcL* and 18S rDNA data. *Am. J. Bot.* 88, 1499—1516.

Sanderson, M.J., 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol. Biol. Evol.* 19, 101—109.

Sanderson, M.J., 2004. r8S, program and documentation. Available at <http://loco.biosci.arizona.edu/r8s/>.

Sanmartín, I. 2005. A paleogeographic history of the Southern Hemisphere. Available from: <http://www.ebc.uu.se/systzoo/staff/sanmartin.html>

Sanmartín, I. 2007. Event-based biogeography: Integrating patterns, processes, and time. Pages 135-159 *in* *Biogeography in a Changing World* (M. C. Ebach, and R. Tangney, eds.). Crc Press-Taylor & Francis Group, Boca Raton.

Sanmartín, I. 2009. Breaking the chains of parsimony: parametric approaches to historical biogeography. In: *Biogeography: An ecological and evolutionary approach* (eds. B. C. Cox & P.D Moore), 7th Edition, Blackwell Publishing, in press.

Sanmartín, I., Enghoff, H., Ronquist, F., 2001. Patterns of animal dispersal, vicariance and diversification in the Holarctic. *Biol. J. Linn. Soc.* 73, 345—390.

Sanmartín, I., Ronquist, F., 2004. Southern hemisphere biogeography inferred by event-based models: plant versus animal patterns. *Syst. Biol.* 53, 1—28.

Sanmartín, I., Wanntorp, L., Winkworth, R., 2007. West Wind Drift Revisited: Testing for directional dispersal in the Southern Hemisphere using event-based tree fitting. *J. Biogeogr.* 34, 398—416.

Sanmartín, I., van der Mark, P., Ronquist, F., 2008. Inferring dispersal: a Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands. *J. Biogeogr.* 35, 428—449.

Santos, J.C., Coloma, L.A., Summers, K., Caldwell, K.P., Ree, R., Cannatella, D.C., 2009.

Amazonian amphibian diversity is primarily derived from Late Miocene Andean lineages. *Plos Biology* 7, 448—461.

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

Smith, S.A., 2006. AReA: area reconstruction analysis. Version 2.1.

<http://blackrim.org/programs/area.html>.

Song, Z., 1988. Late Cenozoic palyno-flora from Zhaotong, Yunnan. *Mem. of Nanjing Inst. of Geol. & Palaeont., Acad. Sinica* 24: 1–52 (in Chinese; English summary).

Stamatakis, A., 2006. RAxML-VI-HPC: Maximum Likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688-2690.

Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML web-servers. *Syst. Biol.* 57, 758-771

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

Takhtajan, A., 1978. The floristic regions of the World. Academy of Sciences of the U.S.S.R.. Leningrad. (In Russian).

Takhtajan, A., 1986. Floristic regions of the World. University of California Press, Berkeley.

Taylor, T.N., Taylor, E.L., 1993. The biology and evolution of fossil plants. Academic Press Inc, New Jersey.

Teodoris, V., 2003. Tertiary flora and vegetation of the locality Záhohří near Žatec (Most Basin, Czech Republic). *Bulletin of Geosciences* 78, 261—276.

Thorne, J.L., Kishino, H., 2002. Divergence time and evolutionary rate estimation with multilocus data. *Syst. Biol.* 51, 689–702.

- Tiffney, B.H., 1985a. Perspectives on the origin of the floristic similarity between eastern Asia and eastern North America. *J. Arnold Arbor.* 66, 73—94.
- Tiffney, B.H., 1985b. The Eocene North Atlantic land bridge: its importance in Tertiary and modern phytogeography the Northern Hemisphere. *J. Arnold Arbor.* 66, 243—273.
- Yang, Z.H., 1993. Maximum-likelihood-estimation of phylogeny from DNA-sequences when substitution rates differ over sites. *Mol. Biol. Evol.* 10, 1396—1401.
- Xiang, Q.Y.J., Thomas, D.T., 2008. Tracking character evolution and biogeographic history through time in Cornaceae-Does choice of methods matter. *JSE* 46, 349—374.
- Zhi-Chen, S., Wei-Ming, W., Huang, F., 2004. Fossil pollen records of extant angiosperms in China. *Bot. Rev.* 70, 425—458.

Figures legends

Figure 1. Biogeographical regions. Abbreviations: (A). Eurasia: from Western Europe to Indochina; (B) Africa; (C) Madagascar, including the Comoro Islands and the Mascarene Islands; (D) South East (SE) Asia: including India, the Malaysian Peninsula, Philippines, Sumatra, Borneo, and the Inner Banda Arc, as well as the Pacific Islands (e.g., Hawaii); (E) Australia: including New Guinea, New Caledonia and New Zealand; (F) North America, and (G) South America including Central America and the West Indies.

Figure 2. Biogeographical model subdivided into four time slices defined on paleogeographic connections. Zeros in dispersal matrices were substitute by 0.01 when performing Lagrange Str analyses. Abbreviations: Bold and dashed arrows represent respectively dispersals probabilities of 1 and 0.5 and dashed circles symbolize subequatorial current and the west wind drift (WWD). See figure 1 for details on circumscription of areas and text for explanations.

Figure 3. Divergence time uncertainties mapped on the Med dated phylogeny of the Sapindaceae. Left: NPRS estimation; right: PL estimation. Lineages-through-time plots are shown. Abbreviations: circles: calibration points (see material and methods for more details); grey lines: limits of time slices.

Figure 4. Scatter plots comparing PL and NPRS divergence time estimations.

Figure 5. A. Pairwise comparison of mean maximum probabilities of assignment to area(s). B. Comparison of mean probabilities to disperse between Node-Node and Node-Tip branches for each biogeographic method. C. Approximation of the number of dispersals per branch according to branch length (see text for more details).

Figure 6. Soft and hard incongruities between methods through time. Abbreviations: soft incongruities are sorted into six distance classes; X: hard incongruent node; bold lines: limits of time slices; star: nodes shared between to time slices. Node numbers are depicted in Fig. S1.

Figure 7. Comparison of biogeographic scenarios obtained with Bayes-DIVA and Lagrange M1 inferences represented on the Med PL dated phylogenetic tree. For Lagrange analysis, pie charts represent ancestral area(s) based on the median PL estimation (see text for more details). Dispersal events are represented on paleogeographic maps according to time slices. Thicknesses of arrows are proportional to the number of dispersals (see Table 5 for more details). Abbreviations: junk (in black): sum of ancestral area probabilities <0.1 ; star: hard incongruent nodes between methods; @: hard incongruent node within methods; grey lines: ranges between time slices and TS: time slice.

Figure 8. Comparison of biogeographic scenarios obtained with Bayes-DIVA and Lagrange Str inferences represented on the Med PL dated phylogenetic tree. For Lagrange analysis, pie charts represent ancestral area(s) based on the median PL estimation (see text for more details). Dispersal events are represented on paleogeographic maps according to time slices. Thicknesses of arrows are proportional to the number of dispersals (see Table 5 for more details). Abbreviations: junk (in black): sum of ancestral area probabilities <0.1 ; star: hard incongruent nodes between methods; @: hard incongruent node within methods; grey lines: ranges between time slices and TS: time slice.

Figure 9. Comparison of biogeographic scenarios obtained with Lagrange M1 and Lagrange Str inferences represented on the Med PL dated phylogenetic tree. Pie charts represent ancestral area(s) based on the median PL estimation (see text for more details). Dispersal events are represented on paleogeographic maps according to time slices. Thicknesses of arrows are proportional to the number of dispersals (see Table 5 for more details).

Chapter 2

Abbreviations: junk (in black): sum of ancestral area probabilities < 0.1 ; star: hard incongruent nodes between methods; @: hard incongruent node within methods; grey lines: ranges between time slices and TS: time slice.

Tables

Table 1. Transition cost matrix used by dispersal-vicariance analysis (DIVA) to estimate the cost of change in geographic range between ancestor and descendants. Note that DIVA used a three-dimensional cost matrix that specifies the cost of moving from the ancestor to the left descendant and to the right descendant. Inheritance of geographic ranges follows two different modes of speciation: for single-area ancestors, the two descendants inherit the entire geographic range (sympatric speciation); for widespread ancestral ranges, the range is divided into two non-overlapping subsets between the two descendants (vicariance). Widespread ranges are always divided by vicariance at speciation events, that is, inheritance of widespread ancestral ranges is not allowed by the model. Extinctions are not inferred unless specific geographic constraints are imposed on the model based on a geological scenario. Transitions between two widespread ranges (e.g., from ab to bc) must be explained by vicariance in the ancestral node (Va|b) followed by range expansion (Dbc); Abbreviations: D: dispersal or range expansion; S: sympatric speciation, V: vicariance; -: not allowed. See text for more details on the method and the coding of geographic ranges.

		To Area(s)																													
		a	b	c	d	e	f	g	ab	ac	ad	ae	af	ag	bc	bd	be	bf	bg	cd	ce	cf	cg	de	df	dg	ef	eg	fg		
From Area(s)	a	S	-	-	-	-	-	-	Dab	Dac	Dad	Dae	Daf	Dag	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	b	-	S	-	-	-	-	-	Dba	-	-	-	-	-	Dbc	Dbd	Dbe	Dbf	Dbg	-	-	-	-	-	-	-	-	-	-		
	c	-	-	S	-	-	-	-	Dca	-	-	-	-	-	Dcb	-	-	-	-	-	Dcd	Dce	Dcf	Dcg	-	-	-	-	-		
	d	-	-	-	S	-	-	-	-	Dda	-	-	-	-	Ddb	-	-	-	-	-	Ddc	-	-	-	Dde	Ddf	Ddg	-	-		
	e	-	-	-	-	S	-	-	-	-	Dea	-	-	-	-	-	Deb	-	-	-	-	Dec	-	-	Ded	-	-	Def	Deg		
	f	-	-	-	-	-	S	-	-	-	-	Dfa	-	-	-	-	-	Dfb	-	-	-	-	Dfc	-	-	Dfd	-	-	Dfe	Dfg	
	g	-	-	-	-	-	-	S	-	-	-	-	Dga	-	-	-	-	-	Dgb	-	-	-	Dgc	-	-	Dgd	-	-	Dge	Dgf	
	ab	Va b	Va b	-	-	-	-	-	Va b+Dab	Va b+Dac	Va b+Dad	Va b+Dae	Va b+Daf	Va b+Dag	Va b+Dbc	Va b+Dbd	Va b+Dbe	Va b+Dbf	Va b+Dbg	-	-	-	-	-	-	-	-	-	-	-	
	ac	Va c	-	Va c	-	-	-	-	Va c+Dab	Va c+Dac	Va c+Dad	Va c+Dae	Va c+Daf	Va c+Dag	Va c+Dcb	-	-	-	-	-	Va c+Dcd	Va c+Dce	Va c+Dcf	Va c+Dcg	-	-	-	-	-		
	ad	Va d	-	-	Va d	-	-	-	Va d+Dab	Va d+Dac	Va d+Dad	Va d+Dae	Va d+Daf	Va d+Dag	-	Va d+Ddb	-	-	-	-	Va d+Ddc	-	-	-	Va d+Dde	Va d+Ddf	Va d+Ddg	-	-		
	ae	Va e	-	-	-	Va e	-	-	Va e+Dab	Va e+Dac	Va e+Dad	Va e+Dae	Va e+Daf	Va e+Dag	-	-	Va e+Deb	-	-	-	-	Va e+Dec	-	-	-	Va e+Ded	-	-	Va e+Def	Va e+Deg	
	af	Va f	-	-	-	-	Va f	-	Va f+Dab	Va f+Dac	Va f+Dad	Va f+Dae	Va f+Daf	Va f+Dag	-	-	-	Va f+Dfb	-	-	-	Va f+Dfc	-	-	-	Va f+Dfd	-	-	Va f+Dfe	Va f+Dfg	
	ag	Va g	-	-	-	-	-	Va g	Va g+Dab	Va g+Dac	Va g+Dad	Va g+Dae	Va g+Daf	Va g+Dag	-	-	-	-	Va g+Dgb	-	-	-	Va g+Dgc	-	-	-	Va g+Dgd	-	-	Va g+Dge	Va g+Dgf
	bc	-	Vb c	Vb c	-	-	-	-	Vb c+Dba	Vb c+Dca	-	-	-	-	Vb c+Dbc	Vb c+Dbd	Vb c+Dbe	Vb c+Dbf	Vb c+Dbg	Vb c+Dcd	Vb c+Dce	Vb c+Dcf	Vb c+Dcg	-	-	-	-	-	-	-	
	bd	-	Vb d	-	Vb d	-	-	-	Vb d+Dba	-	Vb d+Dda	-	-	-	Vb d+Ddb	Vb d+Ddb	Vb d+Dbe	Vb d+Dbf	Vb d+Dbg	Vb d+Ddc	-	-	-	-	Vb d+Dde	Vb d+Ddf	Vb d+Ddg	-	-	-	
	be	-	Vb e	-	-	Vb e	-	-	Vb e+Dba	-	-	Vb e+Dea	-	-	Vb e+Ddb	Vb e+Ddb	Vb e+Dbe	Vb e+Dbf	Vb e+Dbg	-	Vb e+Dec	-	-	-	-	Vb e+Ded	-	-	Vb e+Def	Vb e+Deg	
	bf	-	Vb f	-	-	-	Vb f	-	Vb f+Dba	-	-	-	Vb f+Dfa	-	Vb f+Ddb	Vb f+Ddb	Vb f+Dbe	Vb f+Dbf	Vb f+Dbg	-	-	Vb f+Dfc	-	-	-	Vb f+Dfd	-	-	Vb f+Dfe	Vb f+Dfg	
	bg	-	Vb g	-	-	-	-	Vb g	Vb g+Dba	-	-	-	Vb g+Dga	-	Vb g+Dbc	Vb g+Dbd	Vb g+Dbe	Vb g+Dbf	Vb g+Dbg	-	-	-	Vb g+Dgc	-	-	-	Vb g+Dgd	-	-	Vb g+Dge	Vb g+Dgf
	cd	-	-	Vc d	Vc d	-	-	-	Vc d+Dca	Vc d+Dda	-	-	-	-	Vc d+Dcb	Vc d+Ddb	-	-	-	Vc d+Dcd	Vc d+Dce	Vc d+Dcf	Vc d+Dcg	Vc d+Dde	Vc d+Ddf	Vc d+Ddg	-	-	-	-	
	ce	-	-	Vc e	-	Vc e	-	-	Vc e+Dca	-	Vc e+Dea	-	-	-	Vc e+Dcb	-	Vc e+Deb	-	-	-	Vc e+Dcd	Vc e+Dce	Vc e+Dcf	Vc e+Dcg	Vc e+Ded	-	-	Vc e+Def	Vc e+Deg		
	cf	-	-	Vc f	-	-	Vc f	-	Vc f+Dca	-	-	Vc f+Dfa	-	-	Vc f+Dcb	-	-	Vc f+Dfb	-	-	Vc f+Dcd	Vc f+Dce	Vc f+Dcf	Vc f+Dcg	-	Vc f+Dfd	-	-	Vc f+Dfe	Vc f+Dfg	
	cg	-	-	Vc g	-	-	-	Vc g	Vc g+Dca	-	-	-	Vc g+Dga	-	Vc g+Dcb	-	-	-	Vc g+Dgb	Vc g+Dcd	Vc g+Dce	Vc g+Dcf	Vc g+Dcg	-	-	Vc g+Dgd	-	-	Vc g+Dge	Vc g+Dgf	
	de	-	-	-	Vd e	Vd e	-	-	-	Vd e+Dda	Vd e+Dea	-	-	-	Vd e+Ddb	Vd e+Ddb	Vd e+Deb	-	-	Vd e+Ddc	Vd e+Ddc	-	-	Vd e+Dde	Vd e+Ddf	Vd e+Ddg	Vd e+Ddf	Vd e+Dde	Vd e+Deg	-	
	df	-	-	-	Vd f	-	Vd f	-	-	Vd f+Dda	-	Vd f+Dfa	-	-	Vd f+Ddb	-	-	Vd f+Dfb	-	-	-	-	Vd f+Dfc	-	-	-	Vd f+Dfd	Vd f+Dde	Vd f+Dde	Vd f+Dde	-
	dg	-	-	-	Vd g	-	-	Vd g	-	Vd g+Dda	-	-	Vd g+Dga	-	Vd g+Ddb	-	-	Vd g+Dgb	-	-	-	-	Vd g+Dgc	Vd g+Dde	Vd g+Ddf	Vd g+Ddg	-	-	Vd g+Dge	Vd g+Dgf	
	ef	-	-	-	-	Ve f	Ve f	-	-	-	Ve f+Dea	Ve f+Dfa	-	-	-	Ve f+Deb	Ve f+Dfb	-	-	-	Ve f+Dec	Ve f+Dfc	-	-	-	-	-	Ve f+Def	Ve f+Deg	Ve f+Dfg	
	eg	-	-	-	-	Ve g	-	Ve g	-	-	Ve g+Dea	-	Ve g+Dga	-	-	Ve g+Deb	-	-	Ve g+Dgb	-	Ve g+Dec	-	Ve g+Dgc	Ve g+Ded	-	Ve g+Dgd	-	-	Ve g+Deg	Ve g+Dgf	
fg	-	-	-	-	-	Vf g	Vf g	-	-	-	Vf g+Dfa	Vf g+Dga	-	-	-	Vf g+Dfb	Vf g+Dgb	-	-	-	Vf g+Dfc	Vf g+Dgc	-	-	Vf g+Dfd	Vf g+Dgd	Vf g+Dde	Vf g+Dfg	-		

Table 2. Transition probability matrix (Q matrix) depicting the instantaneous transition rates between geographic ranges used in the dispersal-extinction-cladogenesis (DEC) analyses (Lagrange M1 and Lagrange Str) to infer by maximum likelihood ancestral ranges and biogeographic events. Note that only rates separated by a single dispersal or extinction event are allowed in the matrix; all other transitions have an instantaneous rate of zero. For transitions involving dispersal, the rate is the sum of rates from areas in the starting range r to the area of expansion r' (Ree & Sanmartín, 2009). Abbreviations: D: dispersal or range expansion between areas; E: local extinction within an area. See text for more details on the method and the area definitions.

		To Area(s)																																				
		a	b	c	d	e	f	g	ab	ac	ad	ae	af	ag	bc	bd	be	bf	bg	cd	ce	cf	cg	de	df	dg	ef	eg	fg	bde	bcd	ade	abcd	abcde	abcdeg			
From Area(s)	a	-	0	0	0	0	0	0	Dab	Dac	Dad	Dae	Daf	Dag	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
	b	0	-	0	0	0	0	0	Dba	0	0	0	0	0	Dbc	Dbd	Dbe	Dbf	Dbg	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
	c	0	0	-	0	0	0	0	Dca	0	0	0	0	0	Dcb	0	0	0	0	Dcd	Dce	Dcf	Dcg	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	d	0	0	0	-	0	0	0	0	Dda	0	0	0	0	0	Ddb	0	0	0	Ddc	0	0	0	Dde	Ddf	Ddg	0	0	0	0	0	0	0	0	0	0	0	
	e	0	0	0	0	-	0	0	0	0	Dea	0	0	0	0	Deb	0	0	0	Dec	0	0	Ded	0	0	Def	Deg	0	0	0	0	0	0	0	0	0	0	
	f	0	0	0	0	0	-	0	0	0	0	Dfa	0	0	0	Dfb	0	0	0	Dfc	0	0	Dfd	0	Dfe	0	Dfg	0	0	0	0	0	0	0	0	0	0	
	g	0	0	0	0	0	0	-	0	0	0	0	0	Dga	0	0	0	0	Dgb	0	0	Dgc	0	Dgd	0	Dge	Dgf	0	0	0	0	0	0	0	0	0	0	
	ab	Eb	Ea	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
	ac	Ec	0	Ea	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	ad	Ed	0	0	Ea	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Dae+Dde	0	0	
	ae	Ee	0	0	0	Ea	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Dad+Ded	0	0	0	
	af	Ef	0	0	0	0	Ea	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	ag	Eg	0	0	0	0	0	Ea	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	bc	0	Ec	Eb	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Dbd+Dcd	0	0	0	
	bd	0	Ed	0	Eb	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Dbe+Dde	Dbc+Ddc	0	0	0	
	be	0	Ee	0	0	Eb	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Dbd+Ded	0	0	0	0	
	bf	0	Ef	0	0	0	Eb	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	bg	0	Eg	0	0	0	0	Eb	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	cd	0	0	Ed	Ec	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	Dcb+Ddb	0	0	0	0	
	ce	0	0	Ee	0	Ec	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	cf	0	0	Ef	0	0	Ec	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	cg	0	0	Eg	0	0	0	Ec	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	de	0	0	0	Ee	Ed	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	Ddb+Deb	0	Dda+Dea	0	0	0
	df	0	0	0	Ef	0	Ed	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0
	dg	0	0	0	Eg	0	0	Ed	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0
	ef	0	0	0	0	Ef	Ee	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0
	eg	0	0	0	0	Eg	0	Ee	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0
	fg	0	0	0	0	0	0	Eg	Ef	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0
bde	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Ee	Ed	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0		
bcd	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Ed	Ec	0	0	0	Eb	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	0	0	
ade	0	0	0	0	0	0	0	0	0	Ee	Ed	0	0	0	0	0	0	0	0	0	0	0	0	Ea	0	0	0	0	0	0	0	0	0	0	-	0	0	
abcd	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Ea	0	-	0		
abcde	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Ee	-	Dae+Dbe+Dce+Dde	0
abcdeg	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	Dag+Dbg+Dcg+Deg	0
																																				Eg	-	0

Table 3. Number of soft and hard incongruent nodes within and between methods.

Abbreviations: A: Bayes-DIVA/Lagrange M1; B: Bayes-DIVA/Lagrange Str; C: Lagrange M1/Lagrange Str; D: Lagrange M1 Max/Lagrange M1 Min ; E: Lagrange Str Max/Lagrange Str Min. See text for more details.

Class of distance	Incongruence type	A	B	C	D	E
0-0.1	Soft	50	58	74	93	83
	Hard	0	0	0	0	0
0.1-0.2	Soft	15	12	29	23	21
	Hard	1	3	9	11	14
0.2-0.3	Soft	10	6	3	15	9
	Hard	3	3	5	3	8
0.3-0.4	Soft	9	5	5	1	2
	Hard	2	1	4	1	3
0.4-0.5	Soft	9	13	5	0	1
	Hard	3	4	4	0	2
>0.5	Soft	14	9	5	0	0
	Hard	31	33	4	0	4
N of Soft incongruent nodes		107	103	121	132	116
N of Hard incongruent nodes		40	44	26	15	31

Table 4. Hard incongruent nodes within Lagrange Str shared between two time slices.

Maximum probabilities of ancestral area are indicated for each analysis. See Fig. 6 for more details and S1 for the location of node.

Node	Med Age		Time slices	Lagrange Str		
	(Mya)	Δ PL (Mya)		Max	Med	Min
294	83	20.4	1-2	a (0.54)	af (0.49)	af (0.51)
154	80.48	12.74	1-2	a (0.72)	a (0.72)	ad (0.62)
160	73.59	15.34	1-2	a (0.54)	a (0.42)	ad (0.38)
175	60.02	8.75	2-3	af (0.36)	d (0.65)	de (0.56)
202	58.48	8.48	2-3	fg (0.25)	de (0.67)	e (0.44)
265	57.45	8.4	2-3	g (0.44)	eg (0.59)	eg (0.68)
155	30.75	9.6	3-4	cd (0.22)	d (0.20)	d (0.33)
178	29.5	6.73	3-4	bcd (0.47)	b (0.32)	bd (0.27)
283	25.71	11.43	3-4	g (0.68)	bg (0.57)	bg (0.63)

Table 5. Comparison of dispersal events through time between Bayes-DIVA, Lagrange M1 and Lagrange Str. See figures 7-9 and text for more details.

Time slice	From-To	Dispersals		
		Bayes-DIVA	Lagrange M1	Lagrange Str
1	A-D	1		
	A-F			1
2	A-D		2	1
	D-E	1		
3	A-D	1		3
	A-B	1		
	A-F		1	
	A-G	1		
	B-C	1		
	B-E	1		
	D-A	1		
	D-B	2	1	1
	D-E		1	2
	D-G	1		
	E-D	1		
	E-G			1
	G-A	1		
	G-B	1		
G-E	1			
4	A-B			
	A-D	4	2	2
	A-F	4		1
	A-G			1
	B-C	5		1

B-D	2	1	2
B-E		1	
B-G		1	1
C-B	3	3	5
D-A	12	9	11
D-B			1
D-C			1
D-E	4	3	2
E-D	18	13	15
E-G	2	1	2
G-B		1	3
G-C	1	1	

Table 6. Extinctions per area through time. Numbers between brackets referred to respectively observed extinctions and extinctions predicted by the model. See material and methods for more details on areas and time slice definitions. Abbreviation: E: extinction.

Area	Lagrange M1			Lagrange Str		
	Time slice 3	Time slice 4	ΣE	Time slice 3	Time slice 4	ΣE
a	1 (1/0)	2 (2/0)	3	2 (1/2)	3 (2/1)	5
b	1 (1/0)	6 (5/1)	7	0	11 (9/2)	11
c	1 (1/0)	6 (6/0)	7	0	2 (2/0)	2
d	2 (2/0)	7 (6/1)	9	4 (4/1)	6 (6/0)	10
e	0 (0/0)	4 (2/2)	4	1 (1/0)	4 (3/1)	5
f	0 (0/0)	1 (1/0)	1	1 (1/0)	0	1
g	0 (0/0)	2 (1/1)	2	0	3 (2/1)	3
Total	5	28	33	8	29	37

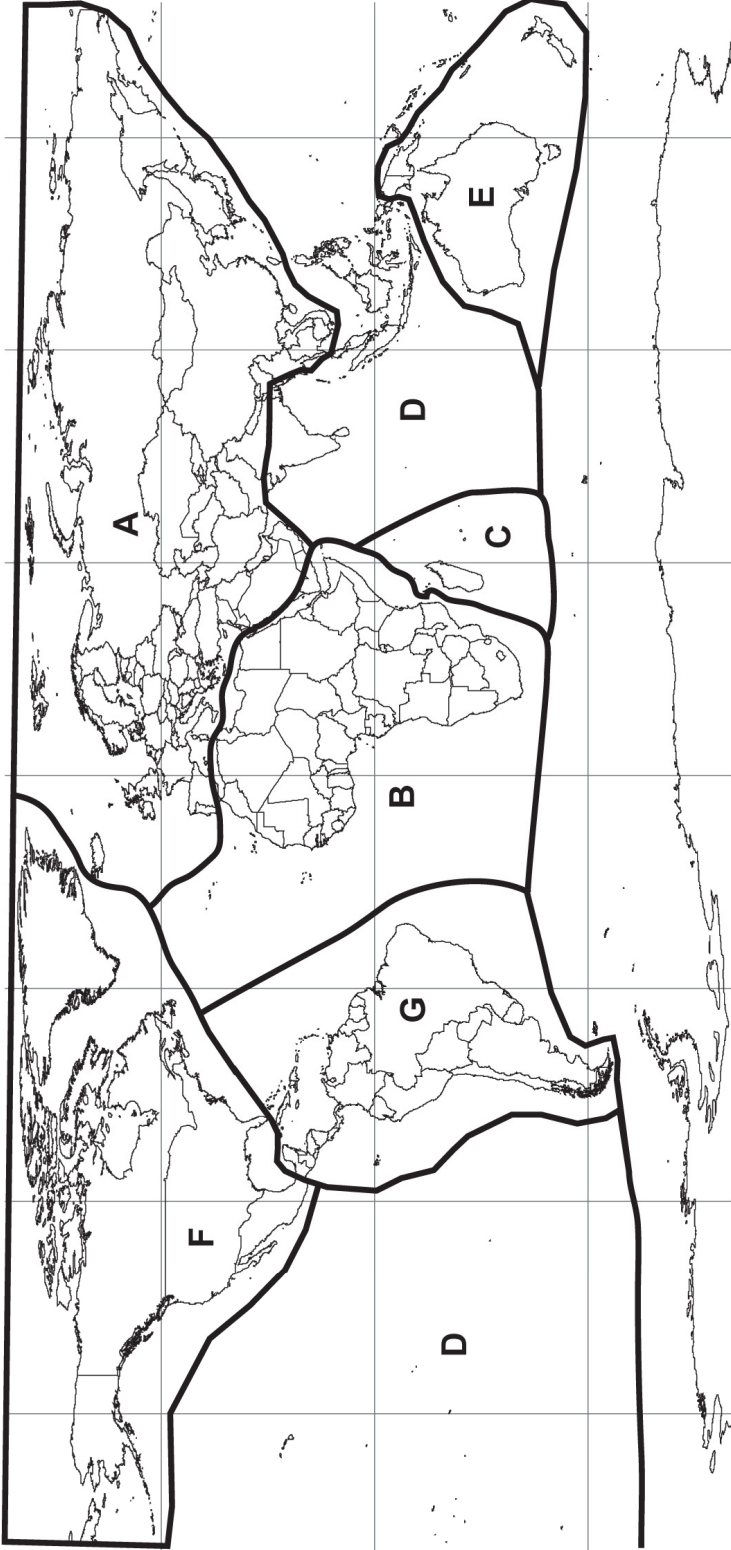


Figure 1

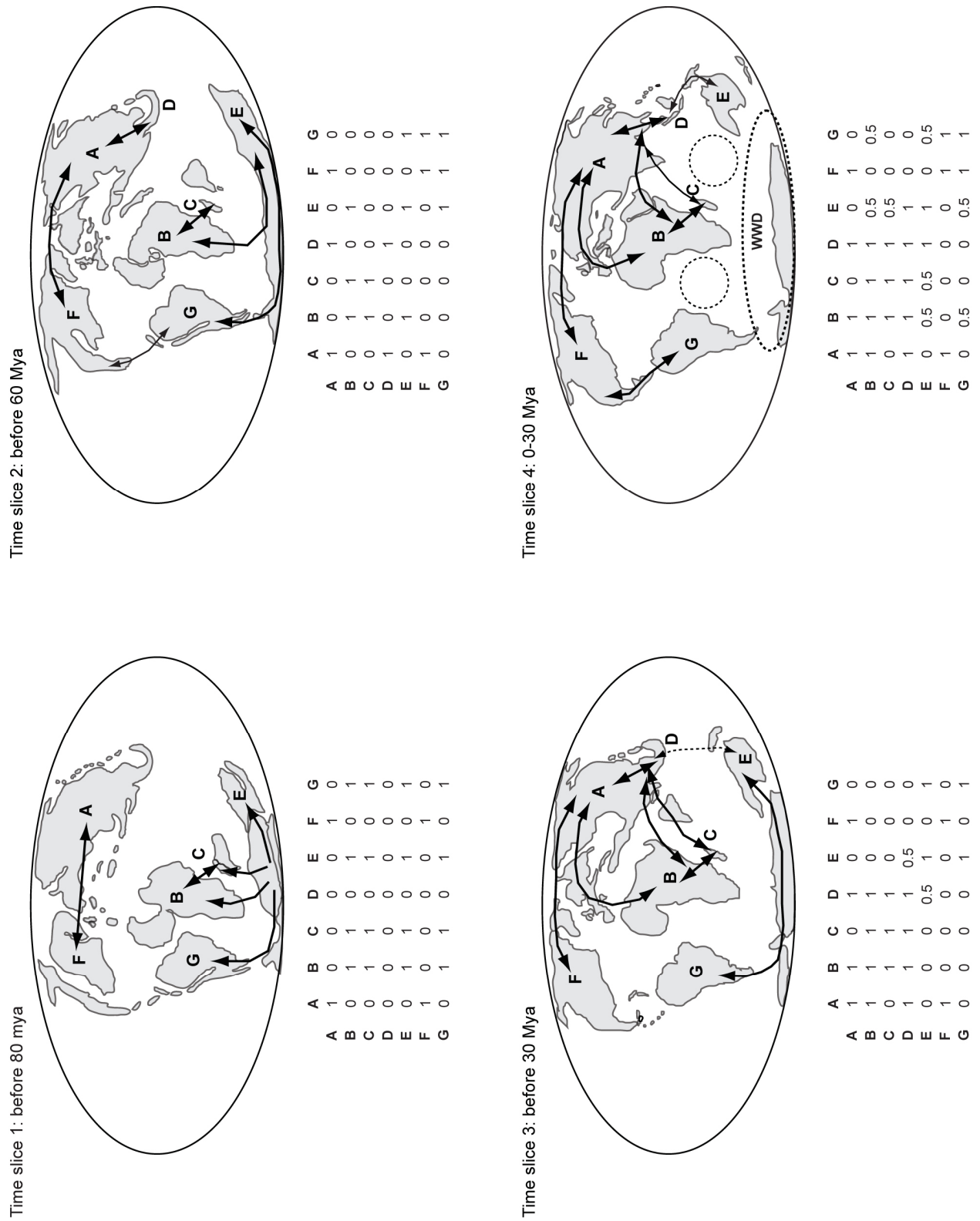


Figure 2

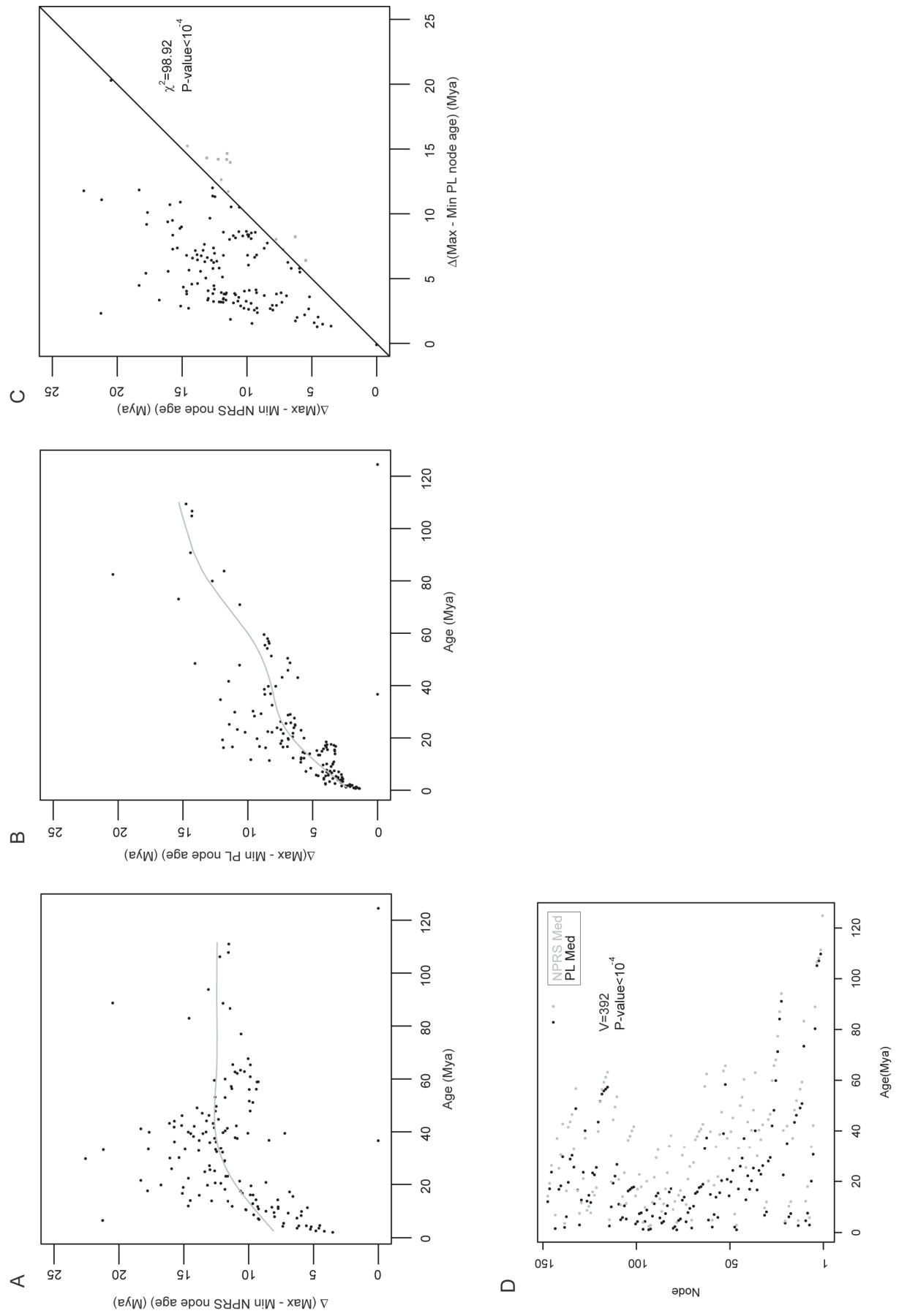


Figure 4

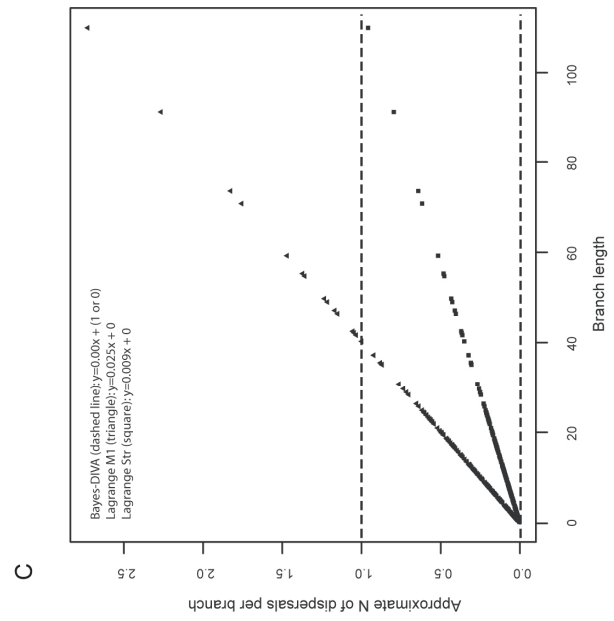
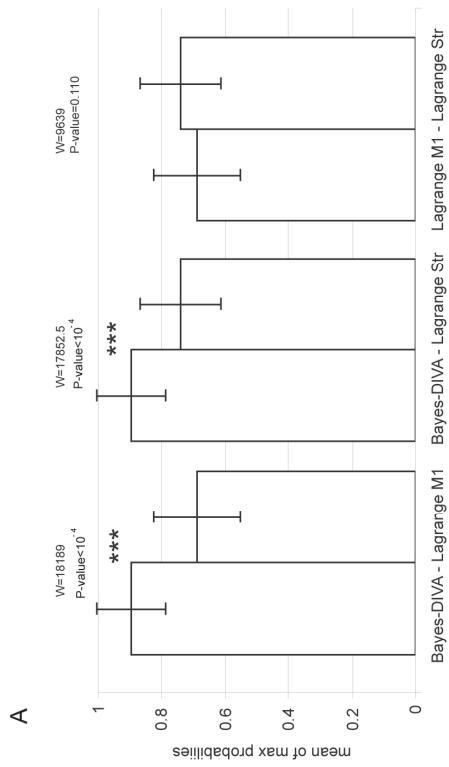
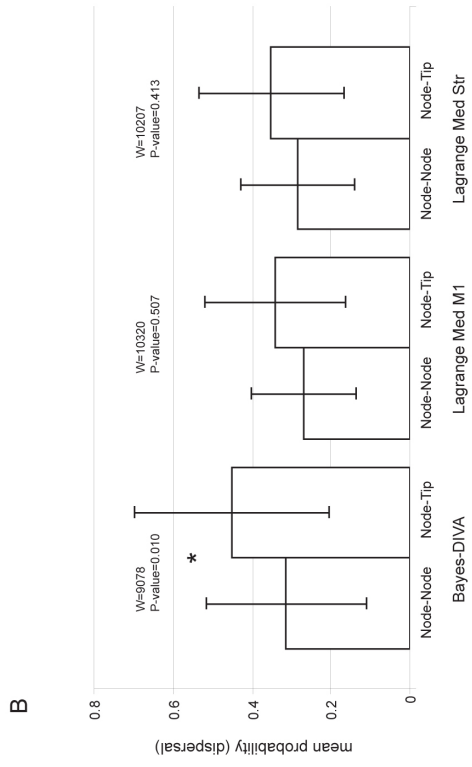


Figure 5

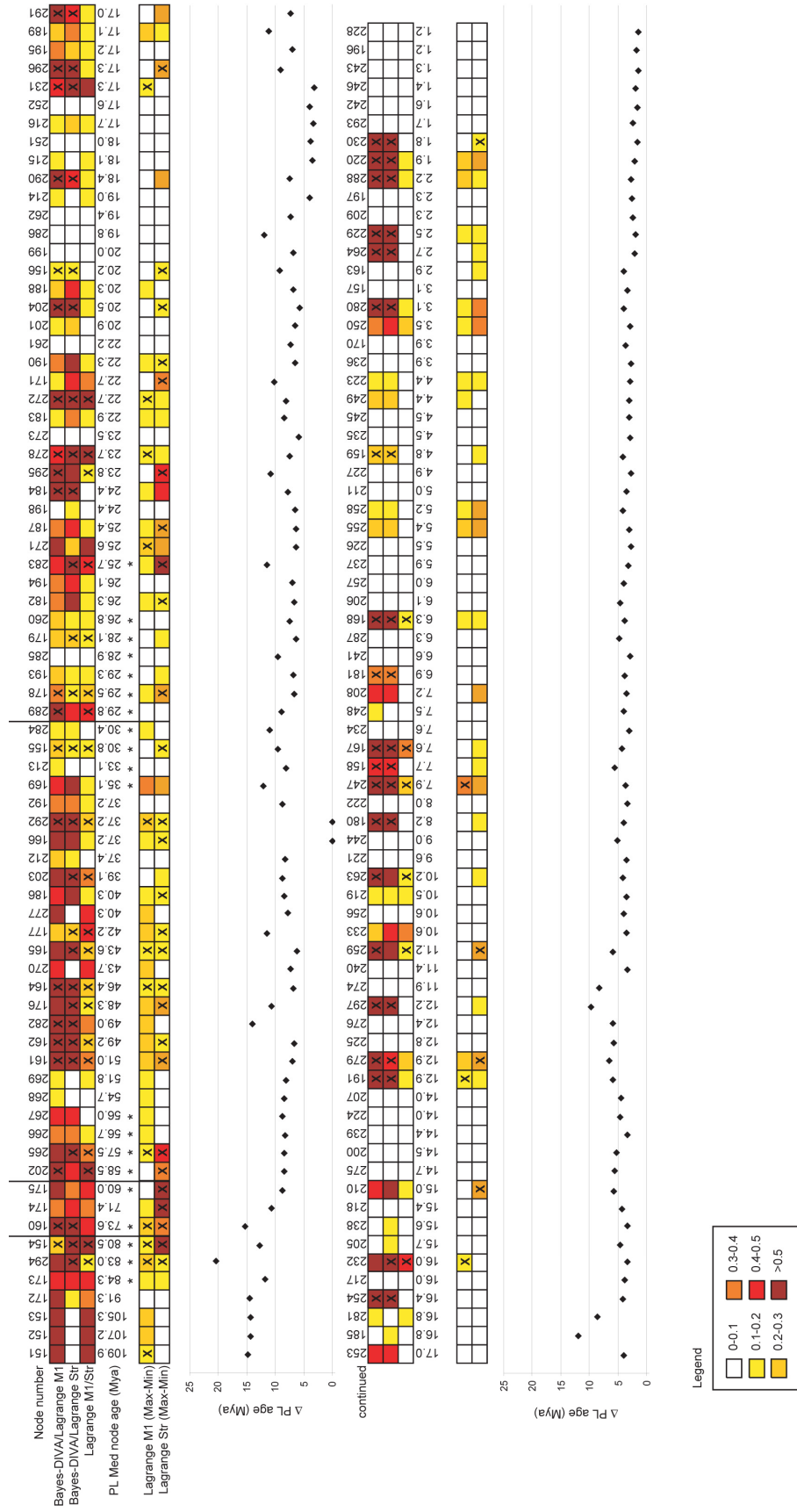


Figure 6

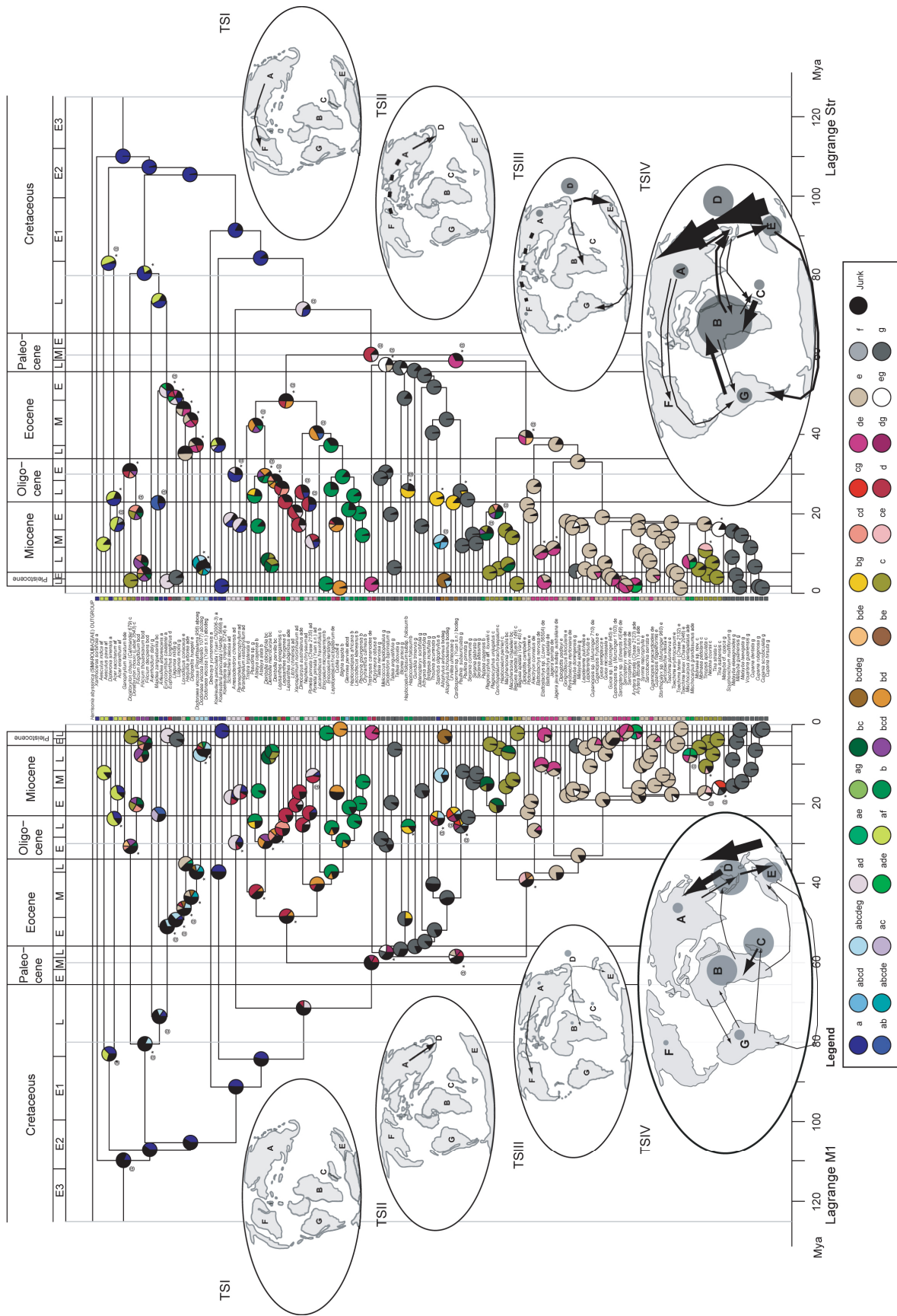


Figure 9

Supplementary material

Table 1. Contingency table showing number and type of biogeographic events (change in geographic range from ancestor to descendant) inferred by dispersal-vicariance analysis (DIVA) integrated over a Bayesian sample of trees obtained from the posterior distribution of the phylogeny of Sapindaceae (i.e., a Bayes-DIVA analysis). Biogeographic events are sorted by time slice (see Fig. 2). Notice that DIVA does not take into account the timing of divergence events between lineages (branch lengths). Biogeographic events are expressed as frequencies of range expansion (Dab), sometimes combined with vicariance events (V a|b) according to the cost transition matrix represented in Table 1. Fonts indicate different types of transition events between geographic ranges: *italic*: vicariance events; roman: range expansion; bold: vicariance and range expansion or two range expansions; red + bold: not considered. See material and methods for more details on areas and time slice definitions.

		To Area(s)																									
Time slice		a	ab	abcd	abcdeg	ad	ade	af	ag	b	bc	bcd	abcdeg	bd	bde	bg	c	cd	ce	cg	d	de	dg	e	eg	g	
From Area(s)	a	2				1 (Dad)																					
	3								1 (Dag)																		
	4		1 (Dab)			4 (Dad)		4 (Daf)					2 (Dab+Dac+Dad+Dae+Daq)														
	ab	3								1 (Va b)																	
	4		1 (Va b)								1 (Va b+Dbc)																
	ad	3																			1 (Va d)						
	4		1 (Va d)																		1 (Va d)						
	ag	3																							1 (Va g+Dge)		
	4		2 (Va g)			1 (Va g+Dad)																				2 (Va g)	
	b	4		1 (Dba+Dbc+Dbd)								4 (Dbc)	3 (Dbc+Dbd)		1 (Dbd)	1 (Dbd+Dbe)											2 (Va g)
	bc	4									2 (Vb c)										2 (Vb c)						
	bd	3									1 (Vb d)											1 (Vb d)					
	4										1 (Vb d)											1 (Vb d)	1 (Vb d+Dde)				
	bg	4									2 (Vb g)																2 (Vb g)
	c	4										3 (Dcb)															
	cd	4																			1 (Vc d)		1 (Vc d)				
	ce	3																								1 (Vc e)	
	4																										
	cg	4																			1 (Vc e)						
	d	3																								1 (Vc g)	
4						5 (Dda)	2 (Dda+Dde)							1 (Ddb)	1 (Ddb)								1 (Dde)			1 (Vc g)	
de	3																									1 (Vd g)	
4																										1 (Vd g)	
dg	3																									1 (Vd g)	
4																										1 (Vd g)	
e	4																					1 (Vd g)				1 (Vd g)	
eg	3																									2 (Deg)	
4																										2 (Deg)	
g	3																									2 (Ve g)	
4										1 (Dga)																2 (Ve g)	

Table 2. Contingency table showing number and type of transition events between geographic ranges (change in geographic range from ancestor to descendant) inferred by a non-stratified likelihood dispersal-extinction-cladogenesis (DEC) model (Lagrange M1) on the Sapindaceae chronogram obtained with penalized likelihood (PL, Fig. 3b). Transition events are sorted by time slice (see Fig. 2). Transitions are expressed as rates of local extinction (E_a) and/or range expansion (D_{ab}), according to the Q transition probability matrix represented in Table 2. Fonts indicate different types of transition events between geographic ranges: *Italic*: local extinction (range contraction); Roman: dispersal (range expansion); **Bold**: transitions between ranges that are separated by more than one dispersal or extinction event; according to the Q matrix, transitions involving dispersal (range expansion) the rate is the sum of rates from areas in the starting range to the target area, e.g., $D_{ab}+D_{bc} = D_{abc}$ (Ree & Sanmartín, 2009); **Bold ***: transitions that involve three or more events of range expansion and for which there is uncertainty in the direction of dispersals or the transition is not present in the Q matrix (e.g., $abc \rightarrow bc$); red: these transition events were not counted because it . See Material and Methods for more details on areas and time slice definitions.

		To Area(s)																			
		a	abcd	ad	ade	ae	af	b	bc	bcd	bd	bde	c	cd	cg	d	de	dg	e	g	
From Area(s)	a	2		1 (Dad)																	
	3					2 (Daf)															
	abc	4	1						1 (Ea)*												
	ad	3														1 (Ea)					
	4	1 (Ed)														1 (Ea)					
	ade	4			1 (Ee)																
	ae	4				2 (Dad+Ded)															
	af	4	1 (Ef)																		
	b	4		1 (Dbc+Dbd+Dcd+Dbc+Dca+Dda)						1 (Dbd)											1 (Dbg+Eb)
	bc	4							1 (Ec)							2 (Eb)					
	bcd	4							1 (Ed+Ec)							2 (Eb)					
	bd	3							1 (Ed)								1 (Eb)				
	4								1 (Ed)								1 (Eb)	1 (Dbe+Dde+Eb)			
	c	4								2 (Dcb)											
	cd	4														2 (Ed)	1 (Ec)				
	ce	3																		1 (Ec)	
	4									1 (Ee+Dcb)											
	cg	4														1 (Eg)					2 (Ec)
	d	3									1 (Ddb)										1 (Dde+Ed)
	4			4 (Dda)	2 (Dda+Dde)*						1										
de	4															1 (Ee)					
dg	3																			1 (Ed)	
e	4			3 (Ded+Dda)*												1	8 (Ded)			1 (Deg+Ee)	
g	4								1 (Dgb+Eg)						1 (Dgc)						

Table 3. Contingency table showing number and type of transition events between geographic ranges (change in geographic range from ancestor to descendant) inferred by a temporally stratified likelihood DEC model (Lagrange Str) on the PL chronogram of Sapindaceae (Fig. 3b). Transition events are sorted by time slice (see Fig. 2). Transitions are expressed as rates of local extinction (*Ea*) and/or range expansion (Dab), according to the Q transition probability matrix represented in Table 2. Fonts indicate different types of transition events between geographic ranges: *Italic*: local extinction (range contraction); Roman: dispersal (range expansion). **Bold**: transitions between ranges that are separated by more than one dispersal or extinction event; for transitions involving dispersal (range expansion), the rate is the sum of rates from areas in the starting range to the target area (e.g., Table 2: from ab to abc = Dab+Dbc (Ree & Sanmartín, 2009). **Bold ***: transitions that involve two or more events of range expansion and for which there is uncertainty in the direction of dispersals (e.g., a to abc); red: these transition events were not considered. See Material and Methods for more details on areas and time slice definitions.

		To Area(s)																							
		Time slice	a	ab	abc	abcd	abcde	abcdeg	ad	ade	af	b	bc	bcd	bcdeg	bd	bde	bg	c	cd	d	de	e	eg	g
From Area(s)	a	1									1 (Daf)														
	2								1 (Dad)																
	3								1 (Dad)													1 (Dad+Ea)			
	4								2 (Dad)		1 (Daf)														1 (Dag+Ea)
	ab	4	1 (Eb)										1 (Ea+Dbc)												
	abcdeg	4	1												1 (Ea)										
	ad	3	1 (Ed)																			1 (Ea)		1 (Dae+Dde+Ea)	
	4	1 (Ed)																				1 (Ea)			
	ade	4							1 (Ee)																
	af	3	1 (Ef)																						
	b	4			1											2 (Dbd)									1 (Dbg+Eb)
	bc	4											1 (Ec)								3 (Eb)				
	bd	3											1 (Ed)												
	4												2 (Ed)								1 (Dbc+Ddc+Eb)	3 (Eb)		1 (Pbd+Dde)	
	bg	4				1						2 (Eg)				1									2 (Eb)
	c	4											4 (Dcb)												
	cd	4												1 (Dcb+Ddb)						2 (Ed)		1 (Ec)			
	d	3														1 (Ddb)								1 (Dde)	
	4								4 (Dda)	2 (Dda+Dde)*							1				1 (Ddc)				
	de	3																						2 (Ed)	1 (Ed+Deg)
4				1					2 (Dda)*												1 (Ee)		1 (Ed)		
e	4									3 (Ded+Dda)*												11 (Ded)	1 (Deg)	1 (Deg+Ee)	
eg	3																							1 (Ee)	
4																						1 (Eg+Ded)		1 (Ee)	
g	4																							3 (Dgb)	

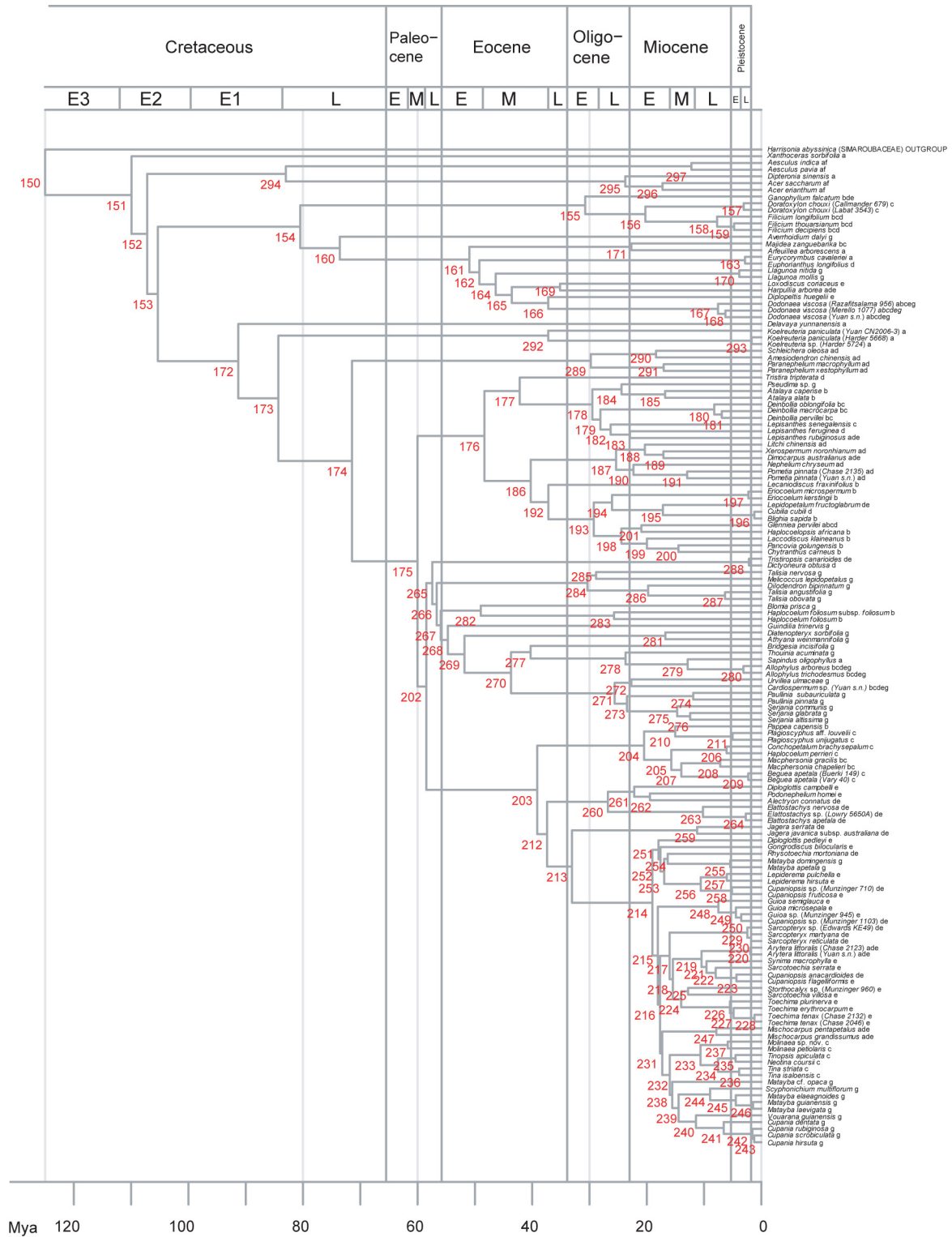


Figure S1. Med PL dated phylogeny of the Sapindaceae showing node numbers. See text for more details.

**Chapter three: Abiotic factors as driving force to promote
speciation in Sapindaceae**

Sven Buerki, Nadir Alvarez, Isabel Sanmartín and Félix Forest

Abiotic factors as driving force to promote speciation in Sapindaceae

Sven Buerki¹, Nadir Alvarez¹, Isabel Sanmartín² and Félix Forest³

¹ Institute of Biology, University of Neuchâtel, Rue Emile-Argand 11, 2009 Neuchâtel, Switzerland.

² Department of Biodiversity and Conservation, Real Jardín Botánico – CSIC, Plaza de Murillo 2, 28014 Madrid, Spain.

³ Molecular Systematics Section, Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, United Kingdom.

Abstract

Biotic and abiotic environmental factors are the drivers behind the evolutionary forces that shaped life on Earth. Their respective impact at large temporal and spatial scales, however, has been shown to be mostly antithetic: whereas abiotic factors (such as intense volcanic activities, meteoritic impacts and interactions between tectonic and orbital forces) were often associated with worldwide extinction processes, biotic factors are frequently related to species diversification (for instance, through diffuse coevolutionary processes). Nonetheless, it is more likely that abiotic factors might have catalyzed large-scale diversification processes of lineages. In this study, state-of-the-art biogeographic methods are applied to the worldwide soapberry family (Sapindaceae) to assess the impact of abiotic factors on diversification rates. Here, we show that abrupt climatic change in the Eocene-Oligocene boundary triggered diversification in Sapindaceae. This paradigm-breaking result is mainly due to the geological and climatic properties of South-East Asia that favoured multiple contacts between lineages and further speciation across Laurasian and Gondwanian continents.

Key-words

Angiosperms, Biogeography, Climate change, Diversification rates, Eocene-Oligocene boundary, Extinction, Key innovations.

Since the origin of flowering plants in the Jurassic or Early Cretaceous (Wikström et al., 2001; Magallon and Sanderson, 2005), an unprecedented diversification process has taken place (Frohlich and Chase 2007), which has resulted in the more than 250,000 species now recognized and distributed all over the world (*i.e.*, the most successful group of plants; Soltis et al., 2005). In the last decades, some researchers have hypothesised that such a high level of species diversity is related to a high rate of diversification during the Mesozoic Era, which might be associated with close interactions with biotic factors such as pollinators or phytophagous insects through the Cretaceous (Raven, 1977; Grimaldi 1999; Wilf et al. 2000; Labandeira, 2006; Hu et al., 2008; Crepet and Niklas, 2009). In contrast, large-scale abiotic factors are mostly seen as responsible of extinction processes, such as the Big five mass extinctions¹ (Raup and Sepkoski, 1982) – partly due to intense volcanic activities and meteoritic impacts (Jablonski, 2001) – or the Pleistocene climatic oscillations – caused by interactions between tectonics and orbital forces (Alley et al., 2003). Although not considered as one of the Big Five mass extinctions, the abrupt cooling near the Eocene-Oligocene boundary (E-O; ca. 33.7 Ma) has also had great impacts on biodiversity (Katz et al., 2008; Zhonghui et al., 2009). At this period, Earth's climate shifted from a relatively ice-free world to one with glacial conditions in Polar regions characterized by substantial ice sheets (Bowen, 2007). In a relatively short span of time, high-latitude (45 degrees to 70 degrees in both hemispheres) temperatures decreased from ca. 20°C to ca. 5°C (Zhonghui et al., 2009). Explanations for this cooling include changes in ocean circulation due to the opening of Southern Ocean gateways, a decrease in atmospheric CO₂, and a decrease in solar insolation (see references in Zhonghui et al., 2009). This important climate change seemed to be again related to a decrease in species diversity, as shown for instance in the decline of the floras in neotropics during the E-O boundary, with a possible spatial shift of tropical ecosystems (Jaramillo et al., 2006).

¹ The Big Five extinction masses: End Ordovician (O-S; 450-440 Ma), Late Devonian (Late D; ca. 364 Ma), End Permian (P-Tr; ca. 251 Ma), End Triassic (Tr-J; ca. 200 Ma) and End Cretaceous (K-T; 65 Ma).

Nonetheless, the biological consequences of this abrupt climate change have not yet been investigated worldwide, and other parts of the world might have undergone different scenarios. By performing state-of-the-art parametric biogeographic methods on the ancient, emblematic and worldwide soapberry family (Sapindaceae), we show that E-O climatic changes (1) promoted the diversification of subfamilial lineages and (2) coincide with the occurrence of several key-innovations in these plants. These results, which strongly contrast with the conventional paradigm of decrease in biodiversity through the E-O boundary, are mainly due to the geological and climatic properties of South-East Asia, which at that time favoured multiple contacts between lineages and further speciation across Laurasian and Gondwanian continents.

Material and Methods

Species names, voucher information, and GenBank accession numbers for all sequences are provided in Buerki et al. (2009) and in Buerki et al. (submitted). The DNA extraction, amplification and sequencing protocols of the nuclear ribosomal region and seven plastid regions are provided in Buerki et al. (2009). The nuclear region is the whole ITS region (ITS1, 5.8S and ITS2) and plastid markers include coding (*matK* and *rpoB*) and non-coding regions (the *trnL* intron and the intergenic spacers *trnD-trnT*, *trnK-matK*, *trnL-trnF* and *trnS-trnG*). Details regarding biogeographic area circumscription, paleogeographic model, phylogenetic analysis and divergence times estimations are presented in Buerki et al. (chapter 2). Numbers of genera per area (including the number of endemic and endemic monotypic genera) were based on Buerki et al. (2009). A Lagrange stratified (Str) biogeographic analysis (Ree and Smith, 2008; Buerki et al., chapter 2) was performed based on penalized likelihood divergence time estimates (Sanderson, 2002). Numbers of dispersal and extinction events were quantified automatically as follows: (i) for each node, the most likely ancestral area was extracted; (ii) branches and node ages were extracted using R (R Development Core Team 2009) and the APE package (Paradis et al., 2004); and finally (iii), each node was assigned the most likely ancestral areas and dispersals/extinction

contingency tables were built. Based on the contingency tables, dispersals (directions and types) and extinctions (observed and inferred by the model) were represented on paleogeographic maps according to the assumptions of the biogeographic model implemented in Lagrange (Ree and Smith, 2008; see Buerki et al., chapter 2 for more details).

Results and Discussion

Most genera in Sapindaceae are restricted in distribution; of the 142 currently recognized genera of Sapindaceae, 96 are restricted to one area, 33 are distributed in two areas, seven in three areas, two in four areas and three in five areas. Current generic diversities for each area are represented in figure 1. The highest generic diversity within Sapindaceae is found in South-East Asia (area D; 45 genera and 24% endemism), whereas the lowest occurred in Northern America (area F; four genera and 50% endemism) (Fig. 1). South America (area G) had the highest percentage of endemism with 91.4% (34.4% of these genera are monotypic; Fig. 1). The results of the Lagrange analysis based on PL median ages are outlined in figure 2. Dispersal and extinction events were quantified based on the PL median divergence time estimates (Fig. 2; Table 1, 2) and overall, 57 dispersals and 37 extinctions were required to explain the biogeographic history of the soapberry family (Tables 1, 2; Table 1 of the Supplementary material). Dispersals through time were distributed as follows: one before 80 million years ago (Ma) (time slice 1; Fig. 3A), one between 80 and 60 Ma (time slice 2; Fig. 3B), seven between 60 and 30 Ma (time slice 3; Fig. 3C) and 48 in the last 30 Ma (time slice 4; Fig. 3D) (Fig. 3, Table 1). South-East Asia and Africa had the highest numbers of extinctions with respectively ten and 11 each, whereas only one extinction event occurred in Northern America (Fig. 3, Table 2). Although South-East Asia and Africa were the scene of similar numbers of extinction events, they took place over a period of 60 Ma in South-East Asia while all occurred in the last 30 Ma in Africa. This suggests that the cooling at the E-O boundary had a more lethal effect on the floras of Africa than those in South-East Asia.

Biogeographic reconstructions suggest that Sapindaceae initially originated in temperate Asia (area A) sometime in the Early Cretaceous (Figs. 2, 3A). Diversification events taking place during this period led to the evolution of all four subfamilies now recognized (ca. 90 Ma in area A; Figs. 2, 3A). Until the Late Paleocene, the main lineages (subfamilies) remained restricted to Laurasia. However, ecological preferences were observed, with the colonization of temperate regions by subfamily Hippocastanoideae (areas A, F; Figs. 2, 3A-B), and subtropical to tropical regions by the remainder of the family (areas A, D; Figs. 2, 3A-B). In the Late Paleocene, a first colonization event of the southern hemisphere from South-East Asia occurred. This region was an important catalyser during the evolutionary history of Sapindaceae by connecting the northern and southern hemispheres (at this period, South-East Asia included the Malay Peninsula and Borneo; Fig. 3C). Furthermore, the spread of Sapindaceae was favoured by the break-up of Gondwana (i.e., Australia and India began their northern migration) and climatic conditions. From the Late Paleocene to the end of the Eocene, the climate was warm and ice probably only occurred in the Antarctic highlands and within and around the Arctic Ocean (Bowen, 2007; Zhonghui et al., 2009). Climatic and tectonic conditions were thus compatible with the dispersal of Sapindaceae from South-East Asia to Australia, Africa and even South America (Figs. 2, 3C).

Unlike the family Cucurbitaceae, which dispersed from Asia to South America through Africa (Schaefer et al., 2009), the colonization of South America by Sapindaceae took place from Australia through Antarctica (Fig. 3C). Once established in South America, ecological conditions and several key innovations triggered high diversification rates in this lineage, which comprises more than 590 extant species (ca. 31% of the family; Buerki et al., 2009). The development of zygomorphic flowers and a liana habit (currently restricted to the Paullinia group; Fig. 2) are likely to have influenced the diversification of the family in South America.² These innovations might have enhanced the complexity of their interactions with pollinating insects in an understorey environment. During the same geological period, a

² Zygomorphic flowers were shown to increase significantly the success of pollination by animals (Sargent, 2004). Gianoli (2004) also established that liana clades had generally higher diversities than their tree or shrub sister clades. It is also likely that liana clades are more likely to demonstrate long-branching patterns.

dispersal event occurred from South-East Asia to Africa (most probably through India; Fig. 3C). This single event was at the origin of most of the African generic diversity (75% of the sampled African genera belong to this clade; e.g., *Chytranthus*, *Eriocoelum*; see the Litchi group; Figs. 2, 3C).

Strikingly, processes of diversification within Sapindaceae are strongly linked with paleoclimatic changes that allowed recurrent events of dispersal and isolation among neighbouring land masses. At the E-O boundary, the abrupt decrease of temperature resulted in the appearance of ice sheets in Antarctica (Coxall et al., 2005; Zhonghui et al., 2009), drought in southern regions (especially in Australia and Africa; Bowen, 2007) and subsequent reduction of the tropical belt. At the same time, intense volcanic activity was recorded in South-East Asia, producing new islands such as Java and the Philippines. Although the spread of some lineages is concomitant with the creation of new habitats in Australia (e.g. biogeographic analyses highlight the dispersal of the most recent common ancestor of *Diplopeltis* from South-East Asia to Australia at this period to colonize subarid regions; Fig. 3D), a majority of the lineages initiated an important northern migration (15 dispersals from Australia to southern Asia and 11 dispersals from southern Asia to Eurasia; Fig. 3D) along an ecological cline in response to the northwards shift of tropical zones. Despite the occurrence of extinctions during this process (especially in Africa and to a lesser extent in South-East Asia; Fig. 3D), the abrupt climate change of the E-O boundary apparently triggered a phase of diversification of Sapindaceae. Geological and climatic conditions encountered in South-East Asia³ at that time might have enhanced ecological interactions between lineages coming from either Australia or Eurasia (Fig. 3D). The multiplication of contacts between such lineages increased the chance of responding favourably to the local environmental diversity (e.g. by favouring hybridization between lineages and creating new genotypes), and augmented the probability of emergence and fixation of key innovations, whereas the insular properties of this area might have favoured

³ Southeast Asia had a constant tropical climate through time and occupied an East-West latitudinal gradient. Moreover, this area was composed by many islands. See Sodhi et al. (2004) for more details.

reproductive isolation and speciation processes (Fig. 3D). This hypothesis is supported by the presence of highest Sapindaceae diversity in South-East Asia and the high number of shared genera with concomitant areas (Fig. 1). In the Miocene, two additional abiotic factors [subequatorial currents and the West Wind Drift (WWD)] might have promoted long-distance dispersals (LDD) in Sapindaceae (Figs. 2, 3D). The South Equatorial Current, which originates off the western coast of South Africa, flows westward towards the Brazilian shelf, and splits with one branch flowing northwards to the Caribbean and the other southwards (Stramma and England, 1999). Biogeographic reconstruction demonstrates that this current might have affected the relationships between the South American and African lineages of Sapindaceae (e.g., by allowing LDD of species of *Allophylus*, *Cardiospermum* and *Haplocoelum* in Africa; Figs. 2, 3D). A similar pattern was, for instance, underlined in the pantropical Malvaceae complex *Hibiscus tiliaceus* (Takayama et al., 2008). Sanmartín et al. (2007) pointed out that eastward LDD mediated by the WWD occurred more frequently between Australian and South American plant lineages. The patterns presented here are in agreement with the hypothesis proposed by Sanmartín et al. (2007). At least two LDDs from Australia to South America were identified at the origin of the most recent common ancestors of the highly diversified genera *Cupania* and *Matayba* (Figs. 2, 3D) and no dispersal in the opposite direction was detected. Our results highlight the positive effect of abiotic factors (especially climate changes) as driving forces promoting diversification in Sapindaceae. Such patterns were also highlighted in placental mammals after the K-T mass extinction, which colonized numerous niches previously occupied by the dinosaurs before their extinction (Springer et al., 2003; but see Bininda-Emonds et al., 2007 for a contrasting view). In the case of Sapindaceae, however, it is not yet possible to assess whether the diversification processes that occurred during the E-O boundary are related to the colonization of “empty” niches rather than with the evolution of key innovations (e.g., development of arillode).

The intricate biogeographic history of Sapindaceae might provide evidence to explain the high level of para/polyphyly at tribal and generic levels and the difficulties encountered by previous authors proposing taxonomic frameworks for the family (Buerki et al., 2009). This

indicates that the processes operating across ecological gradients (along the dispersal routes of Sapindaceae) may be more important than phylogenetic (or even geographic) isolation in generating and/or maintaining phenotypic diversity. Such a pattern has, for instance, been observed in Australian rainforest lizards (Schneider et al., 1999).

Conclusion

It has been widely hypothesized that abiotic factors strongly influencing global climate have been involved in dramatic large-scale extinction processes (Jablonski, 2001; McElwain and Punyasena, 2007). The consequences of these events caused for instance the extinction of ca. 30-60% of plant species as well as a majority of animals during the Cretaceous-Tertiary mass extinction (Jablonski, 2001). Conversely, diversification processes are mostly seen as being driven by biotic factors. For instance, co-evolutionary processes between plant and pollinators or disseminators are likely to have driven key innovations in inflorescence structures of seed plants (Hu et al., 2008). In the present study, however, we show a strongly contrasting pattern in which abiotic factors (i.e., Late-Oligocene climatic oscillations, subequatorial currents and West Wind Drift) appear to have catalyzed the driving forces that enhanced diversification rates in Sapindaceae, and perhaps also in other flowering plants.

References

- Alley, R.B., Marotzke, J., Nordhaus, W.D., Overpeck, J.T., Peteet, D.M., Pielke Jr, R.A., Pierrehumbert, R.T., Rhines, P.B., Stocker, T.F., Talley, L.D., Wallace, J.M., 2003. Abrupt climate change. *Science* 299, 2005—2010.
- Bininda-Emonds, O.R.P., Cardillo, M., Jones, K.E., MacPhee, R.D.E., Beck, R.M.D., Grenyer, R., Price, S.A., Vos, R.A., Gittleman, J.L., Purvis, A., 2007. The delayed rise of present-day mammals. *Nature* 446, 507—512.
- Bowen, G.J., 2007. When the world turned cold. *Nature* 445, 607—608.

- Buerki, S., Forest, F., Acevedo-Rodríguez, P., Callmander, M.W., Nylander, J.A.A., Harrington, M., Sanmartín, I., Küpfer, P., Alvarez, N., 2009. Plastid and nuclear DNA markers reveal intricate relationships at subfamilial and tribal levels in the soapberry family (Sapindaceae). *Mol. Phylogenet. Evol.* 51, 238—258.
- Coxall, K.H., Wilson, P.A., Pälike, H., Lear, C.H., Backman, J., 2005. Rapid stepwise onset of Antarctic glaciation and deeper calcite compensation in the Pacific Ocean. *Nature* 433, 53—57.
- Crepet, W. L., Niklas, K. J., 2009. Darwin's second "abominable mystery": Why are there so many angiosperm species? *Am. J. Bot.* 96, 366—381.
- Frohlich, M.W., Chase, M.W., 2007. After a dozen years of progress the origin of angiosperms is still a great mystery. *Nature* 450, 1184—1189.
- Gianoli, E., 2004. Evolution of a climbing habit promotes diversification in flowering plants. *Proc. R. Soc. Lond. B.* 271, 2011—2015.
- Grimaldi, D., 1999. The co-radiations of pollinating insects and angiosperms in the Cretaceous. *Ann. Missouri Bot. Gard.* 86, 373—406.
- Hu, S., Dilcher, D.L., Jarzen, D.M., Taylor, D.W., 2008. Early steps of angiosperm-pollinator coevolution. *Proc. Natl. Acad. Sci. USA* 105, 240—245.
- Jablonski, D., 2001. Lessons from the past: Evolutionary impacts of mass extinctions. *Proc. Natl. Acad. Sci. USA* 98, 5393—5398.
- Jaramillo, C., Rueda, M.J., Mora, G., 2006. Cenozoic Plant Diversity in the Neotropics. *Science* 311, 1893—1896.
- Katz, M.E., Miller, K.G., Wright, J.D., Wade, B.S., Browning, J.V., Cramer, B.S., Rosenthal, Y., 2008. Stepwise transition from the Eocene greenhouse to the Oligocene icehouse. *Nature Geoscience* 1, 329—334.
- Labandeira, C. C., 2006. The four phases of plant-arthropod associations in deep time. *Geologica Acta* 4, 409—438.
- Magallón, S., Sanderson, M.J., 2005. Absolute diversification rates in angiosperm clades. *Evolution* 55, 1762—1780.

- McElwain, J.C., Punyasena, S.W., 2007. Mass extinction events and the plant fossil record, *Trends Ecol. Evol.* 22, 548—557.
- Paradis, E., Claude, J., Strimmer, K., 2004. APE: analyses of phylogenetic and evolution in R language. *Bioinformatics* 20, 289—290.
- R Development Core Team 2009. A Language and Environment for Statistical Computing. Available at <http://www.R-project.org>.
- Raup, D.M., Sepkoski, J.J. Jr., 1982. Mass extinctions in the marine fossil record. *Science* 215,1501—1503.
- Raven, P.H., 1977. A suggestion concerning the Cretaceous rise to dominance of the angiosperms. *Evolution* 31, 451—452.
- Ree, R.H., Smith, S.A., 2008. Maximum Likelihood Inference of Geographic Range Evolution by Dispersal, Local Extinction, and Cladogenesis. *Syst. Biol.* 57, 4—14.
- Sanderson, M.J., 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol. Biol. Evol.* 19,101—109.
- Sanmartín, I., Wanntorp, L. Winkworth, R.C., 2007. West Wind Drift revisited: testing for directional dispersal in the Southern Hemisphere using event-based tree fitting. *J. Biogeogr.* 34, 398—416.
- Sargent, R.D., 2004. Floral symmetry affects speciation rates in angiosperms. *Proc. R. Soc. Lond. B.* 271, 603—608.
- Schaefer, H., Heibl, C., Renner, S.S. 2009. Gourds afloat: a dated phylogeny reveals an Asian origin of the gourd family (Cucurbitaceae) and numerous oversea dispersal events. *Proc. R. Soc. Lond. B.* 276, 843—851.
- Schneider, C.J., Smith, T.B., Larison, B., Moritz, C., 1999. A test of alternative models of diversification in tropical rainforests: Ecological gradients vs. rainforest refugia. *Proc. Natl. Acad. Sci. USA* 96, 13869—13873.
- Sodhi, N.J., Koh, L.P., Brook, B.W., Ng, P.K.L., 2004. South-East Asian biodiversity: an impending disaster. *Trends Ecol. Evol.* 19, 654—660.

- Soltis, D.E., Soltis, P.S., Endress P.K., Chase, M.W., 2005. Phylogeny and Evolution of Angiosperms. Sinauer, Sunderland, M.A.
- Springer, M.S., Murphy, W.J., Eizirik, E., O'Brien, S.J., 2003. Placental mammal diversification and the Cretaceous–Tertiary boundary. *Proc. Natl. Acad. Sci. USA* 100, 1056—1061.
- Stramma, L., England, M., 1999. On the water masses and mean circulation of the South Atlantic Ocean. *Journal of Geophysical Research* 104, 20863—20883.
- Takayama, K., Tateishi, Y., Murata, J., Kajita, T., 2008. Gene flow and population subdivision in a pantropical plant with sea-drifted seeds *Hibiscus tiliaceus* and its allied species: evidence from microsatellite analyses. *Mol. Ecol.* 17, 2730—2742.
- Wikström, N., Savolainen, S., Chase, M. W., 2001. Evolution of the angiosperms: calibrating the family tree. *Proc. R. Soc. Lond. B.* 268, 2211—2220.
- Wilf, P., Labandeira, C.C., Kress, W.J., Staines, C.L., Windsor, D.M., Allen, A.L., Johnson, K.R., 2000. Timing the radiations of leaf beetles: hispines on gingers from latest Cretaceous to recent. *Science* 289, 291—294.
- Zhonghui, L., Pagani, M., Zinniker, D., DeConto, R., Huber, M., Brinkhuis, H., Shah, S.R., Leckie, R.M., Pearson, A., 2009. Global cooling during the Eocene-Oligocene climate transition. *Science* 323, 1187—1190.

Acknowledgements

We wish to thank Phillipe K pfer and Giordina Bernasconi (University of Neuch tel, Switzerland) for their enthusiasm, comments and help that contributed to the elaboration of this study. We also thank Johan A. A. Nylander (University of Stockholm, Sweden) and Nils Arrigo (University of Neuch tel) for their support in bioinformatics analyses as well as Pete Lowry and Martin Callmander (Missouri Botanical Garden, U.S.A.) for their valuable support. Financial support to SB was provided by the Laboratory of Evolutionary of the University of Neuch tel (Switzerland) and the European Community's Programme Structuring the

European Research Area under a SYNTHESYS grant to visit the Real Jardín Botánico (CSIC) in Madrid.

Figure legends

Fig. 1. Current generic diversity of Sapindaceae per area (see Buerki et al., chapter 2 for more details on the definition of the areas). For each area, numbers separated by back slashes are as follows: number of endemic genera; number of monotypic endemic genera (e.g., A 9/6). The number of genera and their distribution is based on Buerki et al. (2009).

Fig. 2. Lagrange stratified biogeographic reconstruction based on PL median node age estimations. Pie charts on nodes represent ancestral area(s). Hard incongruent nodes are also indicated (the first area(s) referred to the PL maximum, second PL median and third PL minimum node age). The classification proposed in Buerki et al. (2009) is also represented.

Fig. 3. Biogeographic scenario through time inferred from PL median Lagrange Str analysis. A. before 80 Ma; B. before 60 Ma; C. before 30 Ma; D. between 30 and 0 Ma. Dispersals and extinctions are represented in proportion of their occurrence respectively by arrows and circles. See tables 1 and 2 for more details on the number of events.

Tables

Table 1. Numbers of dispersals through time. See text for more details.

Time slice	From-To	Dispersals
1	A-F	1
2	A-D	1
3	A-D	3
	D-B	1
	D-E	2
4	E-G	1
	A-D	2
	A-F	1
	A-G	1
	B-C	1
	B-D	2
	B-G	1
	C-B	5
	D-A	11
	D-B	1
	D-C	1
	D-E	2
	E-D	15
E-G	2	
G-B	3	
Total:		57

Table 2. Numbers of extinctions through time for each area. Numbers between brackets referred to respectively observed extinctions and extinctions predicted by the model. See text for more details on areas and time slice definitions.

Area	Extinction through time		Σ Extinctions
	Time slice 3	Time slice 4	
A	2 (1/2)	3 (2/1)	5
B	0	11 (9/2)	11
C	0	2 (2/0)	2
D	4 (4/1)	6 (6/0)	10
E	1 (1/0)	4 (3/1)	5
F	1 (1/0)	0	1
G	0	3 (2/1)	3
Total:	8	29	37 (29/8)

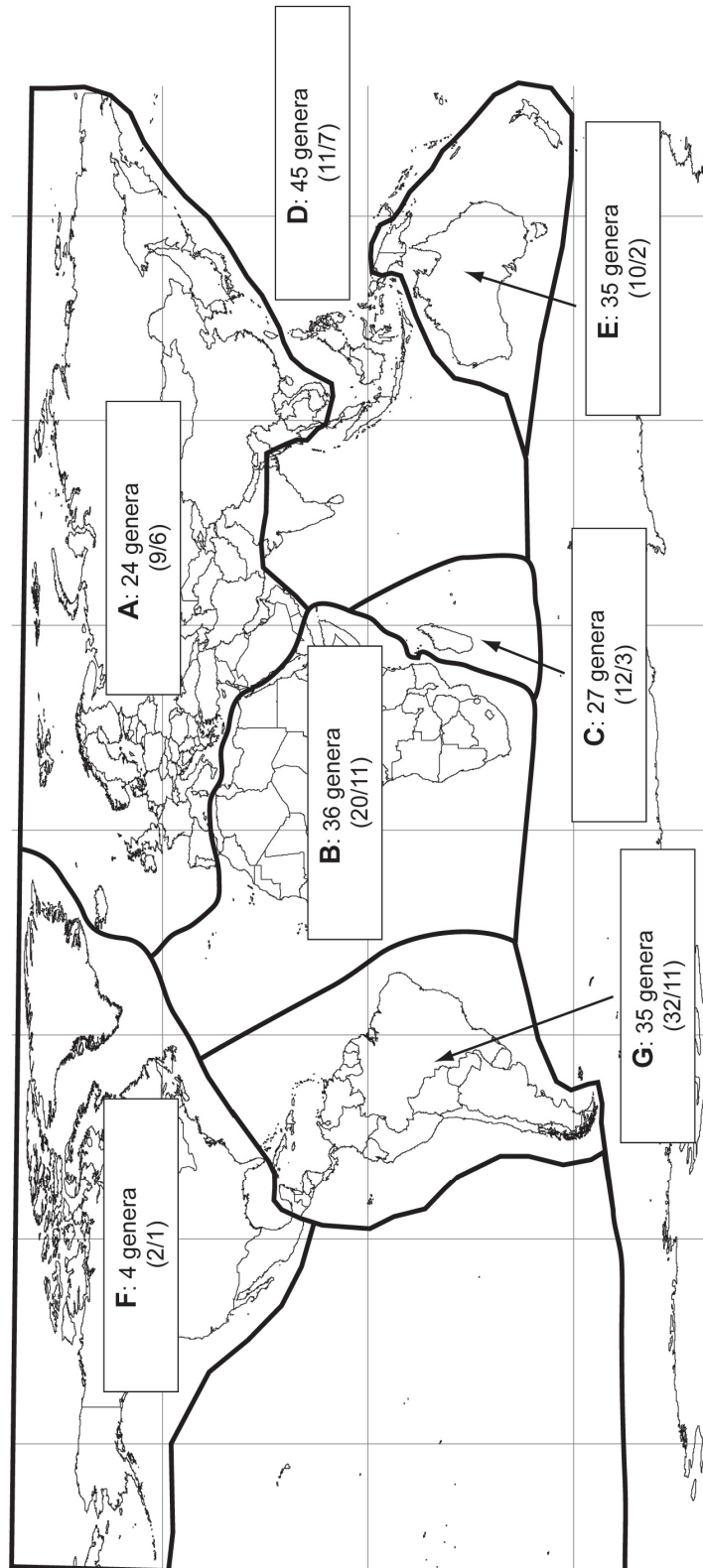


Figure 1

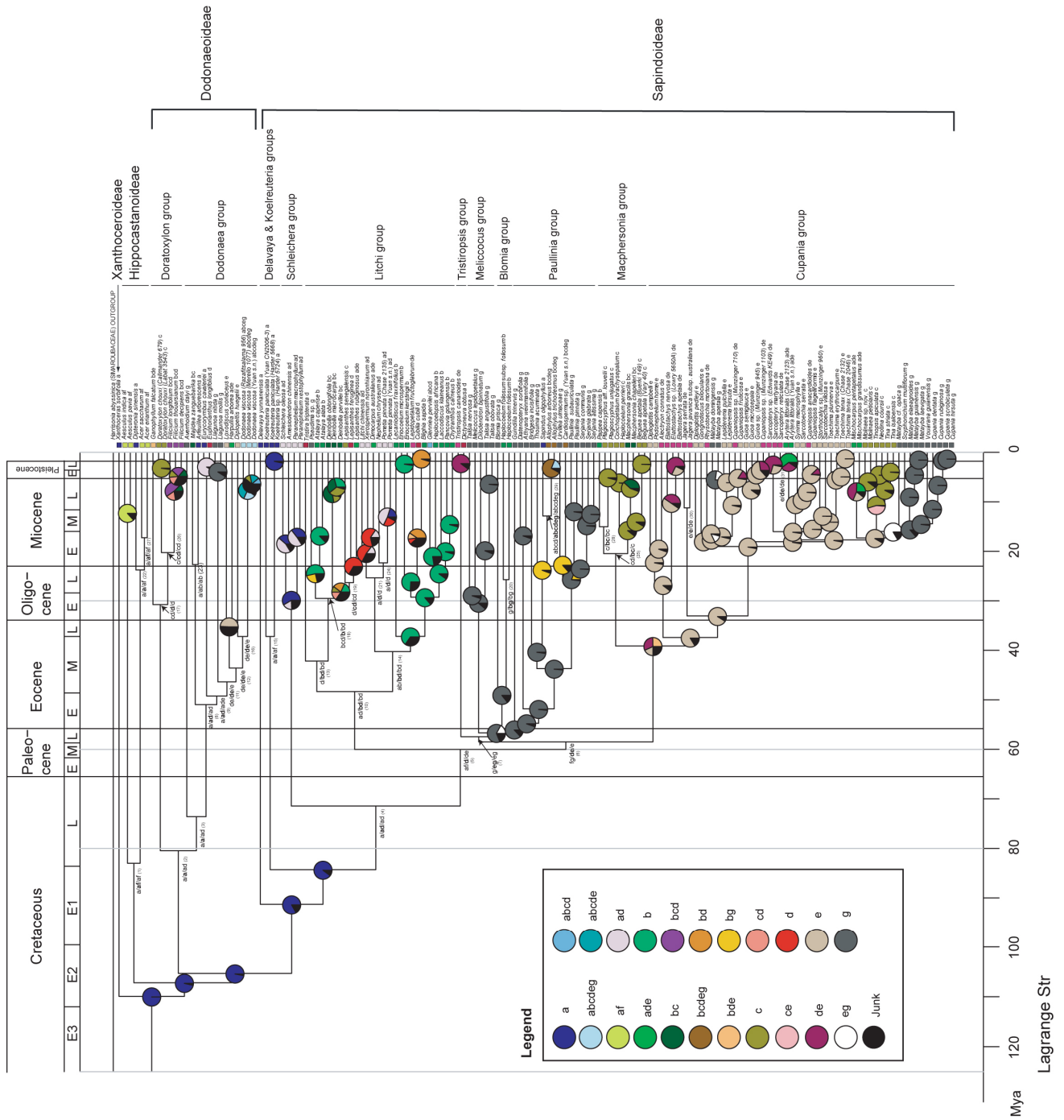


Figure 2

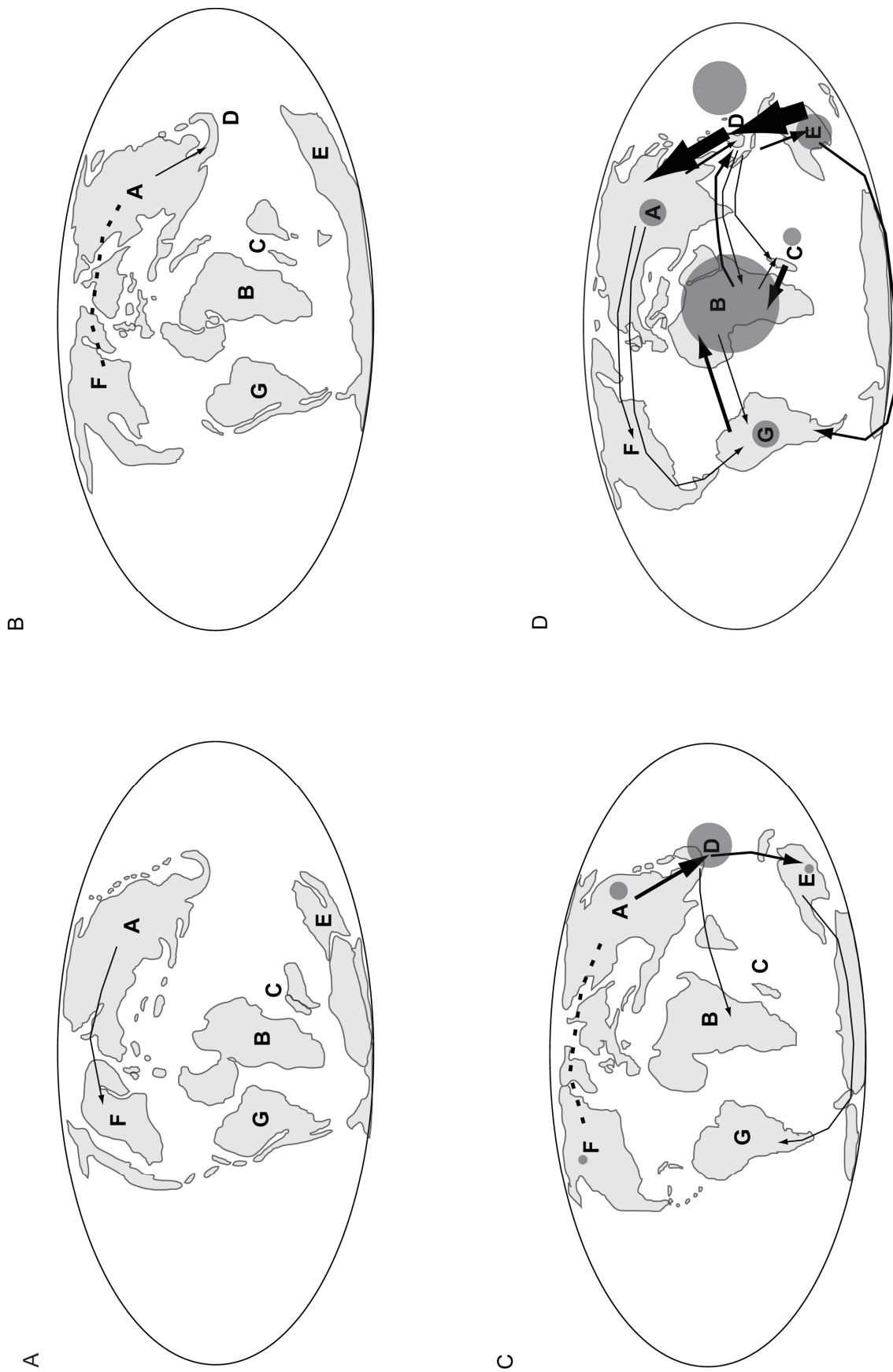


Figure 3

Supplementary material

Table 1. Contingency table showing number and type of transition events between geographic ranges (change in geographic range from ancestor to descendant) inferred by a temporally stratified likelihood DEC model (Lagrange Str) on the PL chronogram of Sapindaceae (Fig. 3b). Transition events are sorted by time slice (see Fig. 2). Transitions are expressed as rates of local extinction (*Ea*) and/or range expansion (*Dab*), according to the Q transition probability matrix represented in Table 2. Fonts indicate different types of transition events between geographic ranges: *Italic*: local extinction (range contraction); Roman: dispersal (range expansion). **Bold**: transitions between ranges that are separated by more than one dispersal or extinction event; for transitions involving dispersal (range expansion), the rate is the sum of rates from areas in the starting range to the target area (e.g., Table 2: from ab to abc = $Dab+Dbc$ (Ree & Sanmartín, 2009)). **Bold ***: transitions that involve two or more events of range expansion and for which there is uncertainty in the direction of dispersals (e.g., a to abc); red: these transition events were not considered. See Material and Methods for more details on areas and time slice definitions.

		To Area(s)																							
		Time slice	a	ab	abc	abcd	abcde	abcdeg	ad	ade	af	b	bc	bcd	bcdeg	bd	bde	bg	c	cd	d	de	e	eg	g
From Area(s)	a	1									1 (Daf)														
		2							1 (Dad)																
		3							1 (Dad)																
		4							2 (Dad)		1 (Daf)											1 (Dad+Ea)			1 (Dag+Ea)
	ab	4	1 (Eb)																						
	abcdeg	4													1 (Ea)										
	ad	3	1 (Ed)																			1 (Ea)			1 (Dae+Dde+Ea)
		4	1 (Ed)																			1 (Ea)			
	ade	4								1 (Ee)															
	af	3	1 (Ef)																						
	b	4				1																			
	bc	4																							
	bd	3											1 (Ec)												1 (Dbq+Eb)
		4											1 (Ed)												
	bg	4											2 (Ed)												
	c	4																							
	cd	4																							
	d	3																							
		4																							
	de	3								4 (Dda)	2 (Dda+Dde)*														
		4																							
	e	4																							
	eg	3																							
		4																							
	g	4																							

**Chapter four: Molecular phylogenetics and morphological
evidence support recognition of *Gereaua*, a new endemic
genus of Sapindaceae from Madagascar**

Sven Buerki, Peter B. Phillipson, Porter P. Lowry II and Martin W. Callmander

Accepted in *Systematic Botany* pending minor revisions

**Molecular phylogenetics and morphological evidence support recognition of
Gereaua, a new endemic genus of Sapindaceae from Madagascar**

Sven Buerki^{1,5}, Peter B. Phillipson^{2,3}, Porter P. Lowry II^{2,3} and Martin W. Callmander^{2,4}

¹ Institute of Biology, University of Neuchâtel, Rue Emile-Argand 11, CH-2009 Neuchâtel, Switzerland.

² Missouri Botanical Garden, P.O. Box 299, St. Louis, MO 63166-0299, U.S.A.

³ Muséum National d'Histoire Naturelle de Paris, Case Postale 39, 57 rue Cuvier, 75231 05 CEDEX, Paris, France.

⁴ Conservatoire et Jardin botaniques de la ville de Genève, ch. de l'Impératrice 1, CH-1292 Chambésy, Switzerland.

⁵ Author for correspondence (sven.buerki@unine.ch)

BUERKI, PHILLIPSON, LOWRYII AND CALLMANDER, GEREUA (SAPINDACEAE) AN ENDEMIC GENUS FROM MADAGASCAR

Abstract— A recent worldwide phylogeny of Sapindaceae inferred from nuclear and plastid DNA regions segregated the Malagasy *Haplocoelum perrieri* Capuron from the African *Haplocoelum foliosum* (Hiern) Bullock. Additional phylogenetic analyses conducted here (including material of *H. inopleum* Radlk., the generi-type) confirmed the result from the previous analysis and showed that maintaining a broad circumscription of *Haplocoelum* to include the Malagasy species would render the genus polyphyletic. In order to maintain monophyly, it is necessary to exclude *H. perrieri*, which we transfer to a new, monotypic genus, described here as ***Gereaua* Buerki & Callm. INED**. This taxon is easily distinguished from the species retained in *Haplocoelum* by the following morphological characters: (1) sexual dimorphic inflorescences in racemules (vs. monomorphic inflorescences in fascicle of cymes); (2) 2-locular ovary (vs. 3-locular ovary); (3) rudimentary ovary in staminate flowers (vs. ovary well developed in staminate flowers); (4) corolla with 4-5 petals (vs. apetalous); (5) glabrous anthers (vs. pubescent anthers); (6) pubescent fruit (vs. glabrous fruit). Relationships with the most-closely related genera, included in the *Macphersonia* group, are discussed in light of molecular, morphological and biogeographic evidence. A preliminary threat assessment of ***G. perrieri* (Capuron) Buerki & Callm. INED** using the IUCN Red List criteria indicates a status of Least Concerned.

Keywords— *Gereaua*, *Haplocoelum*, IUCN criteria, Madagascar, new genus, Sapindaceae.

The economically important soapberry family (Sapindaceae), recently expanded to include several other traditionally recognized families such as Aceraceae and Hippocastanaceae (Soltis et al. 2000; APGII, 2003; Harrington et al. 2005), comprises about 1900 species and 141 genera (Acevedo-Rodríguez, pers. comm.; Buerki et al., 2009) mainly found in the tropics, with only a few genera restricted to temperate areas (e.g., *Acer* L. and *Aesculus* L.).

A recent family-wide phylogeny of Sapindaceae sensu lato inferred from nuclear and plastid markers from representatives of more than 60% of the currently recognized genera revealed intricate relationships at the subfamilial and tribal levels (Buerki et al., 2009). The molecular phylogenetic study confirmed that the monotypic Chinese genus *Xanthoceras* Bunge is sister to the rest of the family, within which subfamily Hippocastanoideae (comprising the previously recognized Aceraceae and Hippocastanaceae) is in turn sister to a clade comprising subfamilies Dodonaeoideae and Sapindoideae (Harrington et al. 2005; Thorne 2007). This study also showed that Dodonaeoideae and Sapindoideae (as defined by Thorne 2007) are paraphyletic and that nearly all of the 14 historically recognized tribes within these two subfamilies are also polyphyletic (Paullinieae appear to be monophyletic and the status of Cossinieae and Koelreuterieae could not be tested because only a single genus from each was included in the analysis). By comparing the results of their molecular analysis with data on morphology, Buerki et al. (2009) redefined Dodonaeoideae and Sapindoideae to render them monophyletic and proposed a revised informal classification for the family (Table 1).

The family-wide phylogenetic analysis also revealed that at least five genera are polyphyletic as currently circumscribed, viz. *Cupaniopsis* Radlk., *Guioa* Cav., *Haplocoelum* Radlk., *Matayba* Aublet and *Sarcotoechia* Radlk. (Buerki et al., 2009). One of these genera, *Haplocoelum*, a member of subfam. Sapindoideae, has puzzled taxonomists for nearly a century (Bullock 1931; Radlkofer 1933; Davies and Verdcourt 1998; Capuron 1969). As currently circumscribed by most authors (e.g., Lebrun and Stork 1992, 1997), *Haplocoelum*

comprises four or five species distributed in central and eastern tropical Africa (Davies and Verdcourt 1998) and one species in Madagascar (Capuron 1969), although delimitation of the African species remains controversial, especially among those occurring in the central and eastern parts of the continent (see Davies and Verdcourt 1998). The Malagasy representative, *H. perrieri* Capuron, differs from the African species in several important ways. For example, its ovary is 2-locular in pistillate flowers but rudimentary in staminate flowers and the anthers are glabrous (vs. a 3-locular ovary in the pistillate and staminate flowers and pubescent anthers in the African taxa; Table 2). This led Capuron (1969) to consider recognizing the Malagasy species as a distinct genus, a view later echoed by Davies and Verdcourt (1998), although ultimately he opted to place it in a separate section, *Cardiophyllariopsis* Capuron. He justified this decision by pointing out that the Malagasy species shares several prominent morphological characters with two African species that were generally included in *Haplocoelum* at that time: *H. jubense* Choiv. (which, like *H. perrieri*, has a developed corolla) and *H. scassellatii* Choiv. (which is similar to *H. perrieri* in having an indument on the fruit). Later, these two African taxa were shown to be misplaced in *Haplocoelum*, and each was placed in synonymy under a previously recognized species in a different genus: *H. jubense* under *Camptolepis ramiflora* Radlk. (Friis and Vollensen 1985) and *H. scassellatii* as a subspecies within *Lecaniodiscus fraxinifolius* Baker (Friis 1984).

The family-wide phylogenetic analysis of Sapindaceae revealed that the Malagasy species currently recognized as *Haplocoelum perrieri* is not closely related to the African *H. foliosum* (Hiern) Bullock (Buerki et al., 2009). The African representative was placed in the *Blomia* group, whereas the Malagasy species belongs to the *Macphersonia* group (Table 1). In the present study, we expanded the data set of Buerki et al. (2009) with the addition of two species of *Haplocoelum* (*H. inopleum* Radlk., the generi- type, and *H. gallense* (Engl.) Radlk.) in order to clarify the relationships with the Malagasy taxon (especially with regard to the type species). Moreover, to provide a refined assessment of the phylogenetic position of *H. perrieri* in the *Macphersonia* group, we also added one species from each of two endemic Malagasy

genera: *Chouxia* Capuron (*C. macrophylla* G.E. Schatz, Gereau & Lowry) and *Pseudopteris* Baill. (*P. decipiens* Baill.) to the data set of Buerki et al. (2009), which originally comprised 8 species representing 6 genera belonging to the *Macphersonia* group.

The Malagasy genus *Conchopetalum* is currently regarded as the closest relative of *Haplocoelum perrieri* (Buerki et al., 2009). However, *Conchopetalum*, which comprises only two species, is strikingly different from *H. perrieri* with its relatively large leaves and flowers. Moreover, *H. perrieri* differs morphologically in many other ways, notably by its inflorescence in racemules (female 1-flowered) (vs. in fascicules or pseudo-umbelliform in *Conchopetalum*), presence of an arillode (vs. only a dry white sarcotesta surrounding the hylum) and the pubescent, 1-seeded fruit (vs. glabrous and 3-6-seeded) (Table 3). It is difficult to envision placing *H. perrieri* within *Conchopetalum*, which suggests that it is best treated as a new monotypic genus. In the present study we thus aim to (1) clarify the phylogenetic relationships and position of *H. perrieri*; (2) describe a new genus to accommodate the endemic species from Madagascar currently placed in *Haplocoelum*; (3) discuss the relationships of this taxon with the other genera in the *Macphersonia* group (in particular those it most closely resembles morphologically); and (4) provide a preliminary threat assessment of this species using the IUCN Red List criteria (IUCN 2001).

MATERIALS AND METHODS

Sampling, sequence data and phylogenetic analyses– The sample employed for the present study includes members of 106 of the 141 currently recognized genera of Sapindaceae (Buerki et al., 2009), representing all subfamilies and tribes recognized in the traditional systems of Radlkofer (1878, 1933) and Müller and Leenhouts (1976), and in updated systems proposed by Harrington et al. (2005) and Buerki et al. (2009). Ingroup sampling comprised the 154 specimens in the dataset of Buerki et al. (2009) along with the addition of two species of *Haplocoelum* [*H. inopleum* (Lap 117) and *H. gallense* (see Harrington et al. 2005 for voucher information)], one species of *Chouxia* (*C. macrophylla*; Lowry et al. 5199) and one species of

Pseudopteris (*P. decipiens*; Service Forestier 12529). The outgroup included one accession each from the families Anacardiaceae (*Sorindeia* sp.; defined as the outgroup in previous analyses; Savolainen et al. 2000; Muellner et al. 2007), Simaroubaceae (*Harrisonia abyssinica*) and Meliaceae (*Malleastrum* sp.).

Seven plastid regions [coding regions *rpoB* and *matK*; *trnL* intron; intergenic spacers (IGS) *trnK-matK*, *trnD-trnT*, *trnL-trnF* and *trnS-trnG*] and one nuclear region (the ITS region comprising the *ITS1* and *ITS2* internal transcribed spacers, and the 5.8S region) were amplified. Information on the primers, PCR and sequencing protocols are given in Buerki et al. (2009). Voucher citations and GenBank accession numbers for most of the specimens included in this study are cited in Buerki et al. (2009) and in Harrington et al. (2005) for a single species (*Haplocoelum gallense*). GenBank numbers for the new sequences generated for the present study are as follows: *Chouxia macrophylla* (*ITS* region: FJ514258; *matK*: FJ514260; *rpoB*: FJ514261; *trnD-trnT* IGS: FJ514262; *trnK-matK*: FJ514263; *trnL* intron: FJ514264; *trnL-trnF* IGS: FJ514266; *trnS-trnG* IGS: FJ514268), *H. inopleum* (*ITS* region: FJ514259; *trnL* intron: FJ514265; *trnL-trnF* IGS: FJ514267) and *Pseudopteris decipiens* (*ITS* region: EU720480; *rpoB*: EU720816). Alignment statistics for the seven plastid regions and the nuclear region are presented in Buerki et al. (2009).

Single-gene, total evidence (sensu Kluge 1989) analyses and their corresponding bootstrap analyses were performed using the maximum likelihood (ML) and maximum parsimony (MP) criteria following the same procedure as in Buerki et al. (2009). In the present study, nodes with bootstrap values (BS) below 50% are considered unsupported, 50-74% are considered weakly supported, 75-89% are moderately supported and 90-100% are strongly supported. Topological differences between single-gene phylogenetic trees were compared using TreeJuxtaposer (Munzner et al. 2003), taking into account the level of resolution obtained by each marker and their bootstrap support. Topological differences with bootstrap support inferior to 75% were not considered.

Herbarium material– For our study we examined all material of *Haplocoelum* (including *H. perrieri* from Madagascar) and the genera belonging to the *Macphersonia* group available at the herbaria in Geneva (G), Kew (K), Missouri (MO), Paris (P) and Antananarivo (TAN & TEF) (herbarium citations follow Holmgren et al. 1990). Historical collections lacking geographic coordinates were post-facto geo-referenced as accurately as possible using the “Gazetteer to Malagasy Botanical Collecting Localities” (Schatz and Lescot 2005; available at: <http://www.mobot.org/MOBOT/Research/madagascar/gazetteer/>) and other sources, and are indicated in square brackets in the citation of material examined in the taxonomic treatment that follows. The species distribution map was generated using ESRI ArcView 3.3 software (2000) and projected on Madagascar’s five broad bioclimatic zones (after Cornet 1974; adapted by Schatz 2000).

Conservation status– The conservation status of *Haplocoelum perrieri* was assessed using the current IUCN Red List Criteria (IUCN 2001). Calculation of the area of occupancy (AOO), extent of occurrence (EOO) and number of subpopulations followed the methods used by Callmander et al. (2007), with a 3 x 3 km grid cell size used to calculate EOO.

RESULTS

Phylogenetic analyses– The best-fit model for all partitions was the general time reversible (GTR) model with an alpha parameter for the shape of the gamma distribution to account for among-site rate heterogeneity (Yang 1993). The only exception was for the ITS region, for which a proportion of invariable sites was added. Although the MP and ML single-gene analyses provided topologies with different levels of resolution within Sapindaceae sensu lato (i.e., several parts of the MP trees were not fully resolved), no moderately to strongly supported differences (BS > 75%) were observed between single-gene trees. The most parsimonious tree for the combined analyses under the MP criterion was 9912 steps in length (CI=0.503, RI=0.726). Under the ML criterion, the best-fit model for the combined matrix was GTR, with a proportion of invariable sites and an alpha parameter for the shape of the gamma

distribution to account for among-site rate heterogeneity (Yang 1993). This model was used to perform the ML search (log likelihood was -69695.60) followed by 1000 rapid bootstrap analyses.

The two total evidence trees were highly congruent and showed the same major groups of Sapindaceae sensu lato as presented in Buerki et al. (2009) (Table 1). Based on the MP and ML total evidence analyses performed here, the four species of *Haplocoelum* included in our sample clustered in three different groups (see Table 1). The African species *H. gallense* (considered a subspecies of *H. foliosum* by Davies and Verdcourt 1998) occupied a position within the *Litchi* group and the two other African taxa sampled, *H. foliosum* and *H. inopleum*, belonged to the *Blomia* group, whereas *H. perrieri* was part of the Afro-Malagasy *Macphersonia* group. Total evidence topologies clearly showed that the exemplar species of *Chouxia* and *Pseudopteris* belong to the *Macphersonia* group. With the addition of these species, the *Macphersonia* group is resolved into two well-supported clades. Clade A includes *H. perrieri* and representatives of the Malagasy genera *Beguea* Capuron, *Chouxia*, *Conchopetalum* and *Pseudopteris* along with members of the East African/Malagasy genus *Macphersonia* Blume. The clade B comprises the Malagasy genus *Plagioscyphus* Radlk. and the South African species *Pappea capensis* Sond. & Harv. The relationships highlighted by these phylogenetic analyses are largely congruent with the views of Capuron (1969), with the notable exception of the placement of *Conchopetalum*, which had previously been included in Dodonaeoideae by Radlkofer (1878, 1933), an interpretation accepted by Capuron (1969). Since our results show a high level of congruence, only the ML total evidence tree will be discussed below because it contains the maximum amount of phylogenetic information (Fig. 1).

DISCUSSION

The phylogenetic results presented here clearly show that maintaining a broad circumscription of *Haplocoelum* to include the African taxa together with the Malagasy species currently recognized as *H. perrieri* would render the genus polyphyletic, and that in order to

maintain monophyly it will be necessary to exclude *H. perrieri* and place it in a new genus, which we describe below as *Gereaua* (Fig.1, Table 1; see Taxonomic treatment below). Morphologically, the new genus can be easily distinguished from the remaining members of *Haplocoelum* (as circumscribed to exclude *H. jubense* and *H. scassellatii*) by several inflorescence features in addition to the characters mentioned above: its staminate flowers are borne in racemules and its pistillate flowers are solitary (or more precisely, borne in reduced, 1-flowered inflorescences), whereas the African species have monomorphic inflorescences comprising fascicles of cymes (Table 2). Phylogenetic data strongly support the placement of the new genus in the *Macphersonia* group of Sapindaceae subfam. Sapindoideae, which comprises eight genera and ca. 30 species, most of which are endemic to Madagascar. This new understanding of relationships necessitates a re-assessment of the morphological features of the single species of *Gereaua*, which until now had always been considered within the context of African *Haplocoelum* and its presumed relatives (Fig. 1). Here we explore the relationships between *Gereaua* and the other genera comprised in the *Macphersonia* group in light of morphological characters (Table 3), the phylogenetic framework provided by Buerki et al. (2009), and the expanded analysis performed for the present study.

Within Clade A, *Beguea* is morphologically distinguished from *Gereaua* by its long inflorescence (exceeding 10 cm), apetalous corolla, 3-locular ovary with 3 well developed stigmas, glabrous fruit and absence of pseudostipules (Table 3). *Chouxia* (recently revised by Schatz et al. 1999) is cauliflorous, with long paniculate inflorescences, glabrous fruit and leaves twice as long as those of *Gereaua* (Table 3). The East African/Malagasy genus *Macphersonia* expresses the greatest range of morphological variation in the group, especially in leaf morphology and inflorescence structure, and is in need of taxonomic revision. Nonetheless, the inclusion within the molecular phylogeny of the somewhat anomalous species *M. chapelieri*, which shares with *G. perrieri* the presence of pinnate leaves and a golden indumentum on the fruit, confirmed its relationships with the previously sampled members of *Macphersonia*, which have bipinnate leaves and glabrous fruits.

Moreover, *Macphersonia* differs from *Gereaua* by the length of the inflorescence (between 10-30 cm long in *Macphersonia*) and the 2-3 stigmatic lines (Table 3). The leaves of *Pseudopteris* are very similar to those of *Gereaua* (e.g., length of the leaf and number of leaflets), but pseudostipules are absent and the leaflets of *Pseudopteris* have an acuminate apex. Moreover, *Pseudopteris* differs from the new genus by its racemose thyrses, 5-lobed disc, and 2-3 locular ovary producing 2-3 seeds (Table 3).

Within Clade B, the South African genus *Pappea* differs from *Gereaua* by its 3-locular ovary and simple leaves, and the geographic distribution of its single species (Table 3). *Plagioscyphus* is distinguished from the new genus by its ramiflorous or cauliflorous habit, and in having a developed petal scale, a highly specialized disc, 3-locular ovary, and leaves >15 cm lacking pseudostipules (Table 3).

Among the Malagasy genera of Sapindaceae, two were not included in our molecular phylogenetic analyses (*Camptolepis* Radlk. and *Tsingya* Capuron), but since they were regarded by Capuron (1969) as possibly related to *G. perrieri*, they must also be considered. The African/Malagasy genus *Camptolepis* can be easily distinguished from *Gereaua* by its ramiflorous habit, with short thyrses, 3-locular ovary with 3 stigmatic lines, glabrous fruit and absence of pseudostipules (Table 3). The monotypic Malagasy genus *Tsingya*, only known from the type material (in flower), differs from *Gereaua* by its axillary racemose thyrse inflorescences, apetalous flowers, 3-locular ovary with 3 stigmatic lines and large leaves (>15 cm long) without pseudostipules (Table 3). The fruits of *Tsingya* are unknown, and as no material is available for molecular analysis, its phylogenetic position can not be assessed, although in many ways it resembles *Beguea* and might simply represent an atypical member of that genus (G.E. Schatz and P.P. Lowry II, unpubl. data). On morphological grounds, *Camptolepis* and *Tsingya* would appear to be best placed in the *Macphersonia* group, but molecular analysis would be desirable to confirm this.

The present study confirms the preliminary conclusions of Buerki et al. (2009) and the earlier ideas expressed by Capuron (1969) and Davies and Verdcourt (1998) that the Malagasy

species long recognized as *Haplocoelum perrieri* should be placed in a separate monotypic genus. This contribution to the systematics of Sapindaceae constitutes the first part of a broader study that aims to identify and circumscribe taxonomic entities within the family that are consistent with results from analyses based on molecular and morphological data and building on the informal classification proposed by Buerki et al. (2009). In Africa, expanded sampling of the genus *Haplocoelum* will be needed to identify morphological characters that support the possible segregation of *H. gallense* from the rest of the genus, as well as other features that clarify limits among the central African members of *Haplocoelum*. In Madagascar, additional investigation of the *Macphersonia* group is required (including of *Camptolepis* and *Tsingya*) in order to provide improved generic circumscriptions and a modern phylogenetic framework for taxonomic revisions needed for the Catalogue of Vascular Plants of Madagascar project (<http://www.efloras.org/madagascar>) that are consistent with a robust phylogenetic framework. Moreover, the *Macphersonia* group, which has fewer than 50 species but exhibits significant morphological diversity and includes taxa with dehiscent fruits lacking an arillode (*Conchopetalum*) along with others that have indehiscent fruits and fleshy arillodes (e.g., *Plagioscyphus* and *Macphersonia*), presents a potentially intriguing opportunity to investigate evolution and diversification in the southwest Indian Ocean, incorporating data on a wide range of characters (e.g., chromosome numbers, anatomy, biochemistry and floral development) along with focused molecular phylogenetic research using broader sampling in order to identify synapomorphies for the group and to clarify its position within a revised classification system for Sapindaceae.

TAXONOMIC TREATMENT

Gereaua Buerki & Callm. **gen. nov. INED** –TYPE: *Gereaua perrieri* (Capuron) Buerki & Callm. INED [= *Haplocoelum perrieri* Capuron].

Haplocoelum Radlk. sect. *Cardiophyllariopsis* Capuron, Mém. Mus. Nat. Hist. Série B, 19: 133 (1969).

Hoc genus inter genera madagascariensia Sapindacearum quoad folia paripinnata foliolis basalibus in pseudostipulas reductis ad Chouxiam Capuron et Macphersoniam Blume maxime accedit, sed ab eis foliolis ad apicem retusis, inflorescentiis brevioribus sexualiter dimorphis (staminatis racemulosis, pistillatis reductis unifloris), floribus unisexualibus sed pistillatis propter staminodia bene evoluta ut videtur bisexualibus, ovario semper biloculari, stigmate bilobo (in pistillodio floris staminati reducto), fructu sphaerico tomento aureo brevi dense oblecto atque semine semper solitario distinguitur.

Evergreen, dioecious trees. Leaves alternate, paripinnately compound, exstipulate; proximal leaflets bract-like or forming pseudostipules at the base of the petiole. Flowers actinomorphic, functionally unisexual (those of female plants morphologically perfect). Inflorescences axillary, staminate flowers borne in racemules, pistillate flowers borne in reduced, 1-flowered inflorescences (thus appearing solitary). Sepals, 4 or 5, equal, free, narrowly imbricate; petals 4 or 5 (or 6), auriculate basally; disc annular; stamens (4 or) 5 to 7, inserted on the disc, long exerted, base of the filaments densely pubescent (stamens slightly reduced in pistillate flowers, lacking pollen. This observation suggests a functional dioecy of this genus), anthers basifixed; ovary 2-carpellate, stipitate, style with 2 stigmatic lobes (reduced to pistillodes in staminate flowers). Fruit nut-like but eventually splitting from the apex to reveal a single arilate seed, exocarp tomentose, (1 or) 2 locular, with a persistent calyx; seed ellipsoid, arillode dorsally or distally split, translucent, fleshy. Fig. 2.

Gereaua perrieri (Capuron) Buerki & Callm., comb. nov. INED ***Haplocoelum perrieri***

Capuron, Mém. Mus. Nat. Hist., Série B, 19: 133 (1969), pl. 26. –TYPE: MADAGASCAR.

Province Toamasina: Centre: vestige de forêt au PK 100 de la route Tananarive-Moramanga, près du village d'Ankarahara, rive gauche du Mangoro, 0-800 m, [18°54'S, 48°09'E], 25 Oct. 1963, *Service Forestier* 22732 (holotype: P! 3 sheets; isotypes: MO!, TEF!), pistillate flowers.

Tree, 7-15 m tall; stems lenticellate, initially pubescent with brown trichomes, eventually glabrescent. Leaves paripinnate, (3-)4-6(-15) cm long, petiole short (1-2 mm), rachis puberulent, slightly winged; leaflets (2 or) 3 to 5 (to 7) pairs, subopposite to alternate (rarely opposite), sessile, apical pair 2-3(-7) x 1-1.5(-2.5) cm, proximal pair (excluding the basal pseudostipules) 1-1.5(-3) x 0.5-0.8(-1.5) cm, elliptic, base acuminate, slightly asymmetrical, margin entire, apex retuse. Staminate racemules 10-30 mm long, with 5 to 10 (to 15) flowers, pubescent with golden trichomes; pistillate inflorescence 10 mm long, pubescent with golden trichomes. Bracts caducous, bracteoles 1.5-2 mm long, ovate-triangular, densely pubescent on adaxial surface, margin ciliate. Pedicel of staminate and pistillate flowers 0.3-0.8 mm long (accrescent in fruit). Sepals 4 or 5, 2.5-3 x 1.5-2 mm, triangular to ovate-triangular, densely pubescent on adaxial surface, puberulent on the abaxial surface, margin ciliate; petals 4 or 5 (or 6), 1.5-2 x 1-1.5 mm, obovate to suborbicular, adaxial surface of the petal laterally folded, pubescent and ciliate; disc annular, lobate, glabrous, 1.5 mm in diameter; stamens (4 or) 5 to 7, ca. 3-3.5 mm long, base of the filament densely pubescent, anther glabrous, 1 mm long; ovary 2-carpellate (vestigial in staminate flowers), ca. 1 mm long, style short, with 2 stigmatic lobes. Fruit spherical, 15-20 mm in diameter, densely pubescent with a short tomentose golden indument, 2-locular (one abortive), stigmas sometimes persistent at the apex of the fruit; seed 1, ellipsoid, flattened, ca. 10-12 x 5-6 x 6-8 mm. Fig. 2.

Material examined– MADAGASCAR. Prov. Antananarivo: Commune d'Ambongamarina, forêt d'Ampatsakandrainivavy, 1356 m, 18°17'6"S, 47°54'30"E, 10 Mar. 2000, fr., *R. Randrianaivo et al. 521* (MO, P, TAN). Mahasoa, Vallée de la Mandraka PK 70, [900 m], [18°56'S, 47°57'E], 8 Nov. 1957, staminate fl., *Service Forestier 18411* (P, TEF). Prov. Antsiranana: Forêt littoral sur sable d'Anaborano près du Lac Sahaka, 25 m, 13°04'42"S, 49°54'13"E, 2 Nov. 2002, buds, *J. Rabenantoandro et al. 1065* (MO, P, TAN). Réserve Spéciale de Manongarivo, est d'Ankaramibe, Bekolosy, 600-800 m, 14°03'05"S, 48°17'07"E, 10 Dec. 1993, fr., *L. Rakotomalala & Fernand 79* (K, MO, P). Prov. Fianarantsoa: Parc National Ranomafana, 1100 m, 21°15'S, 47°27'E, 11-15 Nov. 1991, fr., *Malcomber 1040* (MO, P). Parc National Ranomafana, 900-1100 m, 21°15'S, 47°23'E, 16 Nov. 1994, fr., *J. Randrianasolo & V. Bernardin 187* (P, TEF). Andrambovato, Tolongoina, Fort-Carnot, [21°33'30"S, 47°31'30"E], 19-21 Jan. 1955, imm. fr., *Service Forestier 11586* (P, TEF). Midongy Atsimo, [23°35'00"S, 47°01'00"E], 25 Aug. 1955, pistillate fl., *Service Forestier 13929* (P, TEF). Andrambovato, Fort-Carnot, [21°51'30"S, 47°26'30"E], 15 Mar. 1955, imm. fr., *Service Forestier 14678* (P, TEF). Andrambovato, Tolongoina, Fort-Carnot, 800-1100 m, [21°33'30"S, 47°31'30"E], 1951, st., *Service Forestier 72B-R-230* (P). Andrambovato, Fort-Carnot, [800-1100 m], [21°33'30"S, 47°31'30"E], 26 Sep. 1956, pistillate fl., *Service Forestier 94-R-230* (P, TEF). Prov. Toamasina: Ambatoharanana, près d'Antsevabe, 1000 m, [17°58'S, 48°32'E], 7 Mar. 1955, fr., *Cours 4116* (MO, P). ca. 15 air-km NE of Moramanga, ca. 11 km E of Antanambao, Makaranana, 1078 m, 18°49'30"S, 48°20'47"E, 24 Feb. 2001, fr., *P.-J. Rakotomalaza et al. 1165* (K, MO, P, TAN). Ambohitsitondroinan'i Mahalevona, environ de la baie d'Antongil, 200 m, [15°25'S, 49°58'E], s.d., fr., *Service Forestier 8881* (G, MO, P, TEF). Ankarahara, PK. 100 de la route Antananarivo-Moramanga, [18°54'S, 48°09'E], 25 Oct. 1963, staminate fl., *Service Forestier 22731* (K, P, TEF). Ankarahara, PK 100 de la route Tana-Moramanga, [18°54'S, 48°09'E], 22 Nov. 1967, imm. fr., *Service Forestier 22912* (K, MO, P, TEF). Ankarahara, PK 100 de la route Tana-Moramanga, [18°54'S, 48°09'E], 5 Jan. 1964, fr., *Service Forestier 23200* (G,

MO, P, TEF). Prov. Tulear: Massif de Lavasoa, [25°05'S, 46°44'30E], 24-25 Jan. 1955, fr., *Service Forestier 11820* (K, P, TEF).

Remarks– Several specimens annotated or cited by Capuron (1969) are excluded here and are referred to *Doratoxylon* or *Pteropteris*. *Perrier de la Bâthie 4457* cannot be determined with certainty, and might represent a second species of *Gereaua*, although the material is inadequate to make a definitive assessment.

Distribution and habitat– *Gereaua perrieri* is widely distributed in humid and subhumid forests of eastern and north-western Madagascar. It is known to occur in littoral forests of the northeastern coast (at sea level around Lake Sahaka) and at mid-elevation forests in the Manongarivo massif, and it ranges along the eastern escarpment generally in mid- to high-elevation forest (up to ca. 1350 m above sea level at Anjozorobe) from Antongil Bay to near Fort-Dauphin. Fig. 3.

Etymology– Our new genus is named in honor of Roy Gereau, Assistant Curator at the Missouri Botanical Garden, who has long held an interest in Sapindaceae and who encouraged the authors to investigate the relationships of *Haplocoelum perrieri*, which he regarded as so distinct from the African species that he questioned whether this taxon could belong to the same genus.

Conservation status– With an EOO of 163,665 km², an AOO of 126 km², and 13 subpopulations, 4 of which are situated within protected areas (Manongarivo, Masoala, Midongy du Sud, Ranomafana), *Gereaua perrieri* is assigned a preliminary status of Least Concerned (LC) based on the IUCN Red List criteria (IUCN, 2001).

ACKNOWLEDGEMENTS. We thank Professors Philippe Küpfer and Giorgina Bernasconi of the University of Neuchâtel (Switzerland) for help and financial support provided to the first author; Nadir Alvarez (University of Neuchâtel, Switzerland) and Félix Forest (Royal Botanic Gardens, KEW, U.K.) for their suggestions and permanent support to the first author; Roger Lala Andriamiarisoa for his fine illustrations; Roy Gereau for valuable discussions and for

assistance with the Latin diagnosis; the Malagasy staff of the Missouri Botanical Garden (MBG) based in Antananarivo, especially Ornella Randriambololomamonjy, for help finding specimens in Madagascar; and the curators of the following herbaria for making material available for this study: G, K, MO, P, TAN and TEF. Work in Madagascar was conducted under collaborative agreements between MBG and both the Parc Botanique et Zoologique de Tsimbazaza and the Direction de la Recherche Forestière et Piscicole, FOFIFA (Ministère de l'Education), Antananarivo, Madagascar. We gratefully acknowledge courtesies extended by the Government of Madagascar (Direction Générale de la Gestion des Ressources Forestières) and by the Association Nationale pour la Gestion des Aires Protégées. Research conducted by PBP, PPL and MWC was supported by a grant from the U.S. National Science Foundation (0743355; PPL Co-PI).

LITERATURE CITED

- Angiosperm Phylogeny Group (APG II) 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APGII. *Botanical Journal of the Linnean Society* 141: 399–436.
- Buerki, S., F. Forest, P. Acevedo-Rodríguez, M.W. Callmander, J.A.A. Nylander, M. Harrington, I. Sanmartín, P. Küpfer, and N. Alvarez. 2009. Plastid and nuclear DNA markers reveal intricate relationships at subfamilial and tribal levels in the soapberry family (Sapindaceae). *Molecular Phylogenetics and Evolution* 51: 238–258.
- Bullock, A.A. 1931. XLVII.— The genus *Haplocoelum*. *Bulletin of Miscellaneous information, Royal Botanic Gardens, Kew* 7: 353–357.
- Callmander, M.W., G.E. Schatz, M.O. Laivao, J. Raharimampionona, S. Andriambololona, T. Consiglio, T. Raminosa, and P.P. Lowry II. 2007. Application of IUCN Red List criteria and assessment of Priority Areas for Plant Conservation in Madagascar: rare and threatened Pandanaceae indicate new sites in need of protection. *Oryx* 42: 168–176.

- Capuron, R. 1969. Révision des Sapindacées de Madagascar et des Comores. *Mémoires du Muséum National d'Histoire Naturelle, Série B., Botanique* 19: 1–189.
- Cornet, A. 1974. Essai de cartographie bioclimatique à Madagascar. Notice Explicative. *ORSTOM* 1–55.
- Davies, F.G. and B. Verdcourt. 1998. *Flora of Tropical East Africa: Sapindaceae*. Rotterdam.
- ESRI. 2000. ArcView 3.3. Redlands, California.
- Friis, I. 1984. Additional notes on Somalian Sapindaceae. *Kew Bulletin* 39: 779–783.
- Friis, I. and K. Vollesen. 1985. The identity of *Hypseloderma jubense* (Sapindaceae). *Kew Bulletin* 40: 859–861.
- 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.
- Holmgren, P.K., N.H. Holmgren, and L.C. Barnett. 1990. *Index Herbariorum. Part I: The herbaria of the World*. New York Botanical Garden: New York.
- IUCN Red List Criteria 2001. IUCN Red List Categories and Criteria: Version 3.1. IUCN, Gland, Switzerland and Cambridge, UK, IUCN Species Survival Commission.
- Kluge, A.G. 1989. A Concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Systematic Zoology* 38: 7–25.
- Lebrun, J.-P. and A.L. Stork. 1992. Enumération des plantes à fleurs d'Afrique tropicale, vol. II: Chrysobalanaceae à Apiaceae. Conservatoire et Jardin botaniques de la Ville de Genève, 257 pages.
- Lebrun, J.-P. and A.L. Stork. 1997. Enumération des plantes à fleurs d'Afrique tropicale, vol. IV: Gamopétales: Clethraceae à Lamiaceae. Conservatoire et Jardin botaniques de la Ville de Genève, 712 pages.
- Muellner, A.N., D.D. Vassiliades, and S.S. Renner. 2007. Placing Biebersteiniaceae, a herbaceous clade of Sapindales, in a temporal and geographic context. *Plant Systematics and Evolution* 266: 233–252.

- Müller, J. and P.W. Leenhouts. 1976. A general survey of pollen types in Sapindaceae in relation to taxonomy. Pp. 407–445 in *The evolutionary significance of the exine*, eds. I.K. Ferguson and J. Müller. London, Academic Press.
- Munzner, T., F. Guimbretiere, S. Tasiran, L. Zhang, and Y. Zhou. 2003. TreeJuxtaposer: scalable tree comparison using Focus+ context with guaranteed visibility. Pp. 453–462 in SIGGRAPH.
- Radlkofer, L. 1878. Über *Sapindus* und damit in Zusammenhang stehende Pflanzen. *Sitzungsber. Math.-Phys. Cl. Kön. Bayer. Akad. Wiss. München* 8: 334–338.
- Radlkofer, L. 1933. Sapindaceae. in *Das Pflanzenreich IV*, 165 (Heft 98a-h), ed. A. Engler. Leipzig, Verlag von Wilhelm Engelmann.
- 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 nearly complete familial analysis based on *rbcL* gene sequences. *Kew Bulletin* 55: 257–309.
- Schatz, G.E. 2000. Endemism in the Malagasy tree flora. Pp. 1-11 in *Diversity and Endemism in Madagascar*, eds. W.R. Lourenço and S.M. Goodman. Mémoires de la Société de Biogéographie, MNHN, OSRTOM, Paris.
- Schatz, G.E., R.E. Gereau, and P.P. Lowry II. 1999. A revision of the Malagasy endemic genus *Chouxia* Capuron (Sapindaceae). *Adansonia, série*. 3 21: 51–62.
- Schatz, G.E. and M. Lescot. 2005. Gazetteer to Malagasy Botanical Collecting Localities. Missouri Botanical Garden website (<http://www.mobot.org/MOBOT/Research/madagascar/gazetteer/>).
- Soltis, D.E., P.S. Soltis, M.W. Chase, M.E. Mort, T.D. 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, *rbcL*, and *atpB* sequences. *Botanical Journal of the Linnean Society* 133: 381–461.

Thorne, R.F. 2007. An update classification of the class Magnoliopsida (“Angiospermae”).
Botanical Review 73: 67–182.

Yang, Z.H. 1993. Maximum-likelihood-estimation of phylogeny from DNA-sequences when
substitution rates differ over sites. *Molecular Biology and Evolution* 10: 1396–1401.

TABLES

Table 1. Informal classification of the Sapindaceae sensu lato based on molecular and morphological characters following Buerki et al. (in press). The position of the four sequenced species of *Haplocoelum* within the informal classification is indicated between parentheses (see text for more details).

Subfamilies

1. Xanthoceroideae

Xanthoceras sorbifolia

2. Hippocastanoideae

2.1. Acer group

2.2. Aesculus group

3. Dodonaeoideae

3.1. Doratoxylon group

3.2. Dodonaea group

4. Sapindoideae

4.1. Delavaya group

4.2. Koelreuteria group

4.3. Schleichera group

4.4. Litchi group (incl. *H. gallense*)

4.5. Macphersonia group (incl. *H. perrieri*)

4.6. Cupania group

4.7. Tristiropsis group

4.8. Blomia group (incl. *H. foliosum* and *H. inopleum*)

4.9. Melicoccus group

4.10. Paullinia group

Table 2. Comparison of the Malagasy *Haplocoelum perrieri* Capuron with the other African species of the genus.

	<i>H. perrieri</i>	<i>Haplocoelum</i>
Distribution	Madagascar	Africa
No. species	1 sp.	4-5 sp.
Phylogenetic analyses	Macphersonia group	Sapindus and Blomia groups
Inflorescence		
- type	racemules (♀ 1-flowered)	fasicule of cymes
- sexual dimorphism	yes	no
Flowers		
- petals	4-5	0
- ovary	2-locular	3-locular
- male flower	rudimentary ovary	ovary well developed
- stamens	glabrous	pubescent
Fruits		
- indument	yes	no

Table 3. Comparison of the genus *Gereaua* Buerki & Callm. with the other genera belonging to the *Macphersonia* group. The last two genera (*Camptolepis* and *Tsingya*) were not included in the phylogenetic analyses (see text for more details).

	<i>Gereaua</i>	<i>Beguea</i>	<i>Chouxia</i>	<i>Conchopetalum</i>	<i>Macphersonia</i>	<i>Pseudopteris</i>	<i>Pappea</i>	<i>Plagioscyphus</i>	<i>Camptolepis</i>	<i>Tsingya</i>
Distribution	Madagascar	Madagascar	Madagascar	Madagascar	Madagascar/Africa	Madagascar	Africa	Madagascar	Madagascar/Africa	Madagascar
No. species	1 sp.	1 sp.	6 sp.	2 sp.	about 8 sp.	3 sp.	1 to 4 sp.	about 10 sp.	4 sp.	1 sp.
Leaves										
- type	paripinnate	paripinnate	paripinnate	paripinnate	biparipinnate (1 sp. paripinnate)	paripinnate	simple	paripinnate	paripinnate	paripinnate
- length	<15 cm generally	>15 cm	>30 cm	>15 cm	10 (<i>M. gracilis</i> et <i>M. hildebrandtii</i>) >15 cm	>15 cm	>5 cm	>15 cm	>15 cm	>15 cm
- apex of leaflets	4-6 cm retuse	apiculate	apiculate	rounded to acute	acuminate	acuminate	-	acuminate to caudate	acuminate	acuminate
- pseudostipules	yes	no	yes (rarely absent)	no	yes	no	-	no	no	no
Inflorescence										
- breeding system	dioecious	dioecious	dioecious or monoecious		dioecious	polygamous	dioecious	polygamous	dioecious	monoecious
- type	racemes (♀ 1-flowered)	racemes or thyrses	paniculate	fasciculate, pseudo-umbelliform	racemose or spicate, or rarely paniculate thyrses	racemose thyrses	racemose or paniculate thyrses	racemose thyrses	short thyrses	racemose thyrses
- position	axillary	axillary	cauliflorous	axillary or ramiflorous	axillary	axillary	axillary	ramiflorous or cauliflorous	ramiflorous	axillary
- length	<3 cm	>10cm	>15cm		>10-30 cm	>5 cm	>10 cm	>10 cm	>3-5 cm	>10 cm
- sexual dimorphism	yes	no	no	no	no	no	no	no	no	no
Flowers										
- petals	4-5	0	5	5	5	5	(4)5(6)	4-5	5	0
- petal scale	no	-	yes	no	no	no	yes	yes	yes	-
- disc shape	annular	annular-lobed	annular	annular, double, the central rim sometimes tubular resembling an androgynophore	annular, dish-shaped	5-lobed	annular	specialized	annular-lobed or cup-shaped	annular
- no. locule	2	3	(2-)3	3	(2-)3	2-3	3	2-3	3	3
- stigma	2-lobed	3 stigmatic branches	2-3 stigmatic branches	stigma punctiform	2-3 stigmatic lines	2-3 stigmatic zones along edges	3-sub-lobed	2-3 stigmatic lines	3 stigmatic lines	3 stigmatic lines
Fruits										
- arillode	yes	yes	yes	no (sarcotesta around the hilum)	yes	yes	yes	yes	yes	-
- indument	yes	no	no	no	yes and no	no	yes	yes	no	-
- no. seeds	1	1	1-3	3-6	(1-)2(-3)	2-3	1	1-3	1-3	-

FIGURES LEGENDS

Fig. 1. Maximum likelihood phylogenetic tree for Sapindaceae sensu lato inferred from eight nuclear and plastid nucleotide sequences. Bootstrap support values are indicated above branches. The revised infrafamilial classification based on molecular and morphological characters proposed by Buerki et al. (in press) is indicated in grey (see Table 1).

Abbreviations: Nephelieae (NEP) and Schleichereae (SCH). See text for additional details.

Fig. 2. *Gereaua perrieri* Buerki & Callm.: A. fruit showing the indumentum: scale=1 cm. B. staminate inflorescence on branch: scale=2 cm. C. detail of petal from staminate flower: scale=1 mm. D. staminate flower: scale=1 mm. E. pistillate flower: scale=1 mm. F. detail of petal from pistillate flower: scale=1 mm. G. pistillate inflorescence on branch: scale=2 cm. A. *Service Forestier 23200*; B-D, *Service Forestier 22732*; E-G, *Service Forestier 22731*.

Fig. 3. Distribution of *Gereaua perrieri* (Capuron) Buerki & Callm. mapped on the bioclimatic zones of Madagascar (after Cornet 1974; see Schatz 2000).

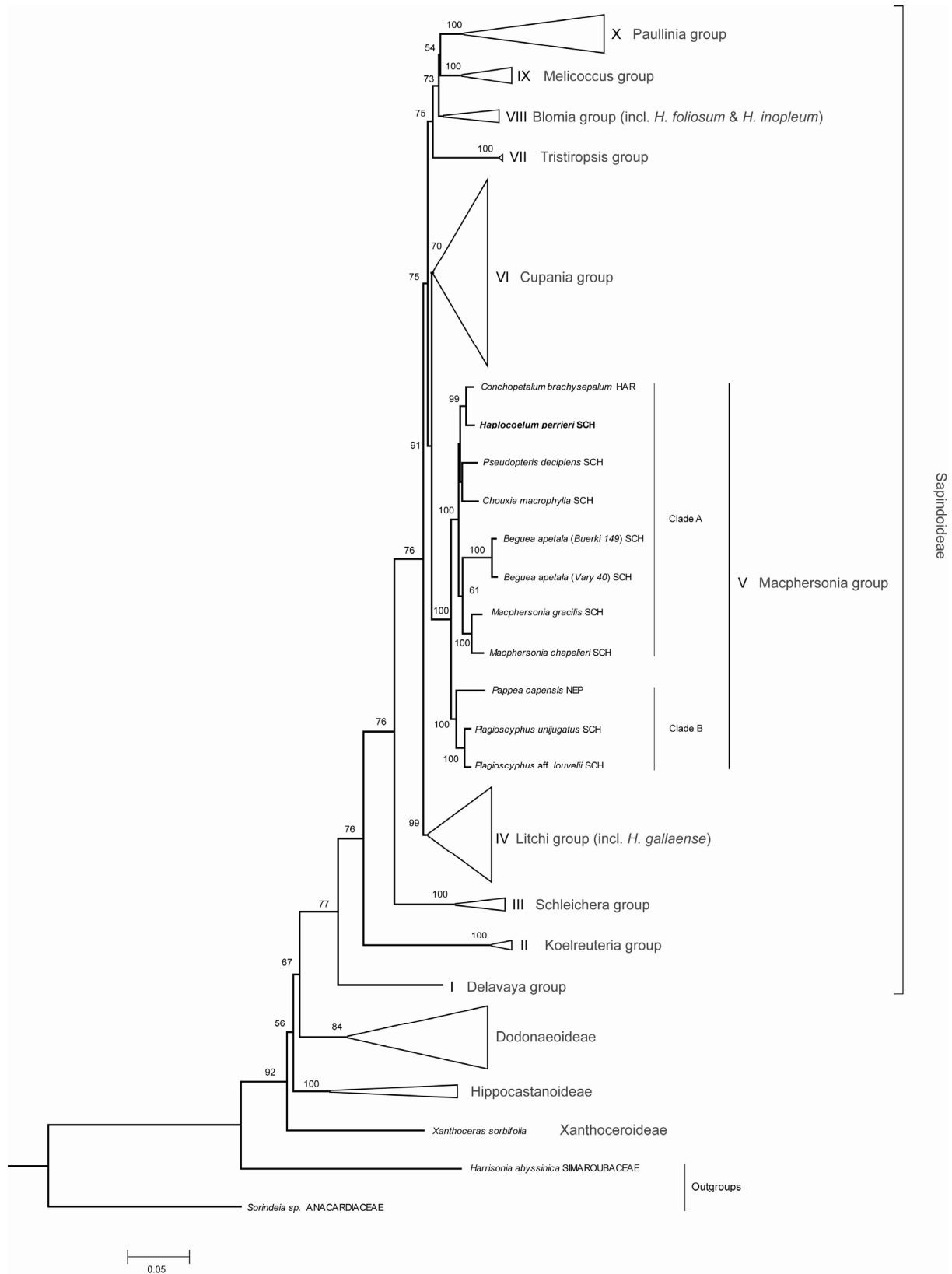


Figure 1

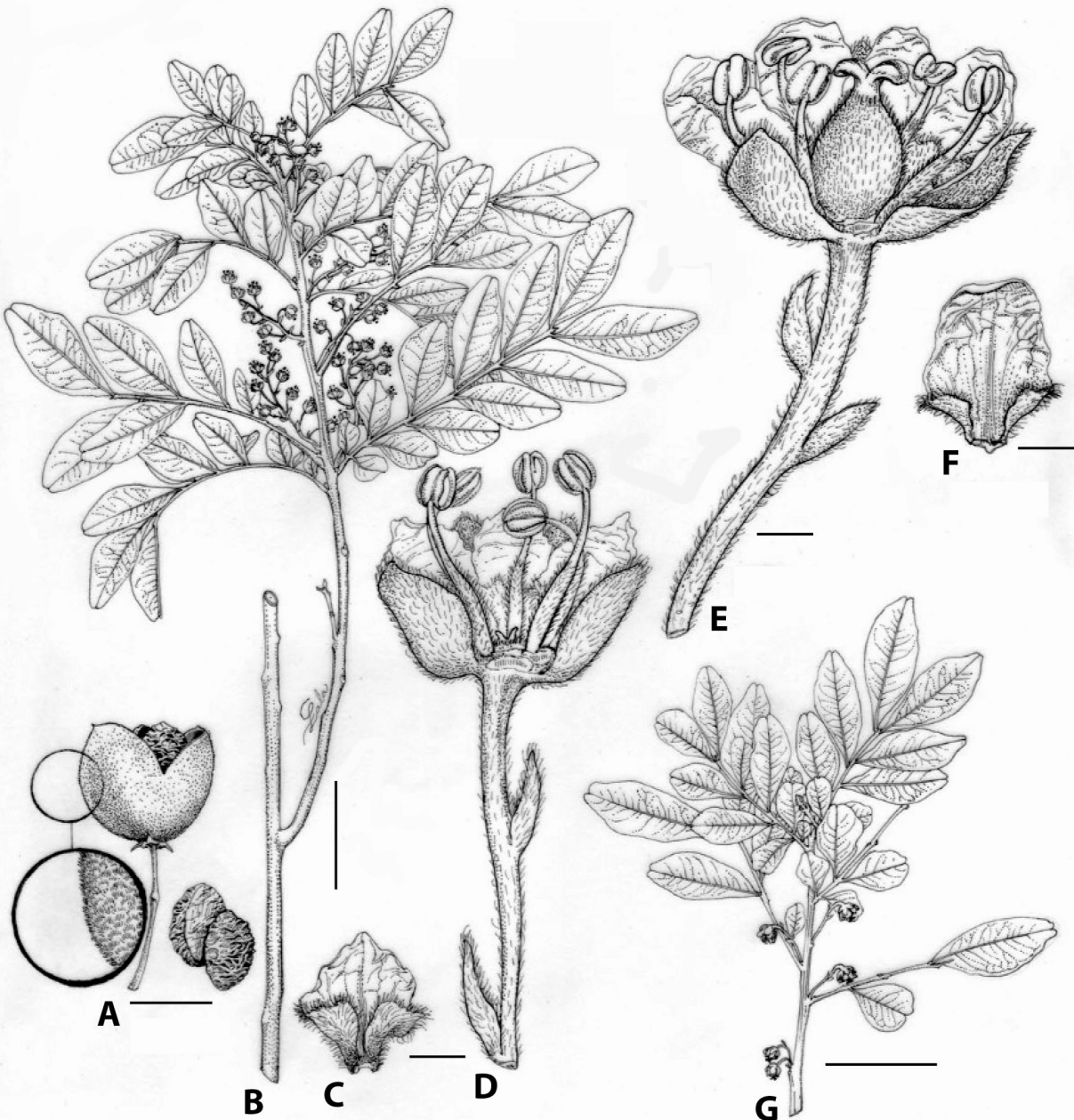


Figure 2

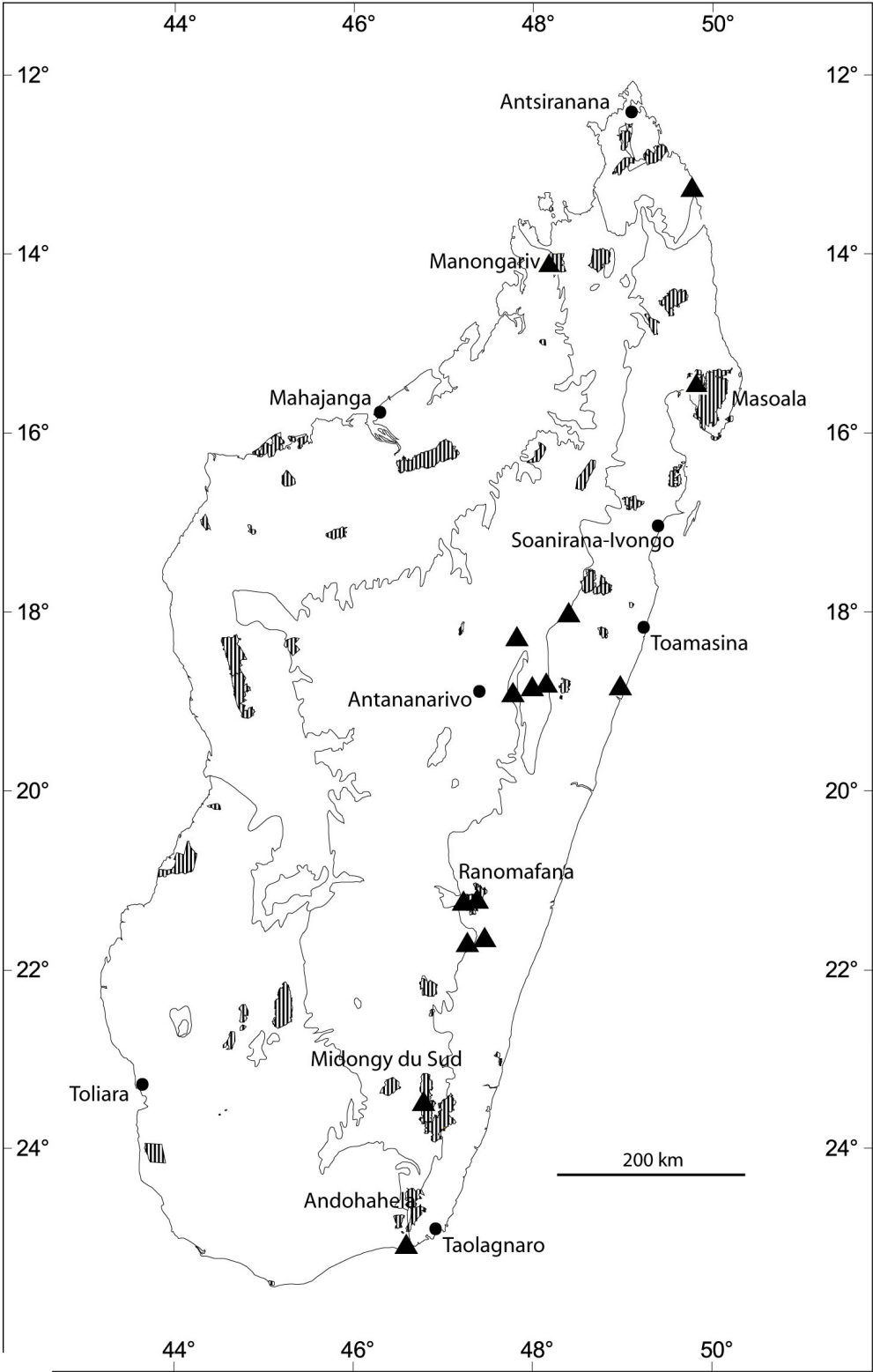


Figure 3

**Chapter five: A synoptic revision of the genus *Lepisanthes*
Blume (Sapindaceae) in Madagascar**

Sven Buerki, Martin W. Callmander, Porter P. Lowry II and Peter B. Phillipson

Submitted to *Adansonia*

A synoptic revision of the genus *Lepisanthes* Blume (Sapindaceae) in Madagascar

Sven Buerki

Institute of Biology, University of Neuchâtel, Rue Emile-Argand 11, 2009 Neuchâtel,
Switzerland.

sven.buerki@unine.ch

Martin W. Callmander

Missouri Botanical Garden, P.O. Box 299, 63166-0299, St. Louis, MO, U.S.A. and
Conservatoire et Jardin botaniques de la ville de Genève, ch. de l'Impératrice 1, 1292
Chambésy, Switzerland.

martin.callmander@mobot-mg.org

Porter P. Lowry II

Missouri Botanical Garden, P.O. Box 299, St. Louis, MO, 63166-0299, U.S.A. and
Muséum national d'Histoire Naturelle, Département Systématique et Évolution, Case Postale
39, 57 rue Cuvier, F-75231 Paris cedex 05, France.

pete.lowry@mobot.org

Peter B. Phillipson

Missouri Botanical Garden, P.O. Box 299, St. Louis, MO, 63166-0299, U.S.A. and
Muséum national d'Histoire Naturelle, Département Systématique et Évolution (UMR 7205),
Case Postale 39, 57 rue Cuvier, F-75231 Paris cedex 05, France.

peter.phillipson@mobot.org

ABSTRACT

A taxonomic revision of the genus *Lepisanthes* Blume in Madagascar is presented. Three species are recognized based on an analysis of morphological characters in combination with eco-geographic parameters. Two infraspecific taxa recognized previously in *Aphania senegalensis* (Juss. ex Poir.) Radlk. [= *Lepisanthes senegalensis* (Juss. ex Poir.) Leenh.] are raised to the species level, viz. *L. chrysotricha* (Capuron) Buerki, Callm. & Lowry INED and *L. perrieri* (Capuron) Buerki, Callm. & Lowry INED. A third species from northwestern Madagascar is described as new, *L. sambiranensis* Buerki, Callm. & Lowry INED. An identification key to the Malagasy species of *Lepisanthes* is presented as well as preliminary assessments of the conservation status for each species.

KEY-WORDS: *Aphania*, *Lepisanthes*, Sapindaceae, Madagascar, new species.

RESUME

Une révision taxonomique du genre *Lepisanthes* Blume à Madagascar est présentée. Trois espèces sont reconnues sur la base de leurs caractères morphologiques, écologiques et géographiques. Deux taxons infraspécifiques reconnus auparavant dans *Aphania senegalensis* (Juss. ex Poir.) Radlk. [= *Lepisanthes senegalensis* (Juss. ex Poir.) Leenh.] sont élevés au rang d'espèce dans le genre *Lepisanthes*: *L. chrysotricha* (Capuron) Buerki, Callm. & Lowry INED et *L. perrieri* (Capuron) Buerki, Callm. & Lowry INED. Une troisième espèce du Nord-Ouest de Madagascar est décrite comme nouvelle: *L. sambiranensis* Buerki, Callm. & Lowry INED. Une clé de détermination des espèces malgaches du genre *Lepisanthes* est présentée ainsi qu'une évaluation préliminaire du statut de conservation de chaque espèce.

MOTS CLES: *Aphania*, *Lepisanthes*, Sapindaceae, Madagascar, espèce nouvelle.

INTRODUCTION

As currently circumscribed the genus *Lepisanthes* (Sapindaceae) comprises c. 25 species in tropical Africa, Madagascar, and southern and southeastern (SE) Asia from Sri Lanka to Hainan, Malesia, and northwestern Australia (Leenhouts 1969; Adema et al. 1994). In the most recent worldwide revision of the genus, Leenhouts (1969) adopted a broad species concept for *L. senegalensis* (Juss. ex. Poir.) Leenh. encompassing material previously placed in the genera *Aphania* Radlk. (*sensu* Radlkofer 1933, with the exception of *A. dictyophylla* Radlk.), *Manongarivea* Choux and *Sapindopsis* F.C. How & C.N. Ho (now referred to as *Howethoa* Rauschert), including a total of 32 heterotypic synonyms in this single taxon. As defined by Leenhouts (1969), *L. senegalensis* ranges widely from Africa (2 synonyms) and Madagascar (1 synonym) to SE Asia and New Guinea (29 synonyms). In his treatment, Leenhouts (1969) mentioned that some elements involved in this complex (especially in West Malesia) could deserve recognition at the species or subspecies rank, stating that they “might even be genetically isolated”. Moreover, he acknowledged that this broad species concept might be difficult to accept, especially for botanists working on the flora of a restricted region. Nonetheless, he justified this potential over-lumping by claiming that the SE Asian specimens form a coherent group recognized by the “reduction of the leaf to 1-jugate, often with a short to very short petiole, pseudoterminal or possibly sometimes truly terminal inflorescences with subsessile cymes”.

The basionym of Leenhouts’s (1969) broadly defined species was first published in Lamarck’s *Encyclopédie Botanique* (1805) as *Sapindus senegalensis* Juss. ex Poir. on the basis of two syntypes from Senegal (*Adanson s.n.* and *Adanson & Geoffrey s.n.*). However, Leenhouts (1969) adopted his concept of *L. senegalensis* without ever having seen these collections, which are deposited at P-JU, nor did he examine any of the other type material in the Paris Herbarium, including that of *Manongarivea perrieri* from Madagascar. In the same year, while preparing a revision of Malagasy Sapindaceae, Capuron (1969), grappling with Leenhouts’s broad species concept, chose to recognize two infraspecific taxa within *Aphania senegalensis* (Juss. ex Poir.) Radlk. (= *Lepisanthes senegalensis sensu* Leenhouts 1969): *A.*

senegalensis subsp. *senegalensis* fa. *perrieri* (Choux) Capuron and *A. senegalensis* subsp. *chrysotricha* Capuron.

In the context of a worldwide phylogenetic study of Sapindaceae (Buerki et al. 2009) and a series of revisions of Malagasy members of the family (Schatz & al. 1999, in prep.; Buerki et al. in prep.; Callmander et al. in prep.), we have re-considered the taxonomic framework of *Lepisanthes* in Madagascar and propose here the following treatment, in which three species are recognized, one of which is described as new whereas the others are based on the infraspecific taxa recognized by Capuron (1969) in the genus *Aphania*.

MATERIAL AND METHODS

We have examined all the available material of *Lepisanthes* at the major herbaria with relevant collections, including those in Antananarivo (TAN & TEF), Geneva (G), Kew (K), Madrid (MA), St. Louis (MO) and Paris (P) (herbarium citations follow Holmgren et al. 1990). Historical collections lacking geographic coordinates were post-facto georeferenced as accurately as possible using the “Gazetteer to Malagasy Botanical Collecting Localities” (Schatz & Lescot 2005) and other sources (placed in square brackets in the citation of material examined in the taxonomic treatment that follows). Species distributions were mapped on the five bioclimatic zones of Madagascar (after Cornet 1974; adapted by Schatz 2000) using ESRI ArcView 3.3 software (2000). The conservation status of each species was assessed using the current IUCN Red List Criteria (2001). Calculations of the area of occupancy (AOO), extent of occurrence (EOO) and number of subpopulation were based on the methods presented in Callmander et al. (2007).

SYSTEMATICS

Genus *Lepisanthes* Blume

Bijdr. Fl. Ned. Ind. 5: 237 (1825). — Type: *L. montana* Blume.

Manongarivea Choux, *Compt. Rend. Hebd. Séances Acad. Sci.* 182: 713. (1926); *Mém. Acad. Malg.* 4: 36 (1927). — Lectotype (here designated): Madagascar, Prov. Mahajunga, Manongarivo (Ambongo), [16°16'S, 45°22'E], X.1909, fl., y fr., *Perrier de la Bâthie 1812* [P (P00624104); iso-, K (K000426188), P (P00624103), TAN].

Careful examination of the available material shows that the Malagasy collections of *Lepisanthes* differ from the African material (which corresponds to *L. senegalensis* sensu stricto) by generally having a glabrous 2-branched inflorescence (vs. a pubescent much-branched inflorescence in specimens from Africa). The only exception is the material assigned below to *L. chrysotricha*, which likewise has a pubescent inflorescence, but differs from the African representatives by its golden indument that is also present on the outer part surface the sepals and petals (vs. brown indument on the inflorescence and glabrous sepals and petals in Africa).

We recognize three well delimited species of *Lepisanthes* in Madagascar, one of which is new, whereas the two others correspond to the infraspecific taxa recognized by Capuron (1969), which we raise to the rank of species. Schatz (2001) provide a comprehensive diagnosis of the genus based on material from Madagascar.

A collection clearly assignable to *Lepisanthes* from the Masoala Peninsula in northeastern Madagascar (*Labat et al.* 3356, MO, P) does not match any of the species recognized here. It has a long inflorescence (up to 32 cm) and leaves with 3 pairs of leaflets that are discoloured and subcoriaceous. Moreover, this is the only Malagasy specimen of the genus known from humid evergreen forest on the east coast (all others were collected in dry to sub-humid forest on the north and western Madagascar). While this collection appears to represent a new species, adequate material is not yet available to describe it.

The monotypic genus *Manongarivea* was described by Choux (1926, with a supplementary description appearing in 1927) on the basis two syntypes (*Perrier de la Bâthie* 1744 and 1812). In the present treatment, however, these collections are regarded as belonging to two different species, *L. chrysotricha* and *L. perrieri*, respectively. Capuron (1969), by designating *Perrier de la Bâthie* 1744 as the type of his new taxon *Aphania senegalensis* subsp. *chrysotricha*, implicitly limited the taxon described by Choux to *Perrier de la Bâthie* 1812, and thereby lectotypified the name, although he did not explicitly indicate his intention to do so. Here we formalize this decision and refine it by designating one of the two sheets of *Perrier de la Bâthie* 1812 in the Paris herbarium as the lectotype.

KEY TO THE MALAGASY SPECIES OF *LEPISANTHES* BLUME

1. Inflorescence and outer surface of the sepals and petals golden pubescent; western Madagascar (Bemaraha and Causse de Kelifely, W of the Mahavavy River)
 1. *L. chrysotricha*
- Inflorescence and outer surface of the sepals and petals glabrous; northern and northwestern Madagascar 2
2. Leaflets 1 or 2 (or rarely 3) pairs, 7–10 (–15) cm long; petiole and rachis slender (1–1.5 mm diameter in dried material), combined length 0.5–2.5 (–11) cm; northern Madagascar (S of Majunga to Montagne d'Ambre, Ankarana AP and Daraina region); on limestone and basement rock 2. *L. perrieri*
- Leaflets (2 or 3 (or 4) pairs, 12–15 (–25) cm long; petiole and rachis stout (ca. 2 mm diameter in dried material), combined length 7–15 (–22) cm; northwestern Madagascar (Kalabenono and Manongarivo massifs); on sandstone
 3. *L. sambiranensis*

1. *Lepisanthes chrysotricha*

(Capuron) Buerki, Callm. & Lowry, comb. nov. and stat. nov. INED

Aphania senegalensis subsp. *chrysotricha* Capuron, *Mém. Mus. Nat. Hist., sér. B*, 19: 82, pl. 19 (1969). — Type: Madagascar, Prov. Mahajunga, forêt de Kasiza, sur le causse du Tampoketsa (Ambongo), rive gauche de la Mahavavy, [16°59'S, 45°47'E], X.1908, fl., y fr., *Perrier de la Bâthie 1744* (holo-, P[mounted on 2 sheets] (P00727174) and (P00727175).

ADDITIONAL MATERIAL EXAMINED. — **Madagascar**. Prov. Mahajunga, bords de la Manambolo, Bemaraha AP, [18°41'S, 44°46'E], 150–750 m, X.1964, fl., *Morat 1224* (G, MA, MO, P, TAN, TEF). — Forêt de l'Antsingy (bordure occidentale du Bemaraha), Bemaraha AP, [18°40'S, 44°44'E], 1.IX.1953, fl., *Service Forestier 8435* (G, MO, P, TEF).

DISTRIBUTION AND ECOLOGY

Lepisanthes chrysotricha is endemic to western Madagascar (Bemaraha and Causse de Kelifely, Ambongo), where it occurs exclusively on limestone substrate (Fig. 1).

REMARKS

Lepisanthes chrysotricha is a small tree to 7 m tall known only from three collections made in deciduous forest on limestone in western Madagascar. This species can easily be distinguished from all other Malagasy members of the genus by the golden indument covering its entire inflorescence and the outer surface of its sepals and petals, by its leaves with two pairs of leaflets that are usually ca. 7 cm long, and by its petiole and rachis, whose combined length does not usually exceed 4.5 cm. Fruits of this species are not known.

CONSERVATION STATUS

With an EOO of 433 km², an AOO of 27 km², and 2 subpopulations, one of which is situated within a protected area (Bemaraha), *Lepisanthes chrysotricha* is assigned a preliminary status of Endangered (EN B1ab(iii); B2ab(iii)).

2. *Lepisanthes perrieri*

(Choux) Buerki, Callm. & Lowry comb. nov. and stat. nov. INED

Manongarivea perrieri Choux, Compt. Rend. Hebd. Séances Acad. Sci. 182: 713. (1926); *Mém.*

Acad. Malg. 4: 36 (1927). — *Aphania senegalensis* subsp. *senegalensis* fa. *perrieri*

(Choux) Capuron, *Mém. Mus. Nat. Hist. Nat., sér. B*, 19: 82. — Lectotype (here designated): Madagascar, Prov. Mahajunga, Manongarivo (Ambongo), [16°16'S, 45°22'E], X.1909, fl., y fr., *Perrier de la Bâthie 1812* [P (P00624104); iso-, K (K000426188), P (P00624103), TAN].

As indicated above, Capuron (1969) used one of the syntypes of *Manongarivea perrieri* (*Perrier de la Bâthie 1744*) as the type of his name *Aphania senegalensis* subsp. *chrysotricha*, implicitly leaving only *Perrier de la Bâthie 1812* associated with Choux's name. Here we formalize this choice, designating one of the two sheets in the Paris Herbarium as the lectotype of *M. perrieri*.

ADDITIONAL MATERIAL EXAMINED. — **Madagascar**. Prov. Antsiranana. — Ankarana AP, 12°54'32"S, 49°06'35"E, 172 m, 7.XII.2006, y fr., *Andriamihajarivo et al. 1047* (MO, P, TAN). — Ankarana AP, 12°49'00"S, 49°01'17"E, 150 m, 12-20.X.1993, fl., *Andrianantoanina 368* (MA, MO, P, TAN). — Ankarana PA, 2 km avant le village d'Ambondromifehy, 12°52'36"S, 49°13'13"E, 276 m, 21.XI.1995, fr., *Andrianantoanina & Bezara 892* (G, MO, P, TAN). — Montagne d'Ambre AP, c. 8 km à l'Est du village de Bobakilandy, 12°37'37"S, 49°6'26"E, 533

m, 5.XII.1995, fr., *Andrianantoanina & Bezara 896* (G, MO, P, TAN). — Ankarana AP, 12°55'23"S, 49°05'11"E, 150 m, 20.X.2001, fl., *Bardot-Vaucoulon & Andrianantoanina 820* (P[2 sheets], TAN). — Ankarana AP, 100 m, 12°57'35"S, 49°07'15"E, 30.X.2001, fr., *Bardot-Vaucoulon & Andrianantoanina 871* (K, MO, P[2 sheets], TAN). — Ankarana AP, 12°58'52"S, 49°06'26"E, 125 m, 12.III.2001, fl., *Bardot-Vaucoulon & Andrianantoanina 901* (K, MO, P[2 sheets], TAN). — Ankarana AP, 12°53'38"S, 49°06'47"E, 130 m, 09.XII.1997, fr., *Bardot-Vaucoulon & Toly 1042* (K, MO, P[2 sheets], TAN). — Abattoir, 22.XII.1916, fl., *Decary 77* (P). — Ankarana AP, 12°54'40"S, 49°06'31"E, 180 m, 19.XI.1996, fl., y fr., *Labat 2772* (K, L, MO, P, TAN, WAG). — Daraina, forêt de Binara, 13°15'S, 49°37'E, 300 m, 5.XI.2001, fl., *Gautier & Ravelonarivo 4045* (G, P). — Daraina, forêt de Bekaraoka, 13°06'S, 49°42'E, 450 m, 22.XI.2006, fl., *Gautier & Chatelain 4912* (G, P). — Ankarana AP, 100-200 m, 12°51'S, 49°04'E, 22-26.XI. 1992, fl., *Malcomber et al. 1821* (K[2 sheets], G, MO, P, TAN). — Ankarana AP, 100–200 m, 12°51'S, 49°04'E, 22–26.XI. 1992, y fr., *Malcomber et al. 1823* (K, MO, P). — *Malcomber et al. 1882* (MO). — Ankarana AP, 100–200 m, 12°51'S, 49°04'E, 22-26.XI. 1992, fl., *Malcomber et al. 1886* (MO, P, TAN, WAG). — Ankarana AP, 100–200 m, 12°51'S, 49°04'E, 22–26.XI. 1992, fr., *Malcomber et al. 1887* (MA, MO, P, TAN). — Ankarana AP, 12°52'S, 49°14'E, 320 m, 24.XI.1993, fr., *McPherson 14525* (MO, P, TAN, TEF). — Nosy Mitsiou, [12°54'S, 48°36'E], X.1936, fl., *Perrier de la Bâthie 18775* (G, MO, P). — Ampondrabe, 12°58'18"S, 49°41'57"E, 200 m, 5. XI. 2005, fl., *Rakotonandrasana et al. 953* (CNARP, MO, P, TAN). — Befarafara, 13°05'26"S, 049°34'32"E, 100 m, 15.XI.2005, fl., *Rakotonandrasana et al. 1002* (CNARP, MO, P, TAN). — Daraina, forêt de Bobankora, 13°11'4"S, 49°45'30"E, 200 m, 11.XI.2005, fl., *Rakotondrajaona et al. 378* (CNARP, MO, P, TAN). — Daraina, forêt d'Antsahabe, 13°13'S, 49°33'E, 340 m, fr., 28.I.2006, *Ranarison et al. 1134* (G, MO, K, P, TEF). — Ananjaka, 13°06'00"S, 049°10'07"E, 155 m, 6. XI. 2004, fl., *Randrianaivo et al. 1093* (CNARP, MO, P, TAN). — Antsisikala, forêt d'Analabe, 12°11'49"S, 049°11'51"E, 184 m, 11.XII.2005, fl., *Randrianaivo et al. 1336* (CNARP, MO, P, TAN). — Ankarana AP, 12°58'57"S, 049°09'50"E, 127 m, 10. XII. 2007, fr., *Randrianasolo.S et al. 630* (CNARP, MO, P, TAN). — Ankijomantsina, 13°07'40"S, 49°27'53"E, 100 m, 31. X. 2005, fl.,

Ratovoson et al. 1044 (CNARP, MO, P, TAN). — Antsisikala, 12°10'02"E, 049°12'38"S, 100 m, 03.XII.2007, fr., *Ratovoson et al. 1416* (CNARP, MO, P, TAN). — 23 km S of Anivorano, 40 km N of Ambilobe, 12°53'S, 49°12'E, 200 m, 24.X.1992, fl., *Schatz et al. 2409* (MO, P, TAN). — Massif de la Montagne d'Ambre, [12°30'S, 49°05'E], c. 300 m., 19.X.1958, fl., y fr., *Service Forestier 11343* (G, K, MO, P, TEF). — Windsor Castle, [12°14'S, 49°10'30"E], c. 100 m., 4.XI.1955, fr., *Service Forestier 15168* (TEF). — Ambondromifehy, [12°53'30"S, 49°12'30"E], c. 300 m., 24.X.1959, fl., y fr., *Service Forestier 15189* (TEF). — Près d'Ambohimagodra, [13°01'30"S, 49°08'E], c. 100 m., 12.XI.1958, fr., *Service Forestier 18951* (TEF). — Ankarana, Forêt de Marovato, Massif de l'Ankerana, [12°48'30"S, 49°09'E], 200 m, 14.XI.1958, fr., *Service Forestier 20008* (P, TEF). — S du Pic Raynaud, [12°28'30"S, 49°26'30"E], 100–400 m, 09.XI.1961, fl., *Service Forestier 20364* (P, TEF). — S du Pic Raynaud, [12°28'30"S, 49°26'30"E], 100–400 m, 09.XI.1961, fl., *Service Forestier 20365* (TEF). — Nord de Vohémar, Mafokovo, [13°18'S, 49°51'E], [50–450 m], 18.XII.1970, y fr., *Service Forestier 27370* (P, TEF).

Prov. Mahajanga, Manongarivo, Ambongo, [16°16'S 45°22'E], 02.X.1910, fl., y fr., *Perrier de la Bâthie 1812* (P). — Soalala, Ambongo, [16°06'S, 45°19'E], XII.1926, fl., y fr., *Perrier de la Bâthie 17842* (P[2 sheets]). — Bongolava, Borizini, 15°38'58"S, 47°35'03"E, 217 m, 16.XI.2004, fl., *Ramananjanahary et al. 133* (MO, P, TEF). — Bongolava, Borizini, 15°34'00"S, 47°29'55"E, 115 m, 27.XI.2004, fr., *Ramananjanahary et al. 166* (MO, P, TEF).

— Ambondro, Analalava, [15°01'S, 47°16'E], 01.X.1958, fr., *Service Forestier 10778* (P). — Massif forestier de Bora, [14°54'S, 48°13'E], 100–411 m, 29.XI.1961, fr., *Service Forestier 18496* (P, TEF). — Forêt d'Ambondro-Ampasy, [15°01'S, 47°16'E], 50 m, 29.X–03.XI.1958, fr., *Service Forestier 18811* (P, TEF). — Soalala, village le plus proche Anjiafitatra, [15°26'S, 46°37'E], 100 m, 13.X.1967, fl., *Service Forestier 21383* (MO, P, TEF). N de Soalala (baie de Baly), [16°05'S, 45°09'E], 100 m, 22.XI.1965, fr., *Service Forestier 24248* (TEF). Plateau de Berivotra, [15°54'S, 46°34'30"E], 100 m, 24.XI.1965, fl., *Service Forestier 24293* (P, TEF).

DISTRIBUTION AND ECOLOGY

Lepisanthes perrieri occurs in semi-deciduous and deciduous forests from western Madagascar (Ambongo, Soalala) to the far north of the island (Montagne d'Ambre PN, Ankarana RS and Daraina) and appears to grow on limestone and basement rock (Fig. 1).

REMARKS

Lepisanthes perrieri can easily be distinguished by its leaves usually with just a single or two (rarely three) pairs of leaflets (vs. consistently two pairs in *L. chrysotricha* and three or rarely two or four in *L. sambiranensis*), a petiole and rachis that combined usually measures less than 2.5 (occasionally up to 11) cm (vs. 4.5 (–6) cm in *L. chrysotricha* and 7 – 15 (–22) cm in *L. sambiranensis*), and a glabrous inflorescence (vs. golden indument covering the whole inflorescence in *L. chrysotricha*, but glabrous in *L. sambiranensis*).

CONSERVATION STATUS

With an EOO of 74,179 km², an AOO of 133 km², and 19 subpopulations, six of which is situated within a protected area (Ankarana, Baie de Baly), *Lepisanthes perrieri* is assigned a preliminary status of Least Concerned (LC).

3. *Lepisanthes sambiranensis*

Buerki, Callm. & Lowry sp. nov. INED (Fig. 2)

Haec species a congeneris madagascariensibus foliorum petiolo cum rhachide (10–) 15 (–20) cm longo, foliolis (2- vel) 3- (vel 4-) jugatis, fructu sphaerico brunneolo atque exocapio sicco differt.

TYPUS. — **Madagascar**. Prov. Antsiranana, Kalabenono, Ambilobe, Beramanja, Anketrabe, forêt de Kalabenono, 13°38'41"S, 48°40'27"E, 730 m, fr., *Callmander, Jo Vasaha & Malaza 627* (holo-, P; iso-, G, K, MO, TAN).

PARATYPES. — **Madagascar**. Prov. Antsiranana, massif du Manongarivo, escarpement gréseux dominant la rive gauche de l'Antsahankolona, à l'Est d'Analantsoa, bords d'un ruisseau, [14°06'E, 48°18'S], 450 m, XI.1954, fr., *Service Forestier 11479* (P, TEF). — Manongarivo RS, à l'Est d'Ankaramibe, Bekolosy, 14°03'05"S, 48°17'07"E, 600–800 m, 10.XII.1993, fr., *Rakotomalala & Fernand 77* (MO, P, TAN).

DESCRIPTION

Shrub 3–4 m tall; stems c. 8–10 mm in diam., glabrous, brownish, striate. Leaves paripinnate, petiole and rachis stout (ca. 2 mm diameter in dried material), 7–15 (–22) cm long (combined), glabrous, striate, light green (brownish at the base), slightly compressed; leaflets (2 or) 3 (or 4) pairs, subopposite to alternate, subcoriaceous, green, 12–15 (–25) × 3–5 (–8) cm, elliptic, base acuminate, margin entire, apex acuminate to cuspidate, acumen 1.5–2 cm, pseudostipules lacking; petiolules 1 cm, glabrous, brown, conduplicate, woody. Inflorescence axillary, 14–18 cm, paniculate, 2-branched, glabrous. Flowers unknown. Fruit spherical, brownish, 2-carpellate, generally only one carpel developing at maturity, 8–15 mm in diam. (the second carpel reduced, 1.5–5 mm in diam.), glabrous, exocarp dry, slightly punctuate, thin (<1 mm), endocarp dry, reddish; style apparently gynobasic, persistent, capitate, 1 mm; peduncle glabrous, 8–12 mm, articulate in the basal third; 1-seeded, subspherical, c. 6–12 mm in diam. Fig. 2.

DISTRIBUTION AND ECOLOGY

Lepisanthes sambiranensis is endemic to sub-humid forests on sandstone substrates in the Kalabenono and Manongarivo massifs in northwestern Madagascar (Fig. 1).

REMARKS

Lepisanthes sambiranensis differs from the other Malagasy members of the genus by its long, stout petiole and rachis, which have a combined length of 7–15 (–22) cm, and a diameter of ca. 2 mm (vs. usually no more than 5 cm long, and 1.5 mm diameter in the other species), usually three pairs of leaflets (vs. usually only one or two pairs in *L. perrieri* and two in *L. chrysotricha*), leaflets 12–15 (–25) cm (vs. rarely more than 7–10 cm in the other taxa), and spherical brownish fruit with a dry exocarp (vs. oblong to ellipsoid, rarely spherical in *L. perrieri*, but always with a fleshy exocarp becoming black when dry; unknown in *L. chrysotricha*).

ETYMOLOGY

The species epithet refers to the Sambirano biogeographic region to which *Lepisanthes sambiranensis* appears to be restricted.

CONSERVATION STATUS

With an EOO of c. 10 km², an AOO of 27 km², and 2 subpopulations, one of which is situated within a protected area (Manongarivo), *Lepisanthes sambiranensis* is assigned a preliminary status of Endangered (EN B1ab(iii); B2ab(iii)).

ACKNOWLEDGMENTS

We thank our collaborators at the Parc Botanique et Zoologique de Tsimbazaza (TAN) and the Département de Recherche Forestière et Piscicole (TEF), FOFIFA for assistance in Madagascar. We are grateful to Laurent Gautier, Louis Nusbaumer and Nicolas Fumeaux for assistance in Geneva, Alyse Rothrock and Zachary Rogers at MBG in St. Louis, and staff of the MBG in Antananarivo, especially Noro Ravololomanana and Marina Rabarimanarivo, for assisting with the management of project collections. We thank Roy Gereau for preparing the Latin diagnosis and Roger Lala Andriamiarisoa for the fine illustration. The authors are also grateful to the curators of the herbaria in Antananarivo, Geneva, St. Louis, Madrid and Paris for providing access to their collections. Financial support to SB was provided by the European Community's Programme Structuring the European Research Area under a SYNTHESYS grant to visit the Real Jardín Botánico (CSIC) in Madrid. Financial support was provided to MWC by Conservation International-Madagascar and the National Geographic Society (grant no. 8114-06). The participation of MWC, PPL and PBP was also supported by grants from the U.S. National Science Foundation (0743355; PPL Co-PI) and the Andrew W. Mellon Foundation (PBP and PPL Co-PIs).

REFERENCES

- ADEMA F., LEENHOUTS P. W. & VAN WELZEN P. C. 1994. — Sapindaceae. *Flora Malesiana*, sér. I, 11: 419–768.
- BUERKI S., FOREST F., ACEVEDO-RODRIGUEZ P., CALLMANDER M. W., NYLANDER J. A. A., HARRINGTON M., SANMARTIN I., KUEPFER, P. & ALVAREZ N. 2009 — Worldwide phylogeny of the soapberry family (Sapindaceae): plastid and nuclear markers reveal intricate relationships at subfamilial, tribal and generic levels. *Molecular Phylogenetics and Evolution* 51: 238–258.
- CALLMANDER M. W., SCHATZ G. E., LOWRY II P. P., LAIVAO M. O., RAHARIMAMPIONONA J., ANDRIAMBOLOLONERA S., RAMINOSOA T. & CONSIGLIO T. 2007. — Application of IUCN

- Red List criteria and assessment of Priority Areas for Plant Conservation in Madagascar: rare and threatened Pandanaceae indicate new sites in need of protection. *Oryx* 42: 168–176.
- CAPURON R. 1969. — Révision des Sapindacées de Madagascar et des Comores. *Mémoires du Muséum National Histoire Naturelle, sér. B, Botanique* 19: 1–189.
- CHOUX P. 1926. — Quelques nouvelles Sapindacées de Madagascar. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences* 182: 713
- CHOUX P. 1927. — Les Sapindacées de Madagascar. *Mémoire de l'Académie Malgache* 4: 1–118.
- CORNET, A. 1974. — Essai de cartographie bioclimatique à Madagascar. Notice Explicative. *ORSTOM*. 1–55.
- ESRI 2000. — *ArcView* 3.3. Redlands, California.
- HOLMGREN P. K., HOLMGREN N. H. & BARNETT L. C. 1990. — *Index Herbariorum. Part I: The herbaria of the World*. New York Botanical Garden: New York.
- IUCN 2001. — *IUCN Red List Categories and Criteria: Version 3.1*. IUCN Species Survival Commission, IUCN, Gland, Switzerland; Cambridge, United-Kingdom.
- LAMARCK J. B. A. P. M. DE. 1805. — *Encyclopédie Méthodique, Botanique, Vol. 6*. H. Agasse, Paris, France.
- LEENHOUTS P. W. 1969. — Florae Malesianae praecursores. 50, a revision of *Lepisanthes* (Sapindaceae). *Blumea* 17: 33–91.
- RADLKOFER L. 1933. — Sapindaceae in ENGLER A. (ed.), *Das Pflanzenreich IV*, 165 (Heft 984 98a–h). Leipzig, Verlag von Wilhelm Engelmann.
- SCHATZ G. E. 2000. — Endemism in the Malagasy tree flora. Pp. 1–11 in *Diversity and Endemism in Madagascar*, eds. W. R. LOURENÇO & GOODMAN S. M., Mémoires de la Société de Biogéographie, MNHN, OSRTOM, Paris.
- SCHATZ G. E. & LESCOT M. 2005. — *Gazetteer to Malagasy Botanical Collecting Localities*. Missouri Botanical Garden website [accessed 30 March 2009]:
<http://www.mobot.org/MOBOT/Research/madagascar/gazetteer>.

SCHATZ G.E., GEREAU R. E. & LOWRY II P. P. 1999. — A revision of the Malagasy endemic genus *Chouxia* Capuron (Sapindaceae). *Adansonia*, sér. 3, 21: 51–62.

LEGENDES

FIG. 1. – Distributions of *Lepisanthes* mapped on the bioclimatic zones of Madagascar (after Cornet 1974; see Schatz 2000): *L. chrysotricha* (squares), *L. perrieri* (circles), and *L. sambiranensis* (triangles).

FIG. 2. – *Lepisanthes sambiranensis* Buerki, Callm. & Lowry: **A**, seed; **B**, fruiting branch; **C**, details of a fruit. A-C, *Callmander, Jo Vasaha & Malaza 627* (holotype, P). Scale bars: A, C, 2 cm; B, 3 cm.

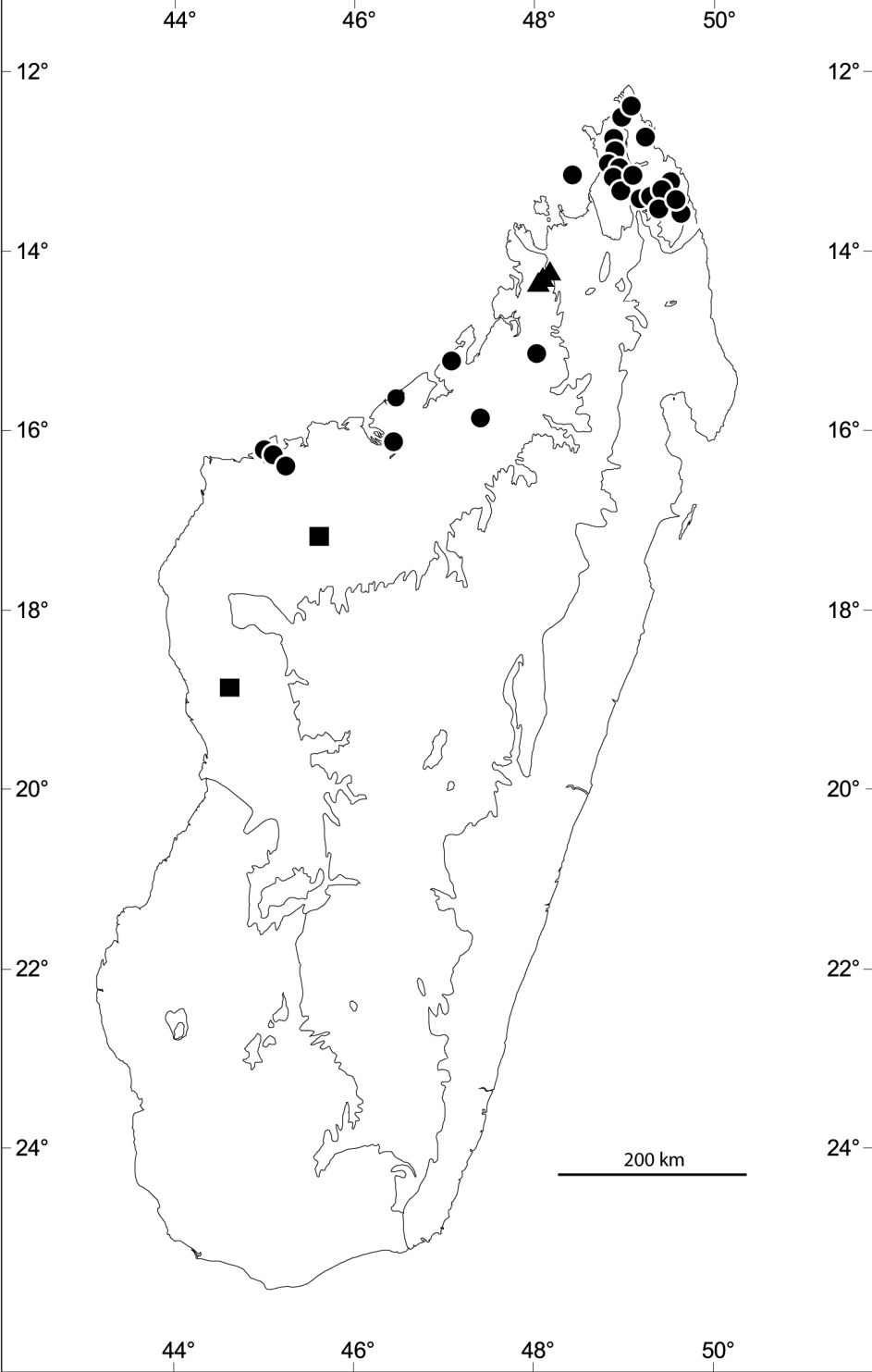


Figure 1

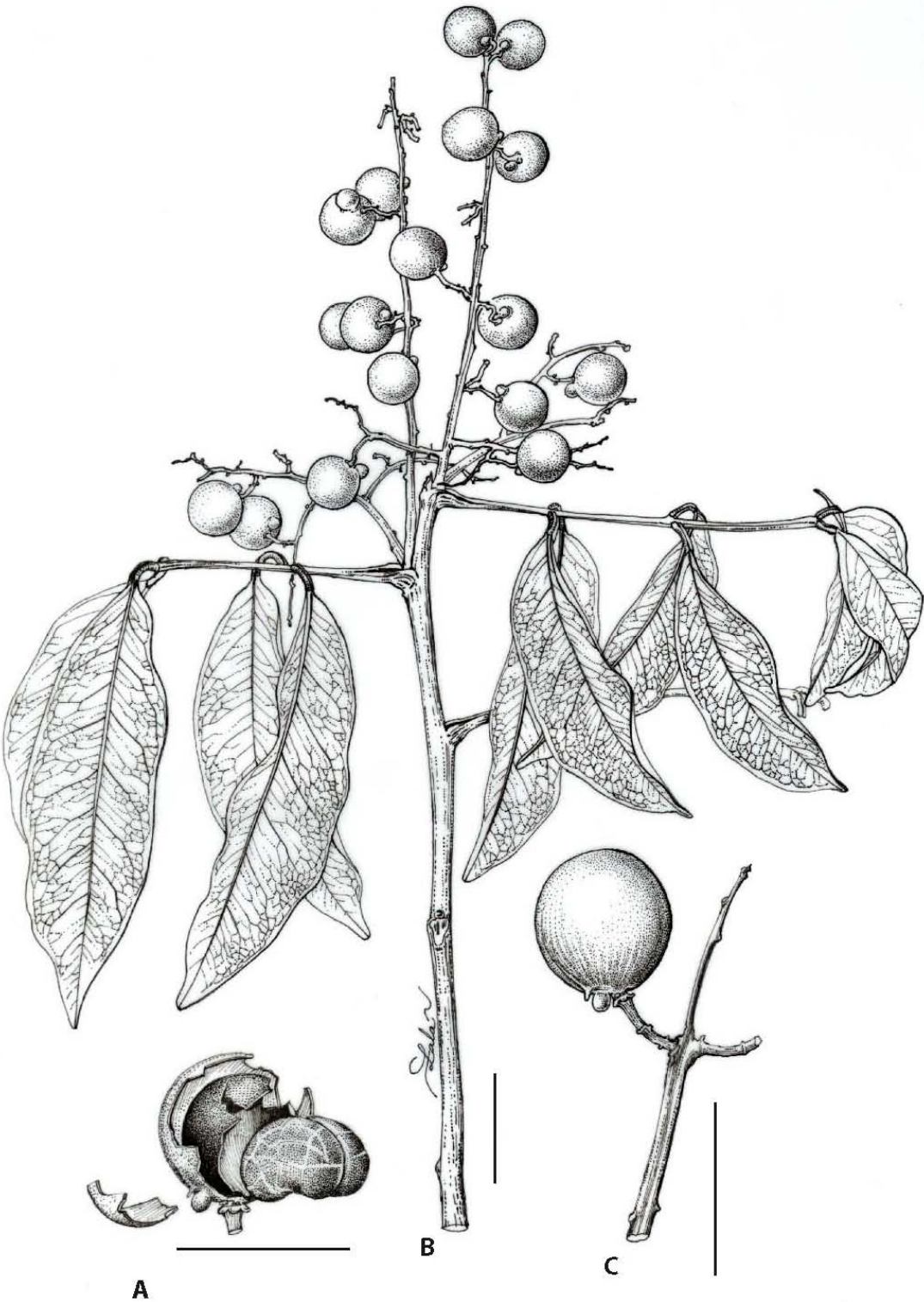


Figure 2

**Chapter six: Comparative Performance of Supertree
Algorithms in Large Data Sets using the Soapberry Family
(Sapindaceae) as a case study**

Sven Buerki, Félix Forest and Nadir Alvarez

Submitted to *Systematic Biology*

Running head: COMPARATIVE PERFORMANCE OF SUPERTREES IN LARGE DATA SETS

Comparative Performance of Supertree Algorithms in Large Data Sets using the Soapberry Family (Sapindaceae) as a case study

SVEN BUERKI¹, FÉLIX FOREST,² and NADIR ALVAREZ¹

¹ *Institute of Biology, University of Neuchâtel, Rue Emile-Argand 11, CH-2000 Neuchâtel, Suisse; E-mail: sven.buerki@unine.ch (S. B.).*

² *Molecular Systematics Section, Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, United Kingdom.*

Abstract Supertrees represent comprehensive phylogenetic hypotheses incorporating all taxa present in partially overlapping input trees. For two decades, supertree methods have been an active research framework and almost 20 major algorithms differing by the kind of criterion and the type of input data have been described [e.g., Matrix Representation with Parsimony (MRP), MinFlip, Average Consensus (AVCON), Most Similar Supertree (MSS), Split fit and MinCut]. Because of the growing popularity of supertree methods and the overwhelming dominance of the MRP method, it becomes necessary evaluating the performance of the different methods in the context of large data sets to determine if the supertree algorithms work well and which are the best options. However, only few studies compared the performance of two or more supertree algorithms when using input trees comprising a large number of taxa. In this study, we investigate for the first time the comparative performance of six major methods, by using a large empirical data set. Based on eight maximum parsimony and maximum likelihood consensus single-gene phylogenies of the soapberry family (Sapindaceae; 240 ingroup samples), we evaluate the performance of each method, according to its compatibility with the input trees and with total evidence trees (reconstructed using the same algorithms as the input trees). Based on the defined criteria, our results indicate that the behaviour of the supertree methods is not influenced by the algorithm used to reconstruct the input trees; however, ML-based supertrees seem to fit the phylogenetic signal of the input trees better than do MP-based supertrees. Moreover, our results demonstrate a large variation in the performance of the different methods: only three methods – MRP, MinFlip and MinCut – performed properly, whereas the others largely failed in producing accurate topologies. We pointed out that the poor performance of AVCON, MSS and Split fit methods could be related to the efficiency of the heuristic searches rather than to the supertree algorithms themselves. Overall, MRP supertrees fit the phylogenetic signal of the input trees better, whereas MinFlip supertrees are more similar to total evidence trees. MinCut supertrees constitute an average solution that is, however, less accurate than the two previous methods.

Chapter 6

Key Words Matrix Representation with Parsimony; MinFlip; Sapindaceae; Supertree; Total evidence tree.

Historically, the development of supertrees has arisen from a need to produce more inclusive phylogenies (e.g., Purvis, 1995a; Bininda-Emonds et al., 1999; Wojciechowski et al., 2000; Jones et al., 2002; Salamin et al., 2002; Stoner et al., 2003; Cardillo et al., 2004; Price et al., 2005). The supertree framework enables the combination of partially overlapping phylogenetic trees obtained using different kind of data (e.g. DNA, protein, morphology) and/or algorithms (e.g. parsimony, likelihood, distance). Since their first description by Gordon (1986), supertree reconstructions have considerably evolved and no less than 14 methods have been described to date (see reviews by Sanderson et al., 1998; Bininda-Emonds et al., 2002; Bininda-Emonds, 2004; Wilkinson et al., 2005a) and some additional new methods have not yet been implemented (e.g. Majority-rule supertrees, Cotton and Wilkinson, 2007).

From a global perspective, supertrees might be viewed as an extension of the consensus techniques for the particular case of partially overlapping input trees. To circumscribe the different tree-building algorithms better, Wilkinson et al. (2001) classified the various supertree algorithms as either “direct” or “indirect”. Direct methods are very similar to the consensus techniques whereby the supertree is directly derived from the input trees without an intermediate discrete step [i.e., the MinCut (MC; Semple and Steel, 2000) and modified MinCut (MMC; Page, 2002) methods]. In contrast, indirect methods reduce the input trees into a matrix representation (MR) that is analysed using an optimization criterion. Although the most widely used indirect method is by far the Matrix Representation with Parsimony (MRP, Baum, 1992, Ragan, 1992), four other indirect methods using different criteria have been developed [MinFlip (Chen et al., 2003; Eulenstein et al., 2004); Average Consensus (AVCON, Lapointe and Cucumel, 1997; Lapointe and Levasseur, 2004); Most Similar Supertree (MSS, Creevey et al., 2004); Split Fit (Rodrigo, 1993; Creevey and McInerney, 2005)].

Although supertrees have been constructed increasingly over the last 15 years (e.g., Bininda-Emonds et al., 1999; Lapointe et al., 1999; Jones et al., 2002; Cardillo et al., 2004; Price et al., 2005), they also have been considerably criticized because they do not directly

rely on the raw data (e.g., Rodrigo, 1993, 1996; Slowinski and Page, 1999; Novacek, 2001; Springer and deJong, 2001; Gatesy et al., 2002), in contrast to total evidence reconstructions (*sensu* Kluge, 1989). However, as underlined by Bininda-Emonds et al. (2002, 2003), the loss of contact with the primary data is a necessary trade-off when seeking to combine all possible sources of phylogenetic information. Several studies have been performed to test the accuracy of a range of supertree methods using simulations (e.g., Bininda-Emonds and Sanderson, 2001; Chen et al., 2003; Eulenstein et al., 2004; Lapointe and Levasseur, 2004) or empirical data sets (Salamin et al., 2002). However, most relied on a small number of taxa, not reflective of the empirical scope of the supertree framework, that aim to reassemble single trees (a source of phylogenetic information that has been building exponentially for the last two decades) into wider phylogenies encompassing higher taxonomic levels. It therefore, seems timely to evaluate the performance of different supertree methods in the context of large empirical data sets.

Following this idea, Salamin et al. (2002) published the first empirical study in which the performance of MRP supertrees was compared to that of a total evidence tree using as case-study the grass family (Poaceae). However, they did not carry out a broad comparative study of different supertree methods. Here, we propose to compare the performance of six major supertree algorithms using a large empirical data set comprising taxa from the soapberry family (Sapindaceae) partially based on Buerki et al. (2009). Since the first family-wide treatment proposed by Radlkofer (1933), the classification of Sapindaceae has remained problematic and successive researchers have failed to provide a convincing family classification based on morphological characters (Muller and Leenhouts, 1976; Umadevi and Daniel, 1991; Thorne, 2000). Recently, however, Harrington et al. (2005) published the first molecular phylogeny of the family (based on 112 taxa and two plastid markers), from which a new subfamily classification was proposed. They recognized four major clades (Sapindoideae, Hippocastanoideae, Dodonaeoideae and Xanthoceroideae) and pointed out the para- or polyphyletic nature of several classical tribes. Following this initial attempt at circumscribing subfamilial entities, Buerki et al. (2009) inferred a worldwide phylogeny of

Sapindaceae based on 154 specimens (representing ca. 60% of the generic diversity) and eight molecular markers. The authors confirmed the results of Harrington et al. (2005) and demonstrated a high degree of para- or polyphyly at the subfamilial and tribal levels. The dataset presented in Buerki et al. (2009) favoured the minimization of missing data against generic diversity, thus preventing the reconstruction of a more extensive phylogeny of the family necessary to propose of a new formal classification.

Here we assess the performance of supertree methods for large datasets and provide a phylogenetic hypothesis for the soapberry family based on 240 specimens representing more than 70% of the generic diversity.

We performed MRP, MinFlip, Split fit, AVCON, MSS and MinCut (MC and MMC) supertree analyses using as input trees (reconstructed with maximum parsimony and maximum likelihood algorithms) the topology obtained for each of the eight partitions, and computed total evidence reconstructions based on the same algorithms as input trees. We further compared the performance of the different supertree algorithms families in our large data set by assessing their level of congruence with the input trees and total evidence trees.

MATERIAL AND METHODS

Sampling and sequence data

The sampling strategy covers 104 of the 141 currently recognized of Sapindaceae genera (Buerki et al., 2009), including all subfamilies and tribes recognized in the classic taxonomy of Radlkofer (1933) and Muller and Leenhouts (1976), and updates proposed by Harrington et al. (2005) and Buerki et al. (2009). The ingroup sampling comprised 240 samples from which 90 were newly produced for this study and the remaining samples corresponded to the dataset of Buerki et al. (2009). Outgroup taxa included one Anacardiaceae (*Sorindeia* sp.; defined as outgroup in all analyses; Savolainen et al., 2000 and Muellner et al., 2007), one Rutaceae (*Harrisonia abyssinica*) and one Meliaceae (*Malleastrum* sp.). Voucher citations are provided in Appendix.

Seven plastid regions [coding regions *rpoB* and *matK*; *trnL* intron; intergeneric spacers (IGS) *trnK-matK*, *trnD-trnT*, *trnL-trnF* and *trnS-trnG*] and one nuclear region (ITS region comprising the *ITS1* and *ITS2* internal transcribed spacers, and the *5.8S* gene) were amplified. The primer information, PCR and sequencing protocols are given in Buerki et al. (2009).

Total evidence analyses

A supermatrix composed of the eight regions was built using CONCATENATE (Alexis Criscuolo, <http://www.lirmm.fr/~criscuolo/>). In the supermatrix, taxa in which no sequences were gathered for a given partition were coded as missing values for the corresponding cells. Total evidence trees (*sensu* Kluge, 1989) were determined using both maximum parsimony (MP) and maximum likelihood (ML) criteria. In the present study, total evidence trees (reconstructed with the same algorithms as the input trees; see below) are used as references to assess the performance of supertree methods.

Maximum Parsimony Analyses. – The supermatrix was analyzed using parsimony ratchet (Nixon, 1999) as implemented in PAUPrat (Sikes and Lewis, 2001). Based on recommendations by Nixon (1999), ten independent searches were performed with 200 iterations and 15% of the parsimony informative characters perturbed. The shortest equally parsimonious trees were combined to produce strict and majority-rule consensus trees. To assess the support at each node, nonparametric bootstrap analyses (Felsenstein, 1985) were performed using PAUP* version 4.0b10 (Swofford, 2002) with 1000 replicates, SPR branch swapping, simple addition sequence, MULTREES option in effect and saving no more than ten trees per replicate. SPR branch swapping was used because it was shown to be twice as fast as TBR and resulting in support values not significantly different from those obtained with TBR (Salamin et al., 2003). In order to quantify the amount of resolution in the consensus trees (i.e., the proportion of resolved nodes; see Price et al., 2005), the consensus fork index (CFI; Colless, 1981) was calculated using PAUP* version 4.0b10

(Swofford, 2002). A normalized CFI index of 1.0 indicates that the tree is fully resolved, whereas a lower value indicates the presence of polytomies.

Maximum Likelihood Analyses. – Model selection for the supermatrix was assessed using Modeltest version 3.7 (Posada and Crandall, 1998) and the Akaike information criterion (Akaike, 1973). ML analyses were performed using RAxML version 7.0.0 (Stamatakis, 2006) with a 1000 rapid bootstrap analyses (Stamatakis et al., 2008) followed by the search of the best-scoring ML tree in one single run. This analysis was done using the facilities offered by the CIPRES portal in San-Diego, U.S.A. (<http://8ball.sdsc.edu:8888/cipres-web/home>).

Input trees reconstruction

Phylogenetic analyses of Sapindaceae were performed using MP and ML criteria for each of the eight DNA partitions, for a total of 16 input trees. The MP, ML and bootstrap analyses were performed using the same settings as for the total evidence analyses (see above). For each single-gene MP analysis, the majority-rule consensus trees were used as input data for supertree reconstructions. For ML analyses, the bootstrap majority-rule consensus trees were used as input data for supertree reconstruction to avoid any bias caused by sub-sampling during the ML heuristic searches and to prevent the use of fully-resolved trees, which could be potentially misleading (e.g. the output of a ML analysis corresponds to the phylogenetic tree that maximize the likelihood, which is fully resolved by definition). Variations in the number of specimens included and the level of resolution of each input tree makes the single-genes phylogenies well suited as starting points for comparing supertree methods.

Supertree reconstruction

Based on the MP and ML input trees, we reconstructed supertrees using the six following supertree methods (the criterion used to find the best tree is indicated in brackets): MRP (parsimony), MinFlip (flip), Split fit (compatibility), AVCON (least-square distance), MSS (distance) and MinCut (adapted from the Aho algorithm; Aho et al., 1981). The settings for

each supertree method detailed below were applied to both MP and ML input trees and when the heuristic search produced more than one phylogenetic tree, a strict consensus tree was built.

MRP analyses were performed according to both matrix representation (MR) coding schemes, Baum and Ragan (1993; hereafter "BR") and Purvis (1995b; hereafter "PU"). The BR MR (developed independently by Baum, 1992 and Ragan, 1992) used additive binary coding (Farris, 1973) to represent the hierarchical structure of source trees as a series of matrix elements (Baum and Ragan, 1993). Each node on a given source tree is represented by a binary matrix element. Terminal taxa delimited by each node are scored as 1 and all other taxa as 0. Missing taxa from an individual source tree are then coded as missing values. Regarding the BR coding, Purvis (1995b) argued that the elements derived from the source trees lack independence and hence add redundant information to the MR. He proposed to remove this apparent redundancy by coding only taxa among sister clades as 0, whereas all the others and the missing taxa are scored as missing values. For each coding type, six binary matrices were constructed using the program SuperTree 0.85b (Salamin et al., 2002), according to the type of parsimony (reversible, hereafter "rev" and irreversible, hereafter "irrev") and weighting procedure (unweighted; weighted by the inverse of the number of nodes present in each source tree, hereafter "nodes"; weighted by the bootstrap support of each node, hereafter "boot") (for more details, see Ronquist, 1996; Bininda-Emonds and Bryant, 1998; Salamin et al. 2002). During the weighting procedure of the bootstrap support of each node, SuperTree 0.85b (Salamin et al., 2002) scales the bootstrap frequencies instead of weighting the matrix with them. Trees were rooted by an all-zero outgroup taxon. The parsimony analyses were performed using a heuristic algorithm in PAUP* version 4.0b10 (Swofford, 2002) with a random addition sequence (nreps=20), TBR branch swapping, STEEPEST and MULTREES options in effect and an unlimited value for MAXTREES. Equally most-parsimonious solutions were summarized using a strict consensus tree. For the MinFlip approach (Eulenstein et al., 2004), based on the two coding schemes mentioned above (BR and PU), two matrices were constructed using SuperTree

0.85b (Salamin et al., 2002). The matrices were analysed with Chen's heuristic supertree software (Chen et al., 2004; <http://genome.cs.iastate.edu/CBL/download/>) with ten replicates, SPR branch swapping and MAXTREES=10000.

Split fit analyses were computed using CLANN (Creevey and McInerney, 2005). Two types of supertrees were constructed using this method, since the contribution of each source tree to the analysis can be normalized according to the number of taxa in each input tree (hereafter, Split fit equal) or not (hereafter, Split fit norm). For each coding scheme, heuristic searches were performed with 10 and 50 replicates, SPR branch swapping, nsteps=10 and maxswaps=10000. For the AVCON approach (Lapointe and Cucumel, 1997; Lapointe and Levasseur, 2004), the average path-length distance matrix was constructed using CLANN (Creevey and McInerney, 2005) and the missing data were estimated with the ultrametric method. The average path-length distance matrix was analysed through both the least-squares algorithm (hereafter, AVCON FITCH), implemented in FITCH (PHYLIP package, v.3.6; Felsenstein, 1993) and the neighbour-joining (NJ) algorithm (hereafter, AVCON NJ), implemented in PAUP* version 4.0b10 (Swofford, 2002). The MSS analyses were performed using CLANN (Creevey and McInerney, 2005). As for the Split fit approach, a normalization was applied (hereafter, MSS equal) or not (hereafter, MSS norm) to the source trees before the analysis and the heuristic searches were conducted with 10 and 50 replicates, SPR branch swapping, nsteps=20 and maxswaps=1000000. The MinCut approach (MC; Semple and Steel, 2000) and modified MinCut (MMC; Page, 2002) analyses were computed using RAINBOW (Chen et al., 2004). For each supertree analysis, the calculation time and CFI index values were compiled.

Supertree evaluation

Average Normalised Partition Metric (NMP) Distance. – The average NPM distance (also known as the Robinson-Foulds topological distance; Robinson and Foulds, 1981) between each supertree and the input trees (pruned to identical taxon sets) and between the supertrees and the total evidence trees was calculated using PartitionMetric version 1.2.1 (O.

R. P. Bininda-Emonds, <http://www.unioldenburg.de/molekularesystematik/>). We used this distance to evaluate the extent to which the supertrees were in agreement with the phylogenetic signal expressed in the input trees. Moreover, the NPM distance was used to quantify the relationships between each supertree and the total evidence trees (see below). The NPM distance is the sum of the components present in one but not both trees. A component refers to the relationships expressed by an internal branch, which separates the members of a clade from the non-members (including the root) (Wilkinson et al., 2005a). Components may entail less inclusive relationships than the quartet distance, which places two terminals closer to each other than a third (and in the case of the quartet distance, this triplet is rooted by a fourth taxon) (see below).

Quartet Distance. – The quartet distance (Estabrook et al., 1985 and Estabrook, 1992), also known as the explicitly agree distance (EA; Wilkinson et al., 2005a), quantifies the differences between trees of same size, thus it was used here to evaluate the distance between the supertrees and the total evidence trees (see below). It evaluates the proportion of quartets [i.e. a quartet expresses relationships between three ingroup taxa rooted by an outgroup taxon as follows: (Outgroup, (A, (B, C)))] that are resolved identically in the two trees (Wilkinson et al., 2005a). As mentioned by Wilkinson et al. (2005), because there are more resolved quartets than components in most trees, this measure is potentially more discerning than NPM distance, and is not as dramatically affected by instability in a single terminal. Here, this distance was used to quantify the similarity level among the 52 supertrees and the three total evidence trees, using Dquad (Alexis Criscuolo, <http://www.lirmm.fr/~criscuolo/>).

Tree of Trees. – On the basis of the NPM and quartet matrices of pairwise distances among trees, two unrooted “tree of trees” were built using the NJ algorithm (PHYLIP package, v.3.6; Felsenstein, 1993). Those trees depict the relationships between the supertrees and the total evidence trees.

Recovery of Key Clades. – In order to assess the quality of the supertree topologies, each supertree was evaluated by comparing it to supported monophyletic clades recovered by the three total evidence trees (Fig. 1):

1. At the “family” level: monophyly of the Sapindaceae sensu lato (including the previously described Aceraceae and Hippocastanaceae and the Sapindaceae sensu stricto).
2. At the “subfamilial” (corresponding to clades) and “tribal” (corresponding to groups) levels: a. monophyly of clades A-C; b. monophyly of groups within each clade. Clades A and B are subdivided into two groups and clade C is subdivided into eight groups (see Supplementary material).

RESULTS

Matrices Overview

The number of sequences included in each single matrix ranged from 80 for *trnS-trnG* IGS to 199 for *rpoB* and the supermatrix was composed by 1242 sequences (Table 1). The number of unique sequences per matrix varied from 0 (e.g., *trnD-trnT* IGS) to 26 for *matK*. To visualize the overlap between the matrices and their contribution to the supermatrix, the number of shared sequences is presented in Table 1. Information about alignment length and number of constant, variable and potentially parsimony-informative characters is also available in Table 1. GenBank accession numbers are provided in the Supplementary material.

Total Evidence Trees & Input Trees

Total Evidence Trees. – The supermatrix compiled for the total evidence analyses comprised 9657 characters (Table 1). The strict and majority-rule consensus trees based on 949 equally parsimonious trees resulting from the MP analysis were relatively well-resolved (CFI indices of 0.614 and 0.909 for the MP strict and MP majority-rule, respectively). For the ML analysis, the best-fit model was the general time reversible model with an alpha

parameter for the shape of the gamma distribution to account for among-site rate heterogeneity (Yang, 1993) and a proportion of invariable sites (GTR+G+I). This model was used to perform the ML search (log likelihood = -79560.8) followed by the 1000 rapid bootstrap analyses (Fig. 1; see Supplementary material for a detailed phylogenetic tree of Sapindaceae sensu lato). Each of the three total evidence analyses showed support for the monophyly of Sapindaceae sensu lato as defined by Thorne (2007), including Aceraceae and Hippocastanaceae. No matter which algorithm is considered, the family is subdivided into three moderately to strongly-supported lineages and a fourth lineage comprising only *Xanthoceras sorbifolia* with the following relationships: (*X. sorbifolia*, (clade A, (clade B, clade C))) (Fig 1; Supplementary material, Figs. S1-6). In the MP strict consensus tree, *X. sorbifolia* is embedded in a polytomy with clade A. The monotypic fourth lineage corresponded to subfamily Xanthoceroideae as described by Thorne (2007). Clade A is subdivided into two groups (*Acer* group and *Aesculus* group) corresponding to the formerly recognized families Aceraceae and Hippocastanaceae (Fig. 1). Clade B corresponded to subfamily Dodonaeoideae as described by Buerki et al. (2009) with the addition of *Diplokeleba* (tribe Cupanieae; subfam. Sapindoideae) and is subdivided into two groups (*Doratoxylon* group and *Dodonaea* group; Fig. 1). Clade C corresponded to subfamily Sapindoideae (Buerki et al., 2009) and is subdivided into ten groups, among which the *Cupania* and *Koelreuteria* groups were not monophyletic in the MP strict consensus tree.

Input Trees. – The best-fit model for all single-gene matrices was GTR+G. The only exception was ITS, for which a parameter for the proportion of invariable sites was necessary. The MP single-gene consensus trees had in general a higher CFI index than the ML trees (Table 1). No strong topological conflicts (bootstrap support >75%) were recognized between the MP/ML input trees and the total evidence trees. Additional characteristics (number of most parsimonious trees; tree length) for each analysis are reported in Table 1.

Supertree evaluation

The 52 supertrees reconstructed in this study using different methods show considerable variation in number of trees retained, resolution and computational time (Table 2). Trees were either fully resolved (e.g., MinCut, AVCON) or partially polytomized (e.g., MRP, MinFlip) (Table 2). Computational times were also highly variable, ranging from less than ten minutes (e.g. MinCut) to more than ten days (e.g. some Split fit analyses; Table 2).

Distance to the Input Trees. – Whichever algorithm is used, two supertree groups were defined according to their agreement with the input trees (Fig. 2): i) MRP, MinFlip and MinCut supertrees clustered together and showed a smaller distance to the input trees, whereas ii) the three other supertree methods clustered together and demonstrated a higher average NPM distance to the input trees. One possible exception is the supertree produced by the MinCut algorithm (MinCut MC), which has a somewhat intermediary NPM value between the two groups under the MP criterion (Fig. 2A). Within the first group, MPR BR rev had the highest agreement with the input trees, both under the MP and ML criteria, whereas MinCut and MRP PU irrev had the highest distances to the input trees. Although supertrees based on MP and ML algorithms behaved similarly in terms of their average NPM distance to the input trees, the NPM values between supertrees and input trees were in general slightly lower when reconstructed under the ML criterion (Fig. 2). MP and ML total evidence trees clustered with MRP, MinFlip and MinCut supertrees and have similar distances to the input trees as supertrees belonging to that group (Fig. 2).

Tree of Trees. – The “trees of trees” based on quartet (Fig. 3) and NPM distances (Fig. S7) depict the relationships between the 55 trees (52 supertrees and three total evidence trees) and were highly congruent. They highlighted the same general patterns than those highlighted by the distance to the input trees (Fig. 2). As for the pattern recovered in the quartet “tree of trees”, the MP MinCut MC is found in a intermediary position between the two groups, which makes the supertree produced by this method difficult to assign to either group. All MinFlip, total evidence trees clustered together with part of MRP and ML MinCut MMC (average quartet distance between the supertrees of 0.07; maximum quartet distance

of 0.159 between ML MinCut MMC and MP MRP BR rev nodes; minimum quartet distance of 0.0015 between ML MRP BR rev and ML MRP BR rev nodes), whereas all MRP PU rev supertrees and the remaining MinCut formed a transitional group close to the second group (Fig. 3; Fig. S7). Within the second group, the average quartet distance between supertrees was 0.51, with a maximum quartet distance of 0.67 between MP MSS equal 50 and ML MSS equal 10. All MP and ML Split fit had exactly the same topology (quartet distance=0) whatever the number of replicates and normalization to the source trees (Fig. 3; Fig. S7). Moreover, quartet and NPM distances to the ML total evidence tree are expressed as a vertical bar chart for all the supertrees and the two MP total evidence trees (Fig. 4).

Recovering Key Clades. – When clades found in supertrees are compared to those recovered in all total evidence trees, none of the supertrees fulfilled the 16 key-clade criteria (Table 2). The best score was obtained by the MP MinFlip PU method with 15 out 16 key clades in topological agreement with the total evidence trees. A bimodal distribution of the number of fulfilled criteria is observed among the supertree methods, with the MRP, MinFlip and MinCut methods showing the highest scores and the other methods demonstrating fewer or zero fulfilled criteria (Table 2). Within the first group, all MinCut and MRP PU rev supertrees had a lower agreement with total evidence trees than the other representatives (with 1 to 6 fulfilled criteria, whereas the others have 8 to 16 fulfilled criteria; Table 2). These patterns are similar to those previously observed with the distance to the input trees (Fig. 2) and the “trees of trees” (Figs. 3, S7).

DISCUSSION

Comparative Performance of the Supertree Algorithms: the Case of the Soapberry Family

This study does not seek to make a case in favour or against supermatrix or supertree methods. Several studies have pointed out weaknesses in these phylogenetic reconstruction methods, supermatrices (e.g., Huelsenbeck, 1991; Wiens and Reeder, 1995; Kearney, 2002; Lecointre and Deleporte, 2005) and supertrees (e.g., Rodrigo, 1993, 1996; Slowinski and Page, 1999; Novacek, 2001; Springer and de Jong, 2001; Gatesy et al., 2002).

The main goal of this study is to compare the most widely used supertree methods and to assess which performed better according to a given set of criteria, including a comparison to total evidence trees. We defined the performance of supertrees by a combination of their agreement with the input trees, relationships with other supertrees and total evidence trees, level of resolution, and computational time. When the first criterion is taken into account, ML supertrees conserved a higher proportion of the phylogenetic signal of the input trees than MP supertrees (Fig. 2). This finding can be related to the efficiency of the ML algorithm coupled with successive heuristic searches. The new heuristic search algorithm implemented in RAxML (Stamatakis et al., 2008) allows an increase of bootstrap replicates (in our case 1000) and produces well supported consensus phylogenetic hypotheses. As a consequence, even if the ML input trees are in general less resolved than the MP consensus trees, they more reliably reflect the raw data (in our case the DNA sequences) (Table 1). However, MP input trees should be favoured if the alternative ML input trees are obtained from “classical” ML heuristic searches (i.e., with only one replicate). Indeed, depending on the degree of variation of a given marker, a fully bifurcated phylogenetic tree might be relatively over-resolved, thus leading to potential conflict during supertree reconstruction.

Several methods (e.g., ML/MP MRP BR rev; ML MinFlip BR) produce supertrees with higher agreement with the input trees than the total evidence trees (especially the ML total evidence tree). The supertrees and total evidence trees that minimize the distance to the input trees (e.g., MRP, MinFlip, MP Strict total evidence tree) are based on consensus methods that include a heuristic search option, which widely explores the tree space (Table 2; number of trees retained and CFI). On the other hand, supertrees that have a higher distance to the input trees are obtained by methods that have a less efficient heuristic strategy and produce only a few trees with high CFI index values (i.e., in most cases only one tree; Table 2). Computational power is also a major drawback for several methods. For example, Split fit analyses were conducted with maxswaps=10000 (instead of 1000000 as set by default in CLANN) due to constraints in computational time; the first trials using default parameters were not completed after more than a month of computation. To compensate for

this technical limit, the number of replicates was increased from 10 to 50 and the nsteps value was increased from 5 to 10, but without increasing the output quality; the same topology was recovered with both numbers of replicates (Table 2; Fig. 3). When the relationships between supertrees and total evidence trees as well as their topological agreement are taken into account, the same pattern is observed which ever comparison is performed: MRP, MinFlip and to some extent MinCut clustered together and provide a result consistent with input trees and total evidence trees, whereas AVCON, Split fit and MSS formed a second group much more distant to input trees and total evidence trees (Figs. 2, 3 and S7; Table 2). Based on the dataset presented in this study and the criteria considered, here, MRP and MinFlip methods were the best performing supertree methods. We must point out that computational time and levels of supertree resolution are not directly linked to the performance of the supertree methods and must be considered in combination with the other criteria (see Table 2).

Poor Performance of AVCON, MSS and Split Fit Methods

In the particular case presented here and using the criteria set above, the AVCON, MSS and Split Fit methods did not perform as well as the three other methods examined (MRP, MinFlip and MinCut) in recovering topologies in agreement with the total evidence trees. Unlike the MRP method, which uses the robust and well tested heuristic search option implemented in PAUP*, most supertree methods have their own build-in heuristic search method (e.g. CLANN, Rainbow). In the case of the MinFlip and MinCut methods (especially regarding MMC; Page, 2002), careful attention was taken to the implementation of heuristic searches that maximize the exploration of tree space and minimize computational time (Chen et al., 2003; Chen 2004; Eulenstein et al., 2004; Page 2002). The reasons explaining the relatively poor performance of AVCON, MSS and Split fit methods might be related to the efficiency of their reconstruction method rather than to the supertree algorithms themselves. On the other hand, one should bear in mind that these methods performed as they were designed by minimizing a given distance measure with the set of input trees (e.g., least-

square distance in AVCON), in which case the accuracy with the total evidence trees might not be compatible with this goal. However, based on the results presented in this study, we would recommend a cautious interpretation of supertrees obtained with AVCON, MSS and Split fit methods using currently available software. Nevertheless, it would be worthwhile to re-evaluate the performance of these supertree methods as soon as more efficient searches become available.

MinFlip: the Method with the Most Affinities with Total Evidence Trees?

Eulenstein et al. (2004) implemented a new heuristic algorithm adapted to the flip supertree method, which was able to manage large input trees. The authors used a series of simulations to compare supertrees constructed with the MinFlip algorithm to those built with MinCut (MC and MMC) and MRP. They argued that MinFlip supertrees were far more accurate than MinCut supertrees and at least as accurate as those built with the MRP algorithm. Based on the dataset and criteria presented in this study, similar trends were recovered. Regardless of which criterion was used to reconstruct the input trees, MinFlip supertrees were far more similar to the total evidence trees than those constructed with the AVCON, MSS, Split fit and MinCut methods and gave results similar to MRP-based supertrees (Figs. 2, 3, 4, S7; Table 2). In the “tree of trees”, MP MinFlip supertrees are more closely related to the total evidence trees than ML MinFlip supertrees, although only marginally (Figs. 3, 4, S7). Indeed, MP MinFlip supertrees share 97% of quartets (independently of the kind of MR) with the ML total evidence tree, whereas ML MinFlip supertrees share respectively 94% (ML MinFlip BR) and 92% (ML MinFlip PU) of the quartets (Fig. 4). Among MP MinFlip, the supertree with PU MR obtained the highest score with regard to the topological agreement with the total evidence trees and the lowest distance to the ML total evidence tree (Fig. 4; Table 2). In addition, both supertrees recovered the *Koelreuteria* group as monophyletic, as did two of the three total evidence trees (see Fig. 1 and Material and Methods). Based on the results presented here, the MinFlip method (using either BR or PU MRs) produces supertrees similar to total evidence

trees in terms of their distance to the input trees (Fig. 2) as well as when compared with all trees (Figs. 3, S7). Further analyses are required to understand the implications of the PU MR on the flip algorithm. However, one possible explanation for the better performance of MinFlip supertrees based on PU MR might be related to the number of states scored as “1” in the MR. Since PU MR have considerably less states scored as “1”, this reduction could significantly reduce the number of possible flips and consequently improve the results.

MRP Method: an Equilibrium between Matrix Representation, Type of Parsimony & Characters Weights

Previous studies based on selected examples (Bininda-Emonds and Bryant, 1998; Ronquist, 1996) and empirical data (Salamin et al., 2002) have shown that the PU MR [which according to Purvis (1995b) aims to remove redundant information; see above] might not always meet its stated objectives and must be used with caution. Results based on the Sapindaceae dataset indicate that the impact of the PU MR on the resulting MRP supertrees is directly related to the type of parsimony used. All the criteria used here reveal that all supertrees based on MRP PU using reversible parsimony performed poorly compared to the other MRP supertrees (Figs. 2-4, S7; Table 2). However, those differences are counterbalanced when the PU matrices are analysed with irreversible parsimony as shown in the “tree of trees” where MRP BR/PU irrev supertrees clustered together (Figs. 3, S7). This increase of accuracy might indicate that the PU matrices contain enough signal when the irreversible parsimony is applied and that in this case, they behave similarly as BR matrices with irreversible parsimony. The poor performance of the MRP PU rev supertrees might be related to the high amount of missing data in the MRs preventing the MRP algorithm to find a proper solution [increase of 2.50 times (from 33% to 82.5% in ML PU) and 2.53 times (from 33.8% to 85.4% in MP PU) of missing data in PU MR compared to the BR MR]. This result concurred with Ronquist (1996) and Bininda-Emonds and Bryants (1998) who argued that the Purvis coding scheme did not remove redundant information but rather essential information, at least when analysed with reversible parsimony (see below). On the other

hand, the performance of MRP BR supertrees seems to be influenced less by the type of parsimony (Figs. 3, 4, S7). The use of irreversible parsimony was initially proposed by Bininda-Emonds and Bryant (1998) to prevent the creation of clades only supported by zeros (in MR, only the “1” represents a phylogenetic signal). Whereas Bininda-Emonds and Sanderson (2001) failed to demonstrate significant differences between reversible and irreversible parsimonies in the context of MRP analyses, Salamin et al. (2002) showed a better agreement of MRP supertrees based on irreversible parsimony with their MP total evidence tree. In our study, we did not obtain a significant difference between the two parsimonies [all the MRP supertrees (except those based on MRP PU rev) have quartet distances to the ML total evidence tree inferior or slightly higher to 0.1; Fig. 4] and attribute the poor performance of MRP PU rev supertrees to the amount of missing data rather than to the creation of unsupported clades. As mentioned above, Ronquist (1996) argued that there is no redundant information in the MR and that the problem Purvis described is related to the size of the input trees. In order to remove or decrease potential misleading relationships in the MR (introduced during the compilation of the input trees), Ronquist (1996) and later Salamin et al. (2002) proposed to weight the characters according to the bootstrap support of each node or inversely weighting each tree (more precisely the characters corresponding to the tree) according to the number of its nodes. In this study, no real tendency in the compatibility with the ML total evidence tree was observed when the characters are weighted or not (Fig. 4). Three patterns were observed according to the compatibility with the ML total evidence tree (Fig. 4): i) no difference between weighted and unweighted trees (ML MRP BR rev), ii) increase of the compatibility in the weighted trees (e.g., MP MRP BR rev), and iii) decrease of the compatibility in the weighted trees (e.g., MP MRP BR irrev). The second pattern that mostly occurs in MRP rev supertrees might be explained by a reduction of homoplasy in the MR and consequently a “reduction” of the distance to the original data. However, this phenomenon is reversed with most supertrees based on irreversible parsimony – i.e. pattern iii) – as first observed by Salamin et al. (2002), although no evidence currently allows us to provide convincing explanations for this pattern.

CONCLUSION

This work evaluates for the first time the performance of six major supertree reconstruction methods using a large empirical data set. It indicates that supertrees based on MP or ML input trees behave similarly; however when the distance to the input trees is taken into account, ML supertrees better fit the phylogenetic signal of the input trees. Overall, our study highlights the better performance (based on the distance to the input trees, relationships between all trees, topological agreement with total evidence trees, resolution, and computational time) of the MRP, MinFlip and MinCut methods. MRP supertrees seem to fit the phylogenetic signal of the input trees better, whereas MinFlip supertrees are generally more similar to total evidence trees. MinCut supertrees constitute an average solution, less accurate than MRP and MinFlip methods, but much more accurate (according to the criteria used in this study) than AVCON, MSS and Split fit methods. We argue that the poor performance of AVCON, Split fit and MSS methods might be related to the efficiency of the heuristic algorithm rather than to the methods themselves. On the other hand, we highlighted the fact that those supertree methods might have performed exactly as they were intended and are not compatible with the criteria defined in this study. Among MRP supertrees, all criteria support the poor performance of MRP PU rev supertrees compared to all other MRP supertrees. However, when irreversible parsimony is applied, the resulting supertrees are similar to those based on BR MR. Our study did not identify substantial differences between MRP rev or irrev and weighted or unweighted supertrees. Future studies based on empirical data sets should investigate the monophyly of subtrees unique to single input trees to assess the performance of supertrees (see Cotton and Page, 2004). Although the majority of our input trees possessed unique taxa, these do not form monophyletic subtrees, and thus do not allow us to apply this criterion here. Another possible approach to assess the performance of supertrees without producing total evidence trees is based on the recognition of unsupported clades. This can be achieved by using several recently published methods providing support for each node according to their agreement with the input trees (e.g.,

Bininda-Emonds, 2003; Wilkinson et al., 2005b; Burleigh et al., 2006). Based on the analyses conducted here, we would recommend starting the exploration of supertree solutions on the base of the MRP BR algorithm, considering reversible parsimony. This represents an excellent trade-off between computational time and a high compatibility with input trees and total evidence trees. However, the comparative examination of the other supertree methods within the MRP and MinFlip frameworks might substantially enhance the compilation of pertinent solutions, when more computational time is available.

ACKNOWLEDGMENTS

We thank Nils Arrigo (University of Neuchâtel, Switzerland) for constructive discussions on former versions of this manuscript. We wish to thank Phillipe Küpfer and Giorgina Bernasconi (University of Neuchâtel, Switzerland) for their enthusiasm, comments and help. We are also grateful to Hervé Sauquet (Royal Botanic Gardens, Kew, U.K.), Isabel Sanmartin (Real Jardín Botánico-CSIC, Spain) and Martin Callmander (Missouri Botanical Garden, U.S.A) for their valuable comments. We wish to thank Alexis Criscuolo (Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier, France) for providing free efficient programs (e.g., CONCATENATE) and help with Dquad. We thank Pedro Acevedo-Rodriguez (Smithsonian Institute, U.S.A.), Mark Harrington (James Cook University, Australia), Caroline Weckerle (Institute of Systematic Botany, Zürich, Switzerland), the Royal Botanic Gardens, Kew (U.K.), the Missouri Botanical Garden (U.S.A) and Laurent Gautier, Louis Nusbaumer and Patrick Ranirison from the Conservatory and Botanical Garden of Geneva, Switzerland for plant material. This work was financially supported by the laboratory of Evolutionary Botany at the University of Neuchâtel (Switzerland).

REFERENCES

Aho, A. V., Y. Sagiv, T. G. Szymanski, and J. D. Ullman. 1981. Inferring a tree from lowest common ancestors with an application to the optimization of relational expressions. *SIAM J. Comput.* 10:405-421.

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. Pages 267-281 *in* Second International Symposium on Information Theory (A. Kiado, ed.). Budapest.

Baum, B. R. 1992. Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. *Taxon* 41:3-10.

Baum, B. R., and M. A. Ragan. 1993. Reply to A. G. Rodrigo's "A comment on Baum's method for combining phylogenetic trees". *Taxon* 42:637-640.

Bininda-Emonds, O. R. P. 2003. Novel versus unsupported clades: assessing the qualitative support for clades in MRP supertrees. *Syst. Biol.* 52:839-848.

Bininda-Emonds, O. R. P. 2004. New uses for old phylogenies: an introduction to the volume. Pages 3-14 *in* *in* Phylogenetic supertrees: Combining information to reveal the Tree of Life (O. R. P. Bininda-Emonds, ed.). Kluwer Academic, Dordrecht, The Netherlands.

Bininda-Emonds, O. R. P., and H. N. Bryant. 1998. Properties of matrix representation with parsimony analyses. *Syst. Biol.* 47:497-508.

Bininda-Emonds, O. R. P., J. L. Gittleman, and A. Purvis. 1999. Building large trees by combining phylogenetic information: A complete phylogeny of the extant Carnivora (Mammalia). *Biol. Rev.* 74:143-175.

Bininda-Emonds, O. R. P., and M. J. Sanderson. 2001. Assessment of the accuracy of matrix representation with parsimony analysis supertree construction. *Syst. Biol.* 50:565-579.

Bininda-Emonds, O. R. P., J. L. Gittleman, and M. A. Steel. 2002. The (Super)tree of life: Procedures, problems, and prospects. *Ann. Rev. Ecol. Syst.* 33:265-289.

Bininda-Emonds, O. R. P., K. E. Jones, S. A. Price, R. Grenyer, M. Cardillo, M. Habib, A. Purvis, and J. L. Gittleman. 2003. Supertrees are a necessary not-so-evil: A comment on Gatesy et al. *Syst. Biol.* 52:724-729.

Buerki, S., F. Forest, P. Acevedo-Rodríguez, M. W. Callmander, J. A. A. Nylander, M. Harrington, I. Sanmartín, P. Küpfer, and N. Alvarez. 2009. Plastid and nuclear DNA markers reveal intricate relationships at subfamilial and tribal levels in the soapberry family (Sapindaceae). *Mol. Phylogenet. Evol.* 51: 238—258.

Burleigh, J. G., A. C. Driskell, and M. J. Sanderson. 2006. Supertrees bootstrapping methods for assessing phylogenetic variation among genes in genome-scale data sets. *Syst. Biol.* 55:426-440.

Cardillo, M., O. R. P. Bininda-Emonds, E. Boakes, and A. Purvis. 2004. A species-level phylogenetic supertree of marsupials. *Journal of Zoology* 264:11-31.

Chen, D., L. Diao, O. Eulenstein, D. Fenandez-Baca, and M. J. Sanderson. 2003. Flipping: A supertree construction method. Pages 135-160 *in* Bioconsensus (M. Janowitz, F.-J. Lapointe, F. R. McMorris, B. Mirkin, and F. S. Roberts, eds.). DIMACS series in discrete mathematics and theoretical computer science. American Mathematical Society, Providence, Rhode Island.

Chen, D. 2004. Rainbow: a toolbox for phylogenetic supertree construction and analysis. *Bioinformatics* 20:2872-2873.

Colless, D. H. 1981. Predictivity and stability in classifications: Some comments on recent studies. *Syst. Biol.* 30:325-331.

Cotton, J. A. and R. D. M. Page. 2004. Tangled trees from multiple markers: reconciling conflict between phylogenies to build molecular supertrees. Pages 107–126 *in* *Phylogenetic supertrees: Combining information to reveal the Tree of Life* (O. R. P. Bininda-Emonds, ed.). Kluwer Academic, Dordrecht, The Netherlands.

Cotton, J. A. and M. Wilkinson. 2007. Majority-rule supertrees. *Syst. Biol.* 56:445-452.

Creevey, C. J., D. A. Fitzpatrick, G. K. Philip, R. J. Kinsella, M. J. O'Connell, M. M. Pentony, S. A. Travers, M. Wilkinson, and J. O. McInerney. 2004. Does a tree-like phylogeny only exist at the tips in the prokaryotes? *Proc. R. Soc. B* 271:2552–2558.

Creevey, C. J., and J. O. McInerney. 2005. Clann: investigating phylogenetic information through supertree analyses. *Bioinformatics* 21:390-392.

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

Estabrook, G. F. 1992. Evaluating undirected positional congruence of individual taxa between two estimates of the phylogenetic tree for a group of taxa. *Syst. Biol.* 41:172-177.

Estabrook, G. F., F. R. McMorris, and C. A. Meacham. 1985. Comparison of undirected phylogenetic trees based on subtrees of four evolutionary units. *Syst. Zool.* 34:193-200.

Eulenstein, O., D. H. Chen, J. G. Burleigh, D. Fernandez-Baca, and M. J. Sanderson. 2004. Performance of flip supertree construction with a heuristic algorithm. *Syst. Biol.* 53:299-308.

Farris, J. S. 1973. Comparing the shapes of taxonomic trees. *Syst. Zool.* 22:50-54.

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

Felsenstein, J. 1993. PHYLIP: Phylogenetic inference package, version 3.5c. University of Washington, Seattle.

Gatesy, J., C. Matthee, R. Desalle, and C. Hayashi. 2002. Resolution of a supertree/supermatrix paradox. *Syst. Biol.* 51:652-664.

Gordon, A. D. 1986. Consensus Supertrees: The synthesis of rooted trees containing overlapping sets of labeled leaves. *J. Classif.* 3:31-39.

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. *Syst. Bot.* 30:366-382.

Huelsenbeck, J. P. 1991. When are fossils better than extant taxa in phylogenetic analysis? *Syst. Zool.* 40:458-469.

Jones, K. E., A. Purvis, A. MacLarnon, O. R. P. Bininda-Emonds, and N. B. Simmons. 2002. A phylogenetic supertree of the bats (Mammalia: Chiroptera). *Biol. Rev.* 77:223-259.

Kearney, M. 2002. Fragmentary Taxa, Missing Data, and Ambiguity: Mistaken Assumptions and Conclusions. *Syst. Biol.* 51:369-381.

Kluge, A. G. 1989. A Concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Syst. Zool.* 38:7-25.

Lapointe, F. J., and G. Cucumel. 1997. The average consensus procedure: Combination of weighted trees containing identical or overlapping sets of taxa. *Syst. Biol.* 46:306-312.

Lapointe, F. J., J. A. W. Kirsch, J. M. Hutcheon. 1999. Total evidence, consensus, and bat phylogeny: a distance-based approach. *Mol. Phylogenet. Evol.* 11:55-66.

Lapointe, F.-J., and C. Levasseur. 2004. Everything you always wanted to know about the average consensus, and more. Pages 87–105 *in* *Phylogenetic supertrees: Combining information to reveal the Tree of Life* (O. R. P. Bininda-Emonds, ed.). Kluwer Academic, Dordrecht, The Netherlands.

Lecointre, G., and P. Deleporte. 2005. Total evidence requires exclusion of phylogenetically misleading data. *Zoologica Scripta* 34:101-117.

Muellner, A. N., D. D. Vassiliades, and S. S. Renner. 2007. Placing Biebersteiniaceae, a herbaceous clade of Sapindales, in a temporal and geographic context. *Pl. Syst. Evol.* 266:233-252.

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.) London, Academic Press.

Nixon, K. C. 1999. The Parsimony Ratchet, a new method for rapid parsimony analysis. *Cladistics* 15:407-414.

Novacek, M. J. 2001. Mammalian phylogeny: Genes and supertrees. *Curr. Biol.* 11:R573-R575.

Page, R. D. M. 2002. Modified mincut supertrees. *Lecture Notes in Computer Science* 2452:537-551.

Price, S. A., O. R. P. Bininda-Emonds, and J. L. Gittleman. 2005. A complete phylogeny of the whales, dolphins and even-toed hoofed mammals (Cetartiodactyla). *Biol. Rev.* 80:445-473.

Purvis, A. 1995a. A composite estimate of primate phylogeny. *Phil. Trans. R. Soc. Lond. B* 348:405-421.

Purvis, A. 1995b. A modification to Baum and Ragan's method for combining phylogenetic trees. *Syst. Biol.* 44:251-255.

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

Ragan, M. A. 1992. Phylogenetic inference based on matrix representation of trees. *Mol. Phylogenet. Evol.* 1:53-58.

Robinson, D. F., and L. R. Foulds. 1981. Comparison of phylogenetic trees. *Math. Biosci.* 53:131-147.

Rodrigo, A. G. 1993. A Comment on Baum method for combining phylogenetic trees. *Taxon* 42:631-636.

Rodrigo, A. G. 1996. On combining cladograms. *Taxon* 45:267-274.

Ronquist, F. 1996. Matrix representation of trees, redundancy, and weighting. *Syst. Biol.* 45:247-253.

Salamin, N., T. R. Hodkinson, and V. Savolainen. 2002. Building supertrees: An empirical assessment using the grass family (Poaceae). *Syst. Biol.* 51:136-150.

Salamin, N., M. W. Chase, T. R. Hodkinson, and V. Savolainen. 2003. Assessing internal support with large phylogenetic DNA matrices. *Mol. Phylogenet. Evol.* 27:528–539.

Sanderson, M. J., A. Purvis and C. Henze. 1998. Phylogenetic supertrees: Assembling the trees of life. *Trends Ecol. Evol.* 13:105-109.

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 *rbcL* gene sequences. *Kew Bull.* 55:257–309.

Semple, C., and M. Steel. 2000. A supertree method for rooted trees. *Discrete Appl. Math.* 105:147-158.

Sikes, D. S., and P. O. Lewis. 2001. PAUPRAT, University of Connecticut, Storrs.

Slowinski, J. B., and R. D. M. Page. 1999. How should species phylogenies be inferred from sequence data? *Syst. Biol.* 48:814-825.

Springer, M. S., and W. W. deJong. 2001. Phylogenetics - Which mammalian supertree to bark up? *Science* 291:1709-1711.

Stamatakis, A. 2006. RAxML-VI-HPC: Maximum Likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688-2690.

Stamatakis, A., P. Hoover and J. Rougemont. 2008. A rapid bootstrap algorithm for the RAxML Web-servers. *Syst. Biol.* 57:758-771.

Stoner, C. J., O. R. P. Bininda-Emonds, and T. M. Caro. 2003. The adaptive significance of coloration in lagomorphs. *Biol. J. Linn. Soc.* 79:309-328.

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

Thorne, R. F. 2000. The classification and geography of the flowering plants: dicotyledons of the class Angiospermae. *Bot. Rev.* 66: 441–647.

Umadevi, I., and M. Daniel. 1991. Chemosystematics of the Sapindaceae. *Feddes Repertorium* 102: 607–612.

Wiens, J.J., and T. W., Reeder. 1995. Combining data sets with different numbers of taxa for phylogenetic analysis. *Syst. Biol.* 44:548-558.

Wilkinson, M., J. L. Thorley, D. T. J. Littlewood, and R. A. Bray. 2001. Towards a phylogenetic supertree of Platyhelminthes? Pages 292-301 *in* Interrelationships of the Platyhelminthes (D. T. J. Littlewood and R. A. Bray, eds.). Taylor and Francis, London.

Wilkinson, M., J. A. Cotton, C. Creevey, O. Eulenstein, S. R. Harris, F. J. Lapointe, C.

Levasseur, J. O. McInerney, D. Pisani, and J. L. Thorley. 2005a. The shape of supertrees to come: Tree shape related properties of fourteen supertree methods. *Syst. Biol.* 54:419-431.

Wilkinson, M., D. Pisani, J. A. Cotton, and I. Corfe. 2005b. Measuring support and finding unsupported relationships in supertrees. *Syst. Biol.* 54:823-831.

Wojciechowski, F., M. J. Sanderson, P. Steele, and A. Liston. 2000. Molecular phylogeny of the "temperate herbaceous tribes" of papilionoid legumes: A supertree approach. Pages 277-298 *in* *Advances in Legume Systematics* (P. Herendeen and A. Bruneau, eds.). Royal Botanic Gardens, Kew, U.K.

FIGURES CAPTIONS

Figure 1. Simplified phylogenetic tree resulting from the maximum likelihood total evidence analysis. Monophyletic clades used to assess the performance of supertree methods are indicated by vertical bars (see text for more details).

Figure 2. Compatibility of each supertree and total evidence trees with the input trees assessed using the average NPM distance (see text). A. Supertrees reconstructed with MP input trees and MP total evidence trees; B. Supertrees reconstructed with ML input trees and the ML total evidence tree. Abbreviations are explained in the text.

Figure 3. NJ “tree of trees” based on pairwise quartet distance between the supertrees and total evidence trees. Abbreviations: black circles refer to MP trees, whereas white circles refer to ML trees. See in the text for definitions and abbreviations of supertree methods.

Figure 4. Histogram showing the quartet (in white) and NPM (in black) distances to the ML total evidence tree for all the supertrees and MP total evidence trees. Abbreviations are explained in the text.

TABLES

Table 1. Characteristics of single-gene matrices used to reconstruct input trees. Characteristics of the supermatrix used to reconstruct total evidence trees are also summarized. Abbreviations: CFI = consensus fork index; IGS = intergenic spacer; PI = potentially parsimony-informative; * = CFI of the MP strict/majority-rule total evidence tree. See text for explanations.

Dataset	No. sequences	Seq. unique to the dataset	Alignment length	No. constant characters (%)	No. variable characters (%)	No. PI characters (%)	No. trees retained	Tree length (step)	CFI (MP consensus tree)	CFI (ML bootstrap consensus tree)	No. shared sequences								
											<i>ITS</i>	<i>matK</i>	<i>rpoB</i>	<i>trnD-trnT</i> IGS	<i>trnK-matK</i> IGS	<i>trnL</i> intron	<i>trnL-trnF</i> IGS	<i>trnS-trnG</i> IGS	
<i>ITS</i>	172	6	1234	584 (47.3)	653 (52.9)	491 (39.8)	214	4666	0.994	0.718	-								
<i>matK</i>	175	26	1614	1062 (65.7)	552 (34.2)	392 (24.3)	1967	1185	0.694	0.618	122	-							
<i>rpoB</i>	199	4	363	222 (61.1)	141 (38.8)	104 (28.7)	2010	281	0.401	0.406	165	144	-						
<i>trnD-trnT</i> IGS	102	0	1925	965 (50.1)	960 (49.9)	505 (26.2)	1975	1858	0.830	0.760	85	86	102	-					
<i>trnK-matK</i> IGS	133	0	931	488 (52.4)	443 (47.6)	277 (29.8)	1703	954	0.718	0.542	112	132	131	80	-				
<i>trnL</i> intron	192	2	773	454 (58.7)	319 (41.2)	194 (25.1)	1500	687	0.658	0.489	154	143	182	102	127	-			
<i>trnL-trnF</i> IGS	189	0	661	328 (49.6)	333 (50.4)	217 (32.8)	667	705	0.631	0.524	154	142	182	102	127	189	-		
<i>trnS-trnG</i> IGS	80	0	2156	1577 (73.1)	579 (26.8)	227 (10.5)	1820	945	0.766	0.623	69	73	80	63	66	80	80	-	
Supermatrix	1242	-	9657	5681 (58.8)	3976 (41.2)	2404 (24.9)	949	11526	0.614/0.909*	1.000	172	175	199	102	133	192	189	80	

Table 2. Characteristics of supertrees and their topological agreement with total evidence trees. Number of criteria per clade are indicated between brackets. See text for more details.

Supertrees	No. Trees retained	CFI	Time (hours:minutes)	Topological agreement with total evidence trees				TOTAL (%)
				Monophyly of Sapindaceae	Clade A (3)	Clade B (3)	Clade C (9)	
MP MRP BR rev	1034	0.621	01:40	1	3	0	7	11 (68.75)
MP MRP BR rev nodes	80	0.858	02:08	1	3	1	7	12 (75)
MP MRP BR rev boot	604	0.7	02:30	1	3	1	7	12 (75)
MP MRP BR irrev	970	0.68	15:57	1	3	0	6	10 (62.5)
MP MRP BR irrev nodes	248	0.755	24:00	1	3	0	7	11 (68.75)
MP MRP BR irrev boot	588	0.693	43:05	1	3	1	7	12 (75)
MP MRP PU rev	92	0.942	03:33	0	0	1	4	5 (31.25)
MP MRP PU rev nodes	99	0.933	04:54	0	0	1	4	5 (31.25)
MP MRP PU rev boot	87	0.942	03:09	0	0	2	6	8 (50)
MP MRP PU irrev	801	0.573	39:25	1	3	0	6	10 (62.5)
MP MRP PU irrev nodes	1043	0.680	56:14	1	3	0	7	11 (68.75)
MP MRP PU irrev boot	166	0.739	81:21	1	3	1	6	11 (68.75)
ML MRP BR rev	1127	0.571	01:04	1	3	1	6	11 (68.75)
ML MRP BR rev nodes	1204	0.608	00:45	1	3	1	7	12 (75)
ML MRP BR rev boot	775	0.633	01:32	1	3	1	6	11 (68.75)
ML MRP BR irrev	832	0.598	06:04	1	3	3	7	14 (87.5)
ML MRP BR irrev nodes	457	0.676	11:01	1	3	1	6	11 (68.75)
ML MRP BR irrev boot	1064	0.627	17:41	1	3	3	6	13 (81.25)
ML MRP PU rev	364	0.217	00:54	0	0	0	5	5 (31.25)
ML MRP PU rev nodes	69	0.946	01:26	0	0	1	4	5 (31.25)
ML MRP PU rev boot	82	0.912	01:31	0	0	0	5	5 (31.25)
ML MRP PU irrev	1035	0.515	16:23	1	3	3	6	13 (81.25)
ML MRP PU irrev nodes	223	0.739	25:56	1	3	1	7	12 (75)
ML MRP PU irrev boot	468	0.610	35:22	1	3	1	5	10 (62.5)
MP MinFlip BR	10	0.963	04:25	1	3	1	5	10 (62.5)
MP MinFlip PU	10	0.925	04:39	1	3	2	9	15 (93.75)
ML MinFlip BR	10	0.979	03:39	1	3	1	6	11 (68.75)
ML MinFlip PU	10	0.979	03:40	1	3	1	7	12 (75)
MP MinCut MC	1	1.000	00:02	1	0	0	0	1 (6.25)
MP MinCut MMC	1	1.000	00:02	1	0	0	3	4 (25)
ML MinCut MC	1	1.000	00:03	1	0	0	1	2 (12.5)
ML MinCut MMC	1	1.000	00:03	1	2	1	2	6 (37.5)
MP AVCON FITCH	1	1.000	23:45	0	0	0	0	0 (0)
MP AVCON NJ	1	1.000	00:02	0	0	0	0	0 (0)
ML AVCON FITCH	1	1.000	23:50	0	0	0	0	0 (0)
ML AVCON NJ	1	1.000	00:02	0	0	0	0	0 (0)
MP Split fit equal 10	2	1.000	95:30	1	0	0	0	1 (6.25)
MP Split fit norm 10	1	1.000	95:55	1	0	0	0	1 (6.25)
MP Split fit equal 50	1	1.000	323:10	1	0	0	0	1 (6.25)
MP Split fit norm 50	1	1.000	323:10	1	0	0	0	1 (6.25)
ML Split fit equal 10	2	0.983	95:30	1	0	0	0	1 (6.25)
ML Split fit norm 10	1	1.000	95:55	1	0	0	0	1 (6.25)
ML Split fit equal 50	2	0.983	323:10	1	0	0	0	1 (6.25)
ML Split fit norm 50	1	1.000	323:10	1	0	0	0	1 (6.25)
MP MSS equal 10	1	0.996	23:10	0	0	0	0	0 (0)
MP MSS norm 10	1	1.000	23:10	0	0	0	0	0 (0)
MP MSS equal 50	1	0.996	118:45	0	0	0	0	0 (0)

Supertrees	No. Trees retained	CFI	Time (hours:minutes)	Topological agreement with total evidence trees				
				Monophyly of Sapindaceae	Clade A (3)	Clade B (3)	Clade C (9)	TOTAL (%)
MP MSS norm 50	1	0.996	118:45	0	0	0	0	0 (0)
ML MSS equal 10	1	1.000	23:10	0	0	0	0	0 (0)
ML MSS norm 10	1	1.000	23:10	0	0	0	0	0 (0)
ML MSS equal 50	1	1.000	118:45	0	0	0	0	0 (0)
ML MSS norm 50	1	0.996	118:45	0	0	0	0	0 (0)

Appendix. Voucher information and GenBank accession numbers for taxa used in the phylogenetic analysis of family Sapindaceae s.l. (including outgroups). Abbreviations: ANH = Andong National University, South Korea; BBG = Bogor Botanic Garden, India, living collections; CSIRO = CSIRO Arboretum, Australia; G = Conservatoire et Jardin Botanique de la ville de Genève, Switzerland; JCT = James Cook University of Northern Queensland, Australia; K = Royal Botanic Gardens, Kew, U.K.; NEU = Neuchâtel, Switzerland; MO = Missouri Botanical Garden, U.S.A.; P = Muséum d'Histoire Naturelle, France; RBG = Royal Botanic Gardens, Kew, U.K., living collections; US = Smithsonian Institution, U.S.A and Z = University of Zürich, Switzerland.

Genera	species	Author	Voucher	Herbarium	Country	GenBank accession number							
						ITS	matK	rpoB	trnD-trnT	trnK-matK	trnL	trnL-F	trnS-trnG
<i>Acer</i>	<i>campestre</i>	L. (Siebold & Zucc.)	<i>s.n.</i>	-	Spain	-	AJ438796	-	-	AJ438796	-	-	-
<i>Acer</i>	<i>cissifolium</i>	K. Koch	-	-	-	AF241483	-	-	-	-	-	-	-
<i>Acer</i>	<i>erianthum</i>	Schwer.	<i>Chase 19983</i>	K	China	EU720501	-	EU720843	EU720980	-	EU721271	EU721459	-
<i>Acer</i>	<i>glabrum</i>	Torr.	<i>Morris Arb. 93-277B</i>	MO	-	-	-	-	-	-	DQ978532	DQ978532	-
<i>Acer</i>	<i>henryi</i>	Pax	-	-	-	AF401141	-	-	-	-	-	-	-
<i>Acer</i>	<i>opalus</i>	Mill.	<i>Grimm GG01305</i>	-	-	AM238317	-	-	-	-	-	-	-
<i>Acer</i>	<i>platanoides</i>	L.	-	-	-	-	AJ438788	-	-	AJ438788	-	-	-
<i>Acer</i>	<i>saccharum</i>	Marshall	<i>Chase 106</i>	K	Cult. source, Orange Co.	EU720502	-	EU720844	-	-	EU721272	EU721460	-
<i>Aesculus</i>	<i>chinensis</i>	Bunge	<i>Xiang305</i>	-	-	-	AY724267	-	-	-	-	-	-
<i>Aesculus</i>	<i>flava</i>	Sol.	<i>98-48</i>	-	-	-	AY968670	-	-	-	-	-	-
<i>Aesculus</i>	<i>glabra</i>	Willd.	<i>JC Raulston Accn. No. 960612</i>	-	-	-	AY968671	-	-	-	-	-	-
<i>Aesculus</i>	<i>hippocastanum</i>	L. (Wall. ex Cambess.) Hook.	<i>livingcollection6911289263</i>	-	-	-	AY724266	-	-	-	-	-	-
<i>Aesculus</i>	<i>indica</i>	Cambess.) Hook.	<i>Chase 19987</i>	K	India	EU927392	-	EU720845	EU720981	-	EU721273	EU721461	-
<i>Aesculus</i>	<i>pavia</i>	Castigl.	<i>Chase 503</i>	K	-	-	-	EU720846	EU720982	-	EU721274	EU721462	-
<i>Aesculus</i>	<i>wangii</i>	Hu	-	-	-	-	-	-	-	-	AF411085	AF411085	-
<i>Alectryon</i>	<i>connatus</i>	Radlk.	<i>Chase 2047</i>	K	Australia	EU720415	EU720577	EU720732	EU720928	EU721025	EU721169	EU721357	EU721534
<i>Alectryon</i>	<i>coriaceus</i>	Radlk.	<i>Edwards KE59</i>	JCT	Australia	-	EU720599	EU720756	-	EU721039	EU721192	EU721380	EU721548
<i>Alectryon</i>	<i>excelsus</i>	Gaertn.	-	-	-	EF635451	-	-	-	-	-	-	-
<i>Allophyllus</i>	<i>arboreus</i>	Choux	<i>Rakotovao 2812</i>	MO	Madagascar	EU720508	-	EU720852	-	-	-	-	-
<i>Allophyllus</i>	<i>arboreus</i>	Choux	<i>Ravelonarivo 1618</i>	MO	Madagascar	EU720515	EU720665	EU720859	-	EU721102	-	-	-
<i>Allophyllus</i>	<i>arboreus</i>	Choux	<i>Rakotovao 3131</i>	MO	Madagascar	EU720531	EU720681	EU720879	-	EU721118	-	-	-
<i>Allophyllus</i>	<i>cobbe</i>	(L.) Rauesch	<i>Callmander 462</i>	MO	Madagascar	EU720517	-	EU720861	-	-	-	-	-
<i>Allophyllus</i>	<i>longipes</i>	Radlk.	-	-	-	-	-	-	-	-	AY207572	AY207572	-
<i>Allophyllus</i>	<i>natalensis</i>	De Winter	<i>Edwards KE227</i>	JCT	South Africa	-	AY724268	-	-	-	-	-	-
<i>Allophyllus</i>	<i>sp.</i>	Radlk.	<i>Edwards KE273</i>	JCT	Tanzania	-	AY724269	-	-	-	-	-	-
<i>Allophyllus</i>	<i>trichodesmus</i>	Radlk.	<i>Rakotovao 2897</i>	MO	Madagascar	EU720511	-	EU720855	-	-	-	-	-
<i>Allophyllus</i>	<i>trichodesmus</i>	Radlk.	<i>Ravelonarivo 1619</i>	MO	Madagascar	EU720516	-	EU720860	-	-	-	-	-
<i>Amesiodendron</i>	<i>chinensis</i>	(Merr.) Hu	<i>Yuan s.n.</i>	NEU	China	EU720403	-	EU720718	EU720917	-	EU721155	EU721344	EU721525
<i>Arfeuillea</i>	<i>arborescens</i>	Pierre	<i>Chase 2122</i>	K	Bogor, BG	EU720461	EU720629	EU720793	EU720962	EU721067	EU721229	EU721417	-
<i>Arytera</i>	<i>divaricata</i>	F. Muell.	<i>Edwards KE010</i>	JCT	Australia	-	AY724271	-	-	-	-	-	-
<i>Arytera</i>	<i>littoralis</i>	Blume	<i>Yuan s.n.</i>	NEU	China	EU720405	EU720566	EU720720	EU720919	EU721018	EU721157	EU721346	EU721527
<i>Arytera</i>	<i>littoralis</i>	Blume	<i>Chase 2123</i>	K	Bogor, BG	EU720462	EU720630	EU720794	EU720963	EU721068	EU721230	EU721418	-
<i>Arytera</i>	<i>microphylla</i>	(Benth.) Radlk.	<i>Edwards KE60</i>	JCT	Australia	-	AY724270	-	-	-	-	-	-
<i>Atalaya</i>	<i>alata</i>	(Sim) H. Forbes	<i>Edwards KE228</i>	JCT	South Africa	EU720425	EU720593	EU720748	EU720939	EU721036	EU721184	EU721372	EU721543

Appendix. Continued

Genera	species	Author	Voucher	Herbarium	Country	GenBank accession number							
						ITS	matK	rpoB	trnD-trnT	trnK-matK	trnL	trnL-F	trnS-trnG
<i>Atalaya</i>	<i>angustifolia</i>	S.T. Reynolds	<i>West 5349</i>	ANH	Australia	-	EU720569	EU720723	-	-	EU721160	EU721348	-
<i>Atalaya</i>	<i>capense</i>	R.A. Dyer	<i>Edwards KE 509</i>	JCT	South Africa	EU720429	-	EU720752	-	-	EU721188	EU721376	-
<i>Atalaya</i>	<i>salicifolia</i>	(DC.) Blume	<i>Edwards KE58</i>	JCT	Australia	-	AY724272	-	-	-	-	-	-
<i>Athyana</i>	<i>weinmannifolia</i>	(Griseb.) Radlk. Acev.-Rodr. & Ferrucci	<i>Pennington 17581</i>	MO	Peru	EU720487	EU720649	EU720824	EU720975	EU721086	EU721257	EU721445	EU721576
<i>Averrhoidium</i>	<i>dalyi</i>		<i>Weckerle 00/03/18-1/1</i>	Z	Peru	EU720495	-	EU720836	-	-	EU721268	EU721456	-
<i>Beguea</i>	<i>apetala</i>	Capuron	<i>Buerki 149</i>	NEU	Madagascar	EU720491	EU720652	EU720828	EU720978	EU721089	EU721261	EU721449	-
<i>Beguea</i>	<i>apetala</i>	Capuron	<i>Vary 40</i>	MO	Madagascar	EU720512	EU720663	EU720856	-	EU721100	EU721281	EU721469	-
<i>Billia</i>	sp.		<i>Hammel 20075</i>	-	-	-	AY724275	-	-	-	-	-	-
<i>Blighia</i>	<i>sapida</i>	K.D. Koenig	<i>Edwards KE86</i>	JCT	West Africa	EU720416	EU720578	EU720733	EU720929	EU721026	EU721170	EU721358	EU721535
<i>Blighia</i>	<i>unijugata</i>	Baker	<i>Edwards KE274</i>	JCT	Tanzania	-	AY724276	-	-	-	-	-	-
<i>Blomia</i>	<i>prisca</i>	(Standl.) Lundell	<i>Acevedo 12242</i>	US	Mexico, Yucatan	EU720444	EU720611	EU720772	-	EU721050	EU721208	EU721396	-
<i>Bridgesia</i>	<i>incisifolia</i>	Bertero ex Cambess.	<i>Killip & Pisano 39778</i>	K	Chile	EU720476	EU720645	EU720811	EU720973	EU721082	EU721247	EU721435	-
<i>Cardiospermum</i>	<i>grandiflorum</i>	Sw.	<i>Edwards KE207</i>	JCT	-	-	EU720588	EU720743	EU720935	-	EU721179	EU721367	-
<i>Cardiospermum</i>	<i>microcarpum</i>	Kunth	<i>Yuan s.n.</i>	NEU	China	-	-	EU720712	EU720911	-	EU721149	EU721338	-
<i>Cardiospermum</i>	sp.	(F. Muell.) F. Muell.	<i>Edwards KE88</i>	JCT	China	EU720399	-	EU720713	EU720912	-	EU721150	EU721339	-
<i>Castanospora</i>	<i>alphanthii</i>		<i>Edwards KE88</i>		Australia	-	AY724279	-	-	-	-	-	-
<i>Chytranthus</i>	<i>carneus</i>	Radlk.	<i>Chase 2868</i>	RBG	-	EU720477	EU720646	EU720812	EU720974	EU721083	EU721248	EU721436	EU721575
<i>Chytranthus</i>	<i>prieurianus</i>	Baill.	<i>Edwards KE272</i>	JCT	Tanzania	-	AY724280	-	-	-	-	-	-
<i>Conchopetalum</i>	<i>brachysepalum</i>	Capuron	<i>Rabarimanarivo 8</i>	MO	Madagascar	EU720530	EU720680	EU720877	-	EU721117	EU721299	EU721487	EU721586
<i>Cossinia</i>	<i>pinnata</i>	Comm. ex Lam.	<i>Lorence 4510</i>	MO	Mauritius	-	-	EU720820	-	-	EU721253	EU721441	-
<i>Cubilia</i>	<i>cubili</i>	(Blanco) Adelb.	<i>Chase 2125</i>	K	Bogor, BG	EU720463	EU720631	EU720795	EU720964	EU721069	EU721231	EU721419	EU721567
<i>Cupania</i>	<i>dentata</i>	DC.	<i>Acevedo 12241</i>	US	Mexico, Yucatan	EU720523	EU720670	EU720867	EU720988	EU721107	EU721289	EU721477	EU721581
<i>Cupania</i>	<i>guatemalensis</i>	(Turcz.) Radlk.	<i>Davidse 35743</i>	MO	Belize	-	EU720678	EU720875	EU720993	EU721115	EU721297	EU721485	-
<i>Cupania</i>	<i>hirsuta</i>	Radlk.	<i>Acevedo 1101</i>	US	French Guiana	EU720521	EU720668	EU720865	-	EU721105	EU721287	EU721475	-
<i>Cupania</i>	<i>rubiginosa</i>	(Poir.) Radlk.	<i>Mori 8868</i>	MO	French Guiana	EU720481	-	EU720817	-	-	EU721251	EU721439	-
<i>Cupania</i>	<i>scrobiculata</i>	Rich.	<i>Acevedo 11100</i>	US	French Guiana	EU720524	EU720671	EU720868	EU720989	EU721108	EU721290	EU721478	-
<i>Cupaniopsis</i>	<i>anacardioides</i>	Radlk.	<i>Chase 217</i>	K	Australia	EU720438	EU720605	EU720763	EU720946	EU721045	EU721199	EU721387	EU721552
<i>Cupaniopsis</i>	<i>flagelliformis</i>	(Bailey) Radlk.	<i>Edwards KE42</i>	JCT	Australia	EU720432	EU720598	EU720755	EU720942	-	EU721191	EU721379	EU721547
<i>Cupaniopsis</i>	<i>fruticosa</i>	Radlk.	<i>Munzinger 564</i>	MO	New Caledonia	EU720533	-	EU720881	-	EU721119	EU721302	EU721490	-
<i>Cupaniopsis</i>	sp.		<i>Munzinger 710</i>	MO	New Caledonia	EU720532	-	EU720880	EU720996	-	EU721301	EU721489	EU721587
<i>Cupaniopsis</i>	sp.		<i>Munzinger 1103</i>	MO	New Caledonia	EU720507	EU720660	EU720851	-	EU721097	EU721278	EU721466	-
<i>Deinbollia</i>	<i>borbonica</i>	Scheff.	<i>Edwards KE197</i>	JCT	Tanzania	EU720412	EU720574	EU720729	-	-	EU721166	EU721354	EU721532
<i>Deinbollia</i>	<i>macrocarpa</i>	Capuron	<i>Antilahimena 4293</i>	MO	Madagascar	-	EU720626	EU720790	-	EU721064	EU721226	EU721414	EU721565
<i>Deinbollia</i>	<i>macrocarpa</i>	Capuron	<i>H. Razafindraibe 118</i>	MO	Madagascar	EU720535	EU720683	EU720883	-	EU721121	EU721304	EU721492	EU721589
<i>Deinbollia</i>	<i>macrocarpa</i>	Capuron (E. Mey. ex Arn.) Radlk.	<i>Buerki 144</i> <i>Edwards KE233</i>	NEU JCT	Madagascar South Africa	EU720503	EU720656	EU720847	-	EU721093	EU721275	EU721463	-
<i>Deinbollia</i>	<i>oblongifolia</i>					EU720427	EU720595	EU720750	-	-	EU721186	EU721374	EU721545
<i>Deinbollia</i>	<i>pervillei</i>	(Blume) Radlk.	<i>Phillipson 5919</i>	MO	Madagascar	EU720395	EU720560	EU720708	-	EU721012	EU721145	EU721334	-
<i>Deinbollia</i>	<i>pervillei</i>	(Blume) Radlk.	<i>Callmander 688</i>	MO	Madagascar	EU720514	-	EU720858	-	-	EU721283	EU721471	-
<i>Delavaya</i>	<i>yunnanensis</i>	Franch.	<i>Forrest 20682</i>	MO	China, Yunnan	EU720484	-	EU720821	-	-	EU721254	EU721442	-
<i>Diatenopteryx</i>	<i>sorbifolia</i>	Radlk.	<i>Zardini 43371</i>	MO	Paraguay	EU720534	EU720682	EU720882	-	EU721120	EU721303	EU721491	EU721588
<i>Diatenopteryx</i>	<i>sorbifolia</i>	Radlk.	<i>Tressens 3504</i>	K	Argentina	-	-	EU720810	-	-	EU721246	EU721434	-
<i>Dictyoneura</i>	<i>obtusata</i>	Blume	<i>Edwards KE142</i>	JCT	Australia	EU720428	-	EU720751	-	-	EU721187	EU721375	-
<i>Dilodendron</i>	<i>bipinnatum</i>	Radlk.	<i>Acevedo 11129</i>	US	Bolivia	-	EU720677	EU720874	-	EU721114	EU721296	EU721484	EU721584
<i>Dimocarpus</i>	<i>australianus</i>	Leenh.	<i>Edwards KE34</i>	JCT	Australia	EU720433	-	EU720757	-	-	-	EU721381	-

Appendix. Continued

Genera	species	Author	Voucher	Herbarium	Country	GenBank accession number							
						ITS	matK	rpoB	trnD-trnT	trnK-matK	trnL	trnL-F	trnS-trnG
<i>Dimocarpus</i>	<i>longan</i>	Lour.	<i>Yuan s.n.</i>	NEU	China	-	EU720563	EU720714	EU720913	EU721015	EU721151	EU721340	EU721521
<i>Dimocarpus</i>	<i>longan</i>	Lour.	<i>Edwards KE502</i>	JCT	Asia	-	EU720590	EU720745	-	-	EU721181	EU721369	-
<i>Dimocarpus</i>	<i>longan</i>	Lour.	-	-	Thailand	EF532337	-	-	-	-	-	-	-
<i>Dimocarpus</i>	<i>longan (=Euphoria cinerea)</i>	Lour.	<i>Chase 1351</i>	K	Bogor, BG	-	EU720615	EU720777	EU720953	EU721053	EU721213	EU721401	EU721559
<i>Diploglottis</i>	<i>campbelli</i>	Cheel	<i>Chase 2048</i>	K	Australian, BG	EU720457	EU720624	EU720788	EU720960	EU721062	EU721224	EU721412	-
<i>Diploglottis</i>	<i>diphylostegia</i>	(F. Muell.) Radlk.	<i>Edwards KE001</i>	JCT	Australia	-	AY724287	-	-	-	-	-	-
<i>Mischarytera</i>	sp.	-	<i>Edwards KE159</i>	JCT	Australia	EU720417	EU720579	EU720734	EU720930	EU721027	EU721171	EU721359	-
<i>Diploglottis</i>	<i>smithii</i>	S.T. Reynolds	<i>BG838</i>	BBG	Australia	-	AY724288	-	-	-	-	-	-
<i>Diplokeleba</i>	<i>floribunda</i>	N.E. Br.	<i>Acevedo 11130</i>	US	Bolivia, St Cruz	-	-	EU720773	EU720950	-	EU721209	EU721397	-
<i>Diplopeltis</i>	<i>huegelii</i>	Endl.	<i>Chase 2192</i>	K	Australia	EU720473	EU720642	EU720807	EU720971	EU721079	EU721243	EU721431	-
<i>Dipteronia</i>	<i>sinensis</i>	Oliv.	<i>Chase 502</i>	RBG	-	EU720445	EU720612	EU720774	-	-	EU721210	EU721398	-
<i>Dodonaea</i>	<i>lanceolata</i>	F. Muell.	<i>Edwards KE120</i>	JCT	Australia	-	AY724290	-	-	-	-	-	-
<i>Dodonaea</i>	<i>madagascariensis</i>	Radlk.	<i>Bocksberger GB028</i>	NEU	Madagascar	EU720518	-	EU720862	EU720984	-	EU721284	EU721472	-
<i>Dodonaea</i>	<i>viscosa</i>	Jacq.	<i>Razafitsalama 956</i>	MO	Madagascar	EU720519	EU720666	EU720863	EU720985	EU721103	EU721285	EU721473	-
<i>Dodonaea</i>	<i>viscosa</i>	Jacq.	<i>Merello 1077</i>	MO	Peru	EU720536	EU720684	EU720884	EU720997	EU721122	EU721305	EU721493	-
<i>Dodonaea</i>	<i>viscosa</i>	Jacq.	<i>Yuan s.n.</i>	NEU	China	EU720406	EU720567	EU720721	EU720920	EU721019	EU721158	EU721347	-
<i>Doratoxylon</i>	<i>chouxii</i>	Capuron	<i>Labat JNL3543</i>	P	Madagascar	EU720394	EU720559	EU720707	EU720908	EU721011	EU721144	EU721333	-
<i>Doratoxylon</i>	<i>chouxii</i>	Capuron	<i>Callmander 679</i>	MO	Madagascar	EU720513	EU720664	EU720857	-	EU721101	EU721282	EU721470	-
<i>Elattostachys</i>	<i>apetala</i>	Radlk.	<i>Munzinger 692</i>	MO	New Caledonia	EU720537	EU720685	EU720885	EU720998	EU721123	EU721306	EU721494	EU721590
<i>Elattostachys</i>	<i>apetala</i>	Radlk.	<i>McPherson 18184</i>	MO	New Caledonia	EU720538	EU720686	EU720886	EU720999	EU721124	EU721307	EU721495	EU721591
<i>Elattostachys</i>	<i>megalantha</i>	S.T. Reynolds	<i>Irvine IRV507</i>	CSIRO	Atherton	-	EU720609	EU720768	-	EU721048	EU721204	EU721392	-
<i>Elattostachys</i>	<i>microcarpa</i>	S.T. Reynolds	<i>Edwards KE98</i>	JCT	Australia	EU720409	EU720571	EU720726	-	-	EU721163	EU721351	-
<i>Elattostachys</i>	<i>nervosa</i>	(F. Muell.) Radlk.	<i>Chase 2022</i>	K	Australian, BG	EU720455	EU720622	EU720786	EU720959	EU721060	EU721222	EU721410	EU721563
<i>Elattostachys</i>	sp.	-	<i>Lowry 5650A</i>	MO	New Caledonia	EU720529	EU720679	EU720876	EU720994	EU721116	EU721298	EU721486	EU721585
<i>Eriocoelum</i>	<i>kerstingii</i>	Gilg ex Engl.	<i>Merello 1586</i>	MO	Ghana	EU720539	EU720687	EU720887	EU721000	EU721125	EU721308	EU721496	EU721592
<i>Eriocoelum</i>	<i>microspermum</i>	Radlk.	<i>Bradley 1025</i>	MO	Gabon	EU720540	EU720688	EU720888	EU721001	EU721126	EU721309	EU721497	EU721593
<i>Erythrophysa</i>	<i>aesculina</i>	Baill.	<i>Randrianasolo 625</i>	MO	Madagascar	-	-	-	-	-	EU721329	-	-
<i>Euphorianthus</i>	<i>longifolius</i>	Radlk.	<i>Chase 2126</i>	K	Bogor, BG	EU720464	-	EU720796	-	-	EU721232	EU721420	-
<i>Eurycorymbus</i>	<i>cavalerieri</i>	(H. Lév.) Rehder & Hand.-Mazz. (Wight & Arn.)	<i>Yuan s.n.</i>	NEU	China	EU720404	EU720565	EU720719	EU720918	EU721017	EU721156	EU721345	EU721526
<i>Filicium</i>	<i>decipiens</i>	Thwaites (H. Perrier)	<i>Chase 2128</i>	K	Bogor, BG	EU720466	EU720633	EU720798	-	-	EU721234	EU721422	-
<i>Filicium</i>	<i>longifolium</i>	Capuron	<i>Rabenantonadro 1113</i>	MO	Madagascar	EU720541	-	EU720889	-	-	EU721310	EU721498	-
<i>Filicium</i>	<i>thouarsianum</i>	(A. DC.) Capuron	<i>Antilahimena 5021</i>	MO	Madagascar	EU720493	-	EU720832	-	-	EU721265	EU721453	-
<i>Ganophyllum</i>	<i>falcatum</i>	Blume	<i>Chase 2129</i>	K	Bogor, BG	EU720467	EU720634	EU720799	-	EU721071	EU721235	EU721423	-
<i>Ganophyllum</i>	<i>falcatum</i>	Blume	<i>BH9269</i>	CSIRO	Australia	-	-	-	-	-	EU721330	-	-
<i>Glennia</i>	<i>pervillei</i>	(Baill.) Leenh.	<i>Andriamihajarivo 1053</i>	MO	Madagascar	EU720490	EU720651	EU720827	EU720977	EU721088	EU721260	EU721448	-
<i>Glennia</i>	<i>pervillei</i>	(Baill.) Leenh.	<i>Andriamihajarivo 1025</i>	MO	Madagascar	-	EU720653	EU720829	-	EU721090	EU721262	EU721450	-
<i>Gongrodiscus</i>	<i>bilocularis</i>	H. Turner	<i>Munzinger 749</i>	MO	New Caledonia	EU720542	EU720689	EU720890	-	EU721127	EU721311	EU721499	-
<i>Guindilia</i>	<i>trinervis</i>	Gilles ex Hook.	<i>Chase 802</i>	K	Chile	-	EU720613	EU720775	EU720951	EU721051	EU721211	EU721399	EU721557
<i>Guioa</i>	<i>acutifolia</i>	Radlk.	<i>Edwards KE14</i>	JCT	Australia	-	AY724297	-	-	-	-	-	-
<i>Guioa</i>	<i>glauca</i>	Radlk.	<i>McPherson 18230</i>	MO	New Caledonia	EU720545	EU720692	EU720893	-	EU721130	EU721315	EU721503	-
<i>Guioa</i>	<i>lasioneura</i>	Radlk.	<i>BG1888</i>	BBG	Australia	-	AY724298	-	-	-	-	-	-
<i>Guioa</i>	<i>microsepala</i>	Radlk.	<i>Munzinger 744</i>	MO	New Caledonia	EU720546	EU720693	EU720894	-	EU721131	EU721316	EU721504	EU721596

Appendix. Continued

														GenBank accession number			
Genera	species	Author	Voucher	Herbarium	Country	ITS	matK	rpoB	trnD-trnT	trnK-matK	trnL	trnL-F	trnS-trnG				
<i>Guioa</i>	<i>semiglauca</i>	(F. Muell.) Radlk.	<i>Chase 2058</i>	K	Australian, BG	EU720458	EU720625	EU720789	-	EU721063	EU721225	EU721413	-				
<i>Guioa</i>	<i>villosa</i>	Radlk.	<i>McPherson 18040</i>	MO	New Caledonia	EU720544	EU720691	EU720892	EU721003	EU721129	EU721314	EU721502	EU721595				
<i>Guioa</i>	sp.		<i>Munzinger 945</i>	MO	New Caledonia	EU720505	EU720658	EU720849	-	EU721095	EU721277	EU721465	-				
<i>Handeliidendron</i>	<i>bodinieri</i>	(H. Lév.) Rehder	<i>QYXiang302 / C. Ming 050923</i>	-	-	-	AY724299	-	-	-	EF186776	-	-				
<i>Haplocoelopsis</i>	<i>africana</i>	F.G. Davies	<i>Edwards KE276</i>	JCT	Tanzania	EU720441	EU720608	EU720767	EU720949	-	EU721203	EU721391	EU721555				
<i>Haplocoelum</i>	<i>foliosum</i>	(Hiem) Bullock	<i>Friis 1894</i>	MO	Ethiopia	EU720479	-	EU720815	-	-	EU721250	EU721438	-				
			<i>Edwards KE195</i>	JCT	Tanzania												
<i>Haplocoelum</i>	<i>foliosum</i> subsp. <i>foliosum</i>	(Hiem) Bullock				EU720410	EU720572	EU720727	EU720924	-	EU721164	EU721352	EU721530				
<i>Haplocoelum</i>	<i>gallaense</i>	(Engl.) Radlk.	<i>Edwards KE501</i>	JCT	South Africa	-	EU720583	-	-	-	-	-	-				
<i>Haplocoelum</i>	<i>perrieri</i>	Capuron	<i>Rakotomalaza 1165</i>	MO	Madagascar	EU720396	-	EU720709	EU720909	-	EU721146	EU721335	EU721519				
<i>Harpullia</i>	<i>arborea</i>	(Blanco) Radlk.	<i>Chase 1353</i>	K	Bogor, BG	EU720448	-	EU720779	-	-	EU721215	EU721403	-				
<i>Harpullia</i>	<i>cupanioides</i>	Roxb.	-	-	-	AY2075669	-	-	-	-	-	-	-				
		C. White & Francis	<i>Edwards KE003</i>	JCT	-	-	AY724303	-	-	-	-	-	-				
<i>Hippobromus</i>	<i>pauciflorus</i>	Radlk.	<i>Edwards KE229</i>	JCT	-	-	AY724305	-	-	-	EU721331	EU721517	-				
<i>Hypelate</i>	<i>trifoliata</i>	Sw.	<i>R.Rankin HABJ72057</i>	K	-	-	-	EU720813	-	-	-	-	-				
		(Blume) Blume ex Kalkman															
<i>Jagera</i>	<i>javanica</i>		<i>Chase 2130</i>	K	Bogor, BG	EU720468	EU720635	EU720800	-	EU721072	EU721236	EU721424	EU721569				
			<i>Edwards KE178</i>	JCT	Australia												
<i>Jagera</i>	<i>javanica</i> subsp. <i>australiana</i>	Leenh.				EU720442	-	EU720769	-	-	EU721205	EU721393	EU721556				
			<i>Edwards KE41</i>	JCT	Australia												
<i>Jagera</i>	<i>pseudorhus</i> var. <i>pseudorhus</i> f. <i>pilosiuscula</i>	Radlk.				-	EU720606	EU720764	EU720947	EU721046	EU721200	EU721388	EU721553				
<i>Koelreuteria</i>	<i>paniculata</i>	Laxm.	<i>Harder 5668</i>	MO	Vietnam	EU720548	EU720695	EU720896	-	EU721133	EU721318	EU721506	-				
<i>Koelreuteria</i>	<i>paniculata</i>	Laxm.	<i>Yuan CN2006-3</i>	NEU	China	EU720397	EU720561	EU720710	-	EU721013	EU721147	EU721336	EU721520				
<i>Koelreuteria</i>	<i>paniculata</i>	Laxm.	<i>Wilson 1476</i>	RBG	-	-	AY724308	-	-	-	-	-	-				
<i>Koelreuteria</i>	sp.		<i>Harder 5724</i>	MO	Vietnam	EU720547	EU720694	EU720895	EU721004	EU721132	EU721317	EU721505	-				
<i>Laccodiscus</i>	<i>klaineanus</i>	Pierre ex Engl.	<i>Walters 1269</i>	MO	Gabon	EU720549	EU720696	EU720897	-	EU721134	EU721319	EU721507	-				
<i>Lecaniodiscus</i>	<i>fraxinifolius</i>	Baker	<i>Edwards KE194</i>	JCT	Tanzania	EU720418	EU720580	EU720735	EU720931	EU721028	EU721172	EU721360	EU721536				
<i>Lepiderema</i>	<i>hirsuta</i>	S.T. Reynolds	<i>Edwards KE36</i>	JCT	Australia	EU720435	EU720601	EU720759	-	EU721041	EU721195	EU721383	EU721549				
<i>Lepiderema</i>	<i>pulchella</i>	Radlk.	<i>Chase 2020</i>	K	Australian, BG	EU720454	-	EU720785	EU720958	-	EU721221	EU721409	-				
<i>Lepidopetalum</i>	<i>fructoglabrum</i>	Welzen	<i>Edwards KE139</i>	JCT	Australia	EU720408	-	EU720724	EU720922	-	EU721161	EU721349	EU721528				
<i>Lepisanthes</i>	<i>alata</i>	(Blume) Leenh.	<i>Chase 1355</i>	K	Bogor, BG	EU720450	EU720618	EU720781	-	EU721056	EU721217	EU721405	-				
<i>Lepisanthes</i>	<i>feruginea</i>	(Radlk.) Leenh.	<i>Chase 1354</i>	K	Bogor, BG	EU720449	EU720617	EU720780	-	EU721055	EU721216	EU721404	-				
<i>Lepisanthes</i>	<i>rubiginosa</i>	(Roxb.) Leenh.	<i>Chase 1350</i>	K	Bogor, BG	EU720446	EU720614	EU720776	EU720952	EU721052	EU721212	EU721400	EU721558				
<i>Lepisanthes</i>	<i>senegalensis</i>	(Poir.) Leenh.	<i>Callmander 627</i>	MO	Madagascar	EU720492	EU720654	EU720830	EU720979	EU721091	EU721263	EU721451	EU721577				
<i>Litchi</i>	<i>chinensis</i>	Sonn.	<i>Yuan s.n.</i>	NEU	China	EU720400	EU720564	EU720715	EU720914	EU721016	EU721152	EU721341	EU721522				
<i>Llagunoa</i>	<i>mollis</i>	Kunth	<i>Jaramilloleija 3199</i>	MO	Colombia	EU720482	-	EU720818	-	-	EU721252	EU721440	-				
<i>Llagunoa</i>	<i>nitida</i>	Ruiz & Pav.	<i>Pennington 17552</i>	MO	Peru	EU720486	-	EU720823	-	-	EU721256	EU721444	-				
<i>Loxodiscus</i>	<i>coriaceus</i>	Hook. f.	<i>Bradford 1136</i>	MO	New Caledonia	EU720488	-	EU720825	-	-	EU721258	EU721446	-				
<i>Macphersonia</i>	<i>chapelierii</i>	(Baill.) Capuron	<i>Buerki 138</i>	NEU	Madagascar	EU720459	EU720627	EU720791	EU720961	EU721065	EU721227	EU721415	EU721566				
<i>Macphersonia</i>	<i>gracilis</i>	O. Hoffm.	<i>Rabenantoandro 1081</i>	MO	Madagascar	EU720550	EU720697	EU720898	EU721005	EU721135	EU721320	EU721508	EU721597				
<i>Magonia</i>	<i>pubescens</i>	A. St.-Hil.	<i>Mori 16966</i>	MO	Brazil	EU720483	-	EU720819	-	-	-	-	-				
<i>Majidea</i>	<i>zanguebarika</i>	Kirk ex Oliv.	<i>TH275</i>	MO	Madagascar	EU720552	-	EU720900	EU721006	-	EU721322	EU721510	-				

Appendix. Continued

Genera	species	Author	Voucher	Herbarium	Country	GenBank accession number							
						ITS	matK	rpoB	trnD-trnT	trnK-matK	trnL	trnL-F	trnS-trnG
<i>Matayba</i>	<i>apetala</i>	Radlk.	<i>Acevedo 11929</i>	US	Jamaica	EU720526	EU720674	EU720871	-	EU721111	EU721293	EU721481	EU721583
<i>Matayba</i>	<i>cf. opaca</i>	Radlk.	<i>Acevedo 11118</i>	US	French Guiana	EU720522	EU720669	EU720866	EU720987	EU721106	EU721288	EU721476	EU721580
<i>Matayba</i>	<i>domingensis</i>	(DC.) Radlk.	<i>Taylor 11819</i>	MO	Caribbean	EU720551	EU720698	EU720899	-	EU721136	EU721321	EU721509	EU721598
<i>Matayba</i>	<i>elaeagnoides</i>	Radlk.	<i>Zardini 43278</i>	MO	Paraguay	EU720553	EU720699	EU720901	-	EU721137	EU721323	EU721511	-
<i>Matayba</i>	<i>guianensis</i>	Aubl.	<i>Acevedo 12342</i>	US	French Guiana	EU720527	EU720675	EU720872	-	EU721112	EU721294	EU721482	-
<i>Matayba</i>	<i>laevigata</i>	Radlk.	<i>Acevedo 12357</i>	US	French Guiana	EU720528	EU720676	EU720873	EU720992	EU721113	EU721295	EU721483	-
<i>Melicococcus</i>	<i>bijugatus</i>	Jacq.	<i>Acevedo s.n.</i>	US	Puerto Rico	EU927391	EU720610	EU720771	-	EU721049	EU721207	EU721395	-
<i>Melicococcus</i>	<i>lepidopetalus</i>	Radlk.	<i>Acevedo 11128</i>	US	Bolivia	EU720443	-	EU720770	-	-	EU721206	EU721394	-
		(F.M. Bailey) H. Turner	<i>Edwards KE1302</i>	JCT	Australia	-	-	EU720742	-	-	-	-	-
<i>Mischarytera</i>	<i>lautereriana</i>	(Merr. & L.M. Perry) H. Turner	<i>BH6631</i>	CSIRO	Australia	-	AY724313	-	-	-	-	-	-
<i>Mischarytera</i>	<i>macrobotrys</i>	-	<i>Edwards KE159</i>	JCT	Australia	EU720417	EU720579	EU720734	EU720930	EU721027	EU721171	EU721359	-
<i>Mischocarpus</i>	<i>exangulatus</i>	(F. Muell.) Radlk.	<i>Edwards KE30</i>	JCT	Australia	EU720434	EU720600	EU720758	EU720943	EU721040	EU721194	EU721382	-
<i>Mischocarpus</i>	<i>grandissimus</i>	Radlk.	<i>Edwards KE37</i>	JCT	Australia	EU720437	EU720604	EU720762	EU720945	EU721044	EU721198	EU721386	EU721551
<i>Mischocarpus</i>	<i>pentapetalus</i>	(Rox.) Radlk.	<i>Chase 2133</i>	K	Bogor, BG	EU720470	EU720637	EU720802	EU720966	EU721074	EU721238	EU721426	EU721571
<i>Mischocarpus</i>	<i>pyriformis</i>	(F. Muell.) Radlk.	<i>Chase 2059</i>	K	Australian, BG	EU720460	EU720628	EU720792	-	EU721066	EU721228	EU721416	-
<i>Molinaea</i>	<i>petiolaris</i>	Radlk.	<i>Rabenantoandro 1448</i>	MO	Madagascar	EU720554	EU720700	EU720902	EU721007	EU721138	EU721324	EU721512	-
<i>Molinaea</i>	<i>sp. nov.</i>		<i>Antilahimena 4301</i>	MO	Madagascar	EU720510	EU720662	EU720854	EU720983	EU721099	EU721280	EU721468	EU721578
<i>Neotina</i>	<i>coursii</i>	Capuron	<i>H. Razafindraibe 119</i>	MO	Madagascar	EU720543	EU720690	EU720891	EU721002	EU721128	EU721313	EU721501	EU721594
<i>Nephelium</i>	<i>lappaceum</i>	L.	<i>Edwards KE222</i>	JCT	Asia	-	EU720584	EU720738	EU720932	EU721030	EU721175	EU721363	EU721537
<i>Nephelium</i>	<i>lappaceum (=N. chryseum)</i>	L.	<i>Yuan s.n.</i>	NEU	China	EU720401	-	EU720716	EU720915	-	EU721153	EU721342	EU721523
<i>Nephelium</i>	<i>mutabile</i>	Blume	<i>Chase 2134</i>	K	Bogor, BG	-	AY724316	-	-	-	-	-	-
		(Hiern) Exell & Mendonça	<i>Edwards KE231</i>	JCT	Tanzania	EU720411	EU720573	EU720728	EU720925	EU721022	EU721165	EU721353	EU721531
<i>Pancovia</i>	<i>golungensis</i>	Eckl. & Zeyh.	<i>Edwards KE232</i>	JCT	South Africa	EU720424	EU720592	EU720747	EU720938	EU721035	EU721183	EU721371	EU721542
<i>Pappea</i>	<i>capensis</i>	King	<i>Chase 1356</i>	K	Bogor, BG	EU720451	EU720619	EU720782	EU720955	EU721057	EU721218	EU721406	-
<i>Paranephelium</i>	<i>macrophyllum</i>	Miq.	<i>Edwards KE503</i>	JCT	Asia	EU720420	EU720582	EU720737	-	EU721029	EU721174	EU721362	-
<i>Paullinia</i>	<i>alata</i> subsp. <i>alata</i>	G. Don	<i>Weckerle 00/03/09-2/1</i>	Z	Peru	-	-	EU720834	-	-	-	-	-
<i>Paullinia</i>	<i>elegans</i>	Cambess.	<i>Weckerle 00/05/27-1/1</i>	Z	Peru	-	-	EU720835	-	-	EU721267	EU721455	-
<i>Paullinia</i>	<i>ericarpa</i>	Triana & Planch.	<i>Weckerle 00/06/13-1/5</i>	Z	Peru	EU720497	-	EU720839	-	-	-	-	-
<i>Paullinia</i>	<i>faginea</i>	(Triana & Planch.) Radlk.	<i>Weckerle 00/05/27-1/5</i>	Z	Peru	-	-	EU720837	-	-	-	-	-
<i>Paullinia</i>	<i>faginea</i>	(Triana & Planch.) Radlk.	<i>Weckerle 00/06/13-1/3</i>	Z	Peru	EU720496	-	EU720838	-	-	-	-	-
<i>Paullinia</i>	<i>pachycarpa</i>	Benth.	<i>Weckerle 01/01/26-1/1</i>	Z	Peru	EU720500	-	EU720842	-	-	-	-	-
<i>Paullinia</i>	<i>pinnata</i>	L.	<i>Edwards KE199</i>	JCT	Tanzania	EU720413	EU720575	EU720730	EU720926	EU721023	EU721167	EU721355	-
<i>Paullinia</i>	<i>subauriculata</i>	Radlk.	<i>Weckerle 00/03/19-1/1</i>	Z	Peru	EU720494	-	EU720833	-	-	EU721266	EU721454	-
<i>Plagioscyphus</i>	<i>aff. louvelii</i>	Danguy & Choux	<i>Lowry 6034</i>	MO	Madagascar	EU720555	EU720701	EU720903	EU721008	EU721139	EU721325	EU721513	EU721599
<i>Plagioscyphus</i>	<i>unijugatus</i>	Capuron	<i>Buerki 145</i>	NEU	Madagascar	EU720475	EU720644	EU720809	EU720972	EU721081	EU721245	EU721433	EU721574
<i>Podonephelium</i>	<i>homei</i>	Radlk.	<i>McPherson 18156</i>	MO	New Caledonia	-	-	-	-	-	EU721312	EU721500	-
<i>Podonephelium</i>	<i>homei</i>	Radlk.	<i>Pillon 156</i>	MO	New Caledonia	EU720489	EU720650	EU720826	EU720976	EU721087	EU721259	EU721447	-

Appendix. Continued

		GenBank accession number											
Genera	species	Author	Voucher	Herbarium	Country	ITS	matK	rpoB	trnD-trnT	trnK-matK	trnL	trnL-F	trnS-trnG
<i>Pometia</i>	<i>pinnata</i>	J.R. Forst. & G. Forst.	<i>Chase 2135</i>	K	Bogor, BG	EU720471	EU720638	EU720803	EU720967	EU721075	EU721239	EU721427	EU721572
<i>Pometia</i>	<i>pinnata</i>	J.R. Forst. & G. Forst.	<i>Yuan s.n.</i>	NEU	China	EU720402	-	EU720717	EU720916	-	EU721154	EU721343	EU721524
<i>Pseudima</i>	sp.		<i>McPherson 15867</i>	MO	Panama	EU720556	EU720702	EU720904	EU721009	EU721140	EU721326	EU721514	EU721600
<i>Pseudopteris</i>	<i>decipiens</i>	Baill.	<i>Kakazomannjary 12529-SF</i>	MO	Madagascar	EU720480	-	EU720816	-	-	-	-	-
<i>Rhysotoechia</i>	<i>mortoniana</i>	(F. Muell.) Radlk.	<i>Edwards KE117</i>	JCT	Australia	EU720414	EU720576	EU720731	EU720927	EU721024	EU721168	EU721356	EU721533
<i>Rhysotoechia</i>	<i>robertsonii</i>	(F. Muell.) Radlk.	<i>Edwards KE277</i>	JCT	Australia	-	EU720570	EU720725	EU720923	EU721021	EU721162	EU721350	EU721529
<i>Sapindus</i>	<i>oligophyllus (=Aphania oligophylla)</i>	Merr. & Chun	<i>Yuan s.n.</i>	NEU	China	EU720407	EU720568	EU720722	EU720921	EU721020	EU721159	-	-
<i>Sapindus</i>	<i>saponaria</i>	L.	<i>Chase 2136</i>	K	Bogor, BG	-	EU720639	EU720804	EU720968	EU721076	EU721240	EU721428	-
<i>Sapindus</i>	<i>trifoliatus</i>	L.	<i>Edwards KE504</i>	JCT	Asia	-	EU720586	EU720740	EU720934	EU721032	EU721177	EU721365	EU721538
<i>Sarcopteryx</i>	<i>martyana</i>	(F. Muell.) Radlk.	<i>Irvine IRV1810</i>	CSIRO	Australia	EU720426	EU720594	EU720749	EU720940	EU721037	EU721185	EU721373	EU721544
<i>Sarcopteryx</i>	<i>reticulata</i>	S.T. Reynolds	<i>Gray BG1137</i>	CSIRO	Australia	EU720421	EU720587	EU720741	-	EU721033	EU721178	EU721366	EU721539
<i>Sarcopteryx</i>	sp.	-	<i>Edwards KE49</i>	JCT	Australia	EU720439	EU720607	EU720765	EU720948	EU721047	EU721201	EU721389	EU721554
<i>Sarcotoechia</i>	<i>serrata</i>	S.T. Reynolds	<i>Edwards KE31</i>	JCT	Australia	EU720436	EU720603	EU720761	EU720944	EU721043	EU721197	EU721385	EU721550
<i>Sarcotoechia</i>	<i>villosa</i>	S.T. Reynolds	<i>Edwards KE102</i>	JCT	Australia	EU720419	EU720581	EU720736	-	-	EU721173	EU721361	-
<i>Schleichera</i>	<i>oleosa</i>	(Lour.) Oken	<i>Chase 2137</i>	K	Bogor, BG	EU720423	EU720591	EU720746	EU720937	-	EU721182	EU721370	EU721541
<i>Scyphonichium</i>	cf. <i>multiflorum</i>	(Mart.) Radlk.	<i>Acevedo 11102</i>	US	French Guiana	-	EU720672	EU720869	EU720990	EU721109	EU721291	EU721479	-
<i>Serjania</i>	<i>altissima</i>	(Poepp.) Radlk.	<i>Weckerle 00/07/02-1/4</i>	Z	Peru	EU720498	-	EU720840	-	-	EU721269	EU721457	-
<i>Serjania</i>	<i>communis</i>	Cambess.	<i>Chase 2138</i>	K	Bogor, BG	EU720472	EU720640	EU720805	EU720969	EU721077	EU721241	EU721429	-
<i>Serjania</i>	<i>glabrata</i>	Kunth	<i>Merello 1058</i>	MO	Peru	EU720557	EU720703	EU720905	EU721010	EU721141	EU721327	EU721515	-
<i>Serjania</i>	<i>mexicana</i>	(L.) Willd.	<i>Davidse 35748</i>	MO	Belize	-	EU720704	EU720906	-	EU721142	-	-	-
<i>Serjania</i>	<i>triquetra</i>	Radlk.	-	-	-	AY207571	-	-	-	-	-	-	-
<i>Smelophyllum</i>	<i>capense</i>	Radlk.	<i>Edwards KE506</i>	JCT	South Africa	-	AY724330	-	-	-	-	-	-
<i>Stadmannia</i>	<i>oppositifolia</i>	(Lam.) Poir.	<i>Edwards KE505</i>	JCT	Madagascar	-	AY724331	-	-	-	-	-	-
<i>Storthocalyx</i>	<i>leioneurus</i>	Radlk.	<i>Munzinger 1100</i>	MO	New Caledonia	EU720506	EU720659	EU720850	-	EU721096	-	-	-
<i>Storthocalyx</i>	sp.		<i>Munzinger 960</i>	MO	New Caledonia	EU720504	EU720657	EU720848	-	EU721094	EU721276	EU721464	-
<i>Synima</i>	<i>cordieri</i>	(F. Muell.) Radlk.	<i>Edwards KE29</i>	JCT	Australia	-	EU720602	EU720760	-	EU721042	EU721196	EU721384	-
<i>Synima</i>	<i>macrophylla</i>	S.T. Reynolds	<i>Edwards KE19</i>	JCT	Australia	EU720430	EU720596	EU720753	EU720941	-	EU721189	EU721377	EU721546
<i>Talisia</i>	<i>angustifolia</i>	Radlk.	<i>Zardini 43668</i>	MO	Paraguay	EU720558	EU720705	EU720907	-	EU721143	EU721328	EU721516	-
<i>Talisia</i>	<i>nervosa</i>	Radlk.	<i>Pennington 628</i>	MO	-	EU720474	EU720643	EU720808	-	EU721080	EU721244	EU721432	-
<i>Talisia</i>	<i>obovata</i>	A.C. Sm.	<i>R.Lombello 13</i>	MO	Brazil	EU720485	EU720648	EU720822	-	EU721085	EU721255	EU721443	-
<i>Thouinia</i>	<i>acuminata</i>	S. Watson	<i>Liston 633-2</i>	MO	Mexico, Jalisco	EU720478	EU720647	EU720814	-	EU721084	EU721249	EU721437	-
<i>Tina</i>	<i>isaloensis</i>	Drake	<i>Ranirison PR827</i>	G	Madagascar	EU720520	EU720667	EU720864	EU720986	EU721104	EU721286	EU721474	EU721579
<i>Tina</i>	<i>striata</i>	Radlk.	<i>Vary 45</i>	MO	Madagascar	EU720509	EU720661	EU720853	-	EU721098	EU721279	EU721467	-
<i>Tinopsis</i>	<i>apiculata</i>	Radlk.	<i>Buerki 131</i>	NEU	Madagascar	EU720422	EU720589	EU720744	EU720936	EU721034	EU721180	EU721368	EU721540
<i>Toechima</i>	<i>daemelianum</i>	Radlk.	<i>JC66</i>	CSIRO	Australia	-	AY724334	-	-	-	-	-	-
<i>Toechima</i>	<i>erythrocarpum</i>	(F. Muell.) Radlk.	<i>Edwards KE20</i>	JCT	Australia	EU720431	EU720597	EU720754	-	EU721038	EU721190	EU721378	-
<i>Toechima</i>	<i>plurinerve</i>	Radlk.	<i>Chase 1357</i>	K	Bogor, BG	EU720452	EU720620	EU720783	EU720956	EU721058	EU721219	EU721407	EU721561
<i>Toechima</i>	<i>tenax</i>	(Cunn. ex Benth.) Radlk.	<i>Chase 2046</i>	K	Australian, BG	EU720456	EU720623	EU720787	-	EU721061	EU721223	EU721411	EU721564
<i>Toechima</i>	<i>tenax</i>		<i>Chase 2132</i>	K	Bogor, BG	EU720469	EU720636	EU720801	EU720965	EU721073	EU721237	EU721425	EU721570
<i>Tristira</i>	<i>triptera</i>	(Blanco) Radlk.	<i>Chase 2139</i>	K	Asia	-	EU720585	EU720739	EU720933	EU721031	EU721176	EU721364	-
<i>Tristiropsis</i>	<i>acutangula</i>	Radlk.	<i>Chase 1358</i>	K	Bogor, BG	EU720453	EU720621	EU720784	EU720957	EU721059	EU721220	EU721408	EU721562

Appendix. Continued

Genera	species	Author	Voucher	Herbarium	Country	GenBank accession number							
						ITS	matK	rpoB	trnD-trnT	trnK-matK	trnL	trnL-F	trnS-trnG
<i>Ungnadia</i>	<i>speciosa</i>	Endl.	<i>Chase 2854</i>	RBG	-	-	EU720706	-	-	-	EU721332	EU721518	-
<i>Urvillea</i>	<i>ulmaceae</i>	Kunth	<i>Weckerle 00/07/05-1/1</i>	Z	Peru	EU720499	EU720655	EU720841	-	EU721092	EU721270	EU721458	-
<i>Vouarana</i>	<i>guianensis</i>	Aubl.	<i>Lucas 109</i>	MO	French Guiana	EU720525	EU720673	EU720870	EU720991	EU721110	EU721292	EU721480	EU721582
<i>Xanthoceras</i>	<i>sorbifolium</i>	Bunge	<i>Yuan CN2006</i>	NEU	China	EU720398	EU720562	EU720711	EU720910	EU721014	EU721148	EU721337	-
<i>Xerospermum</i>	<i>noronhianum</i>	(Blume) Blume	<i>Chase 2130</i>	K	Bogor, BG	-	EU720641	EU720806	EU720970	EU721078	EU721242	EU721430	EU721573
Outroup													
<i>Sorindeia</i>	sp.		<i>Buerki 137</i>	NEU	Madagascar	-	-	EU720831	-	-	EU721264	EU721452	-
<i>Harrisonia</i>	<i>abyssinica</i>	Oliv.	<i>Edwards KE510</i>	JCT	Tanzania	EU720440	-	EU720766	-	-	EU721202	EU721390	-
<i>Malleastrum</i>	sp.		<i>Rakotovao 2609</i>	MO	Madagascar	-	-	EU720878	EU720995	-	EU721300	EU721488	-

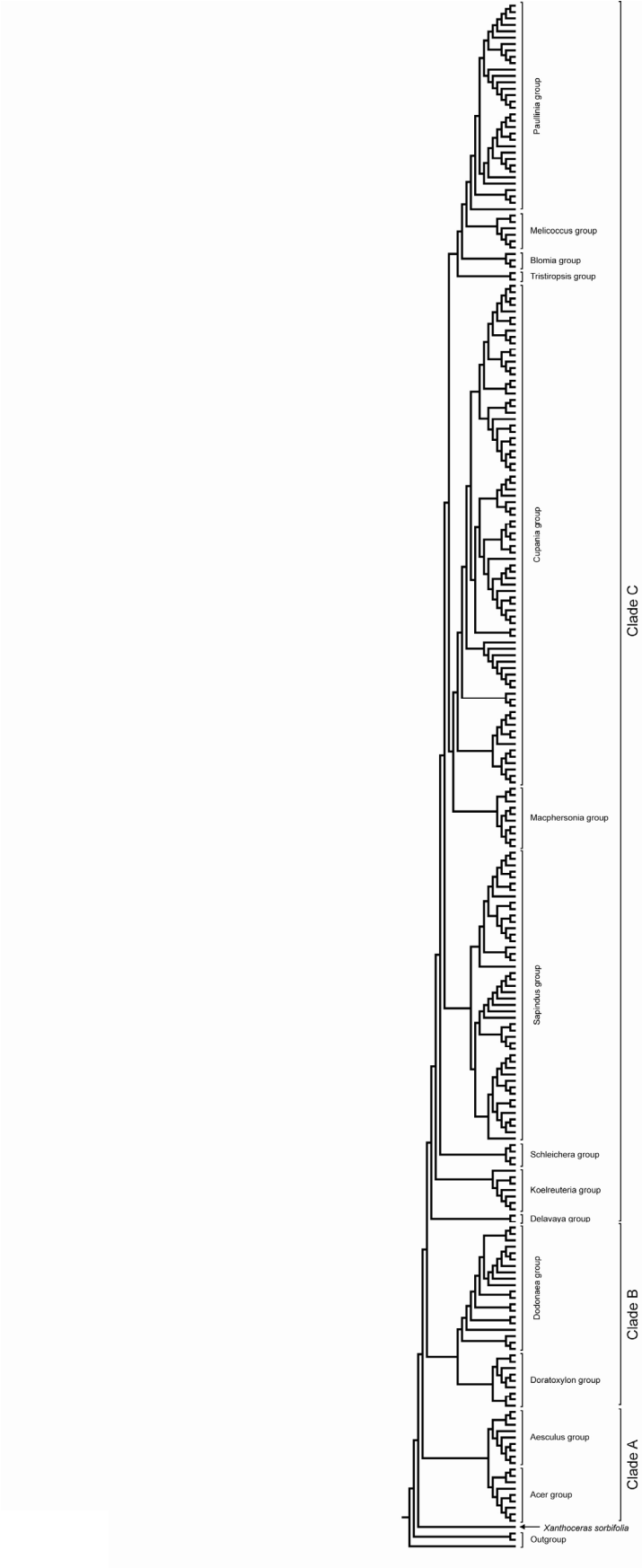


Figure 1

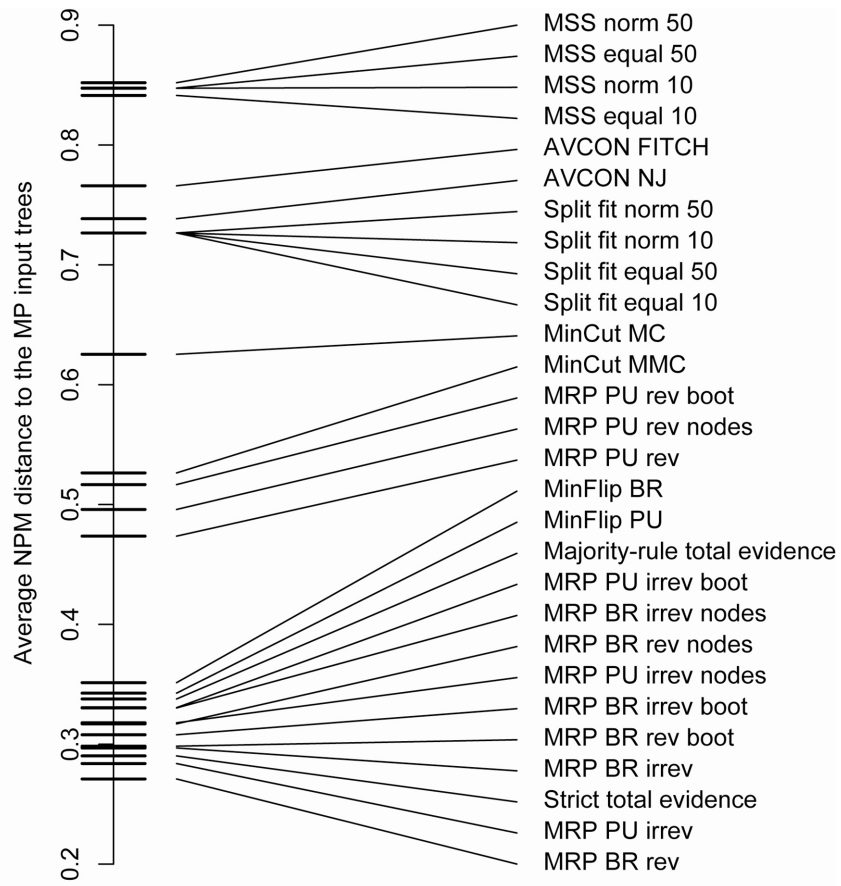


Figure 2A

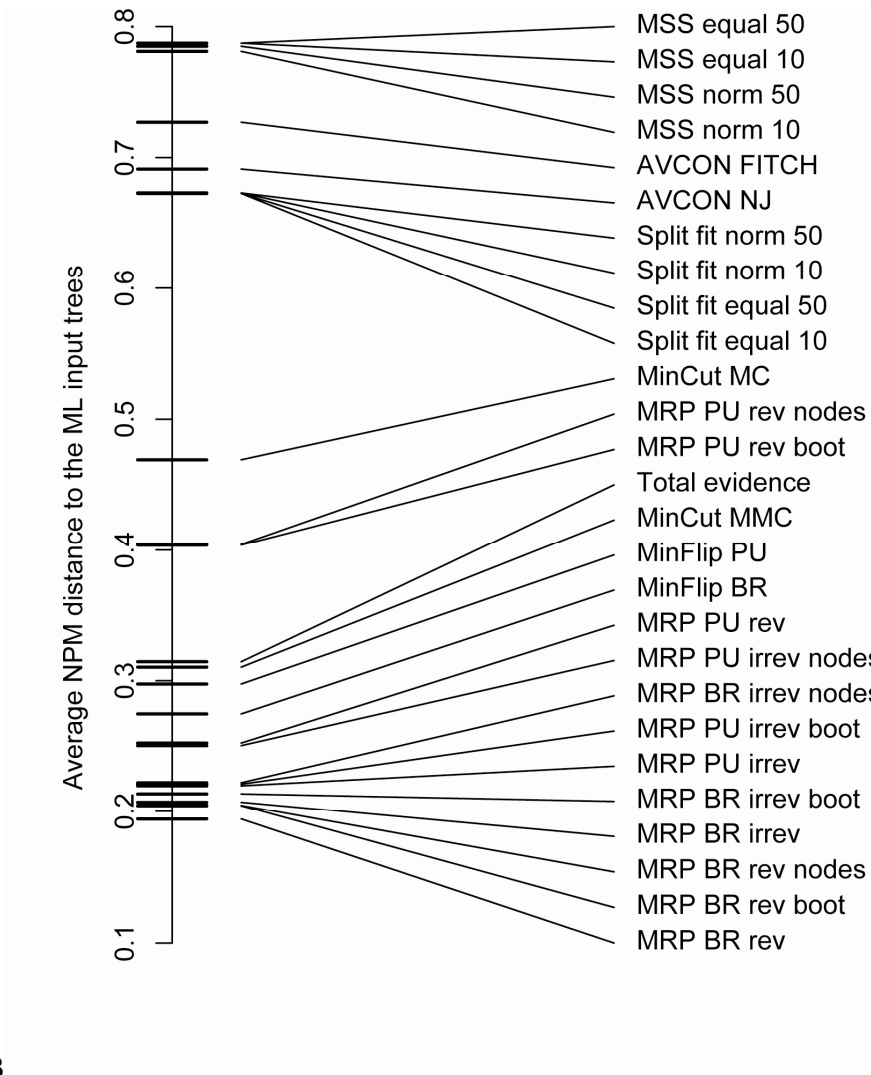


Figure 2B

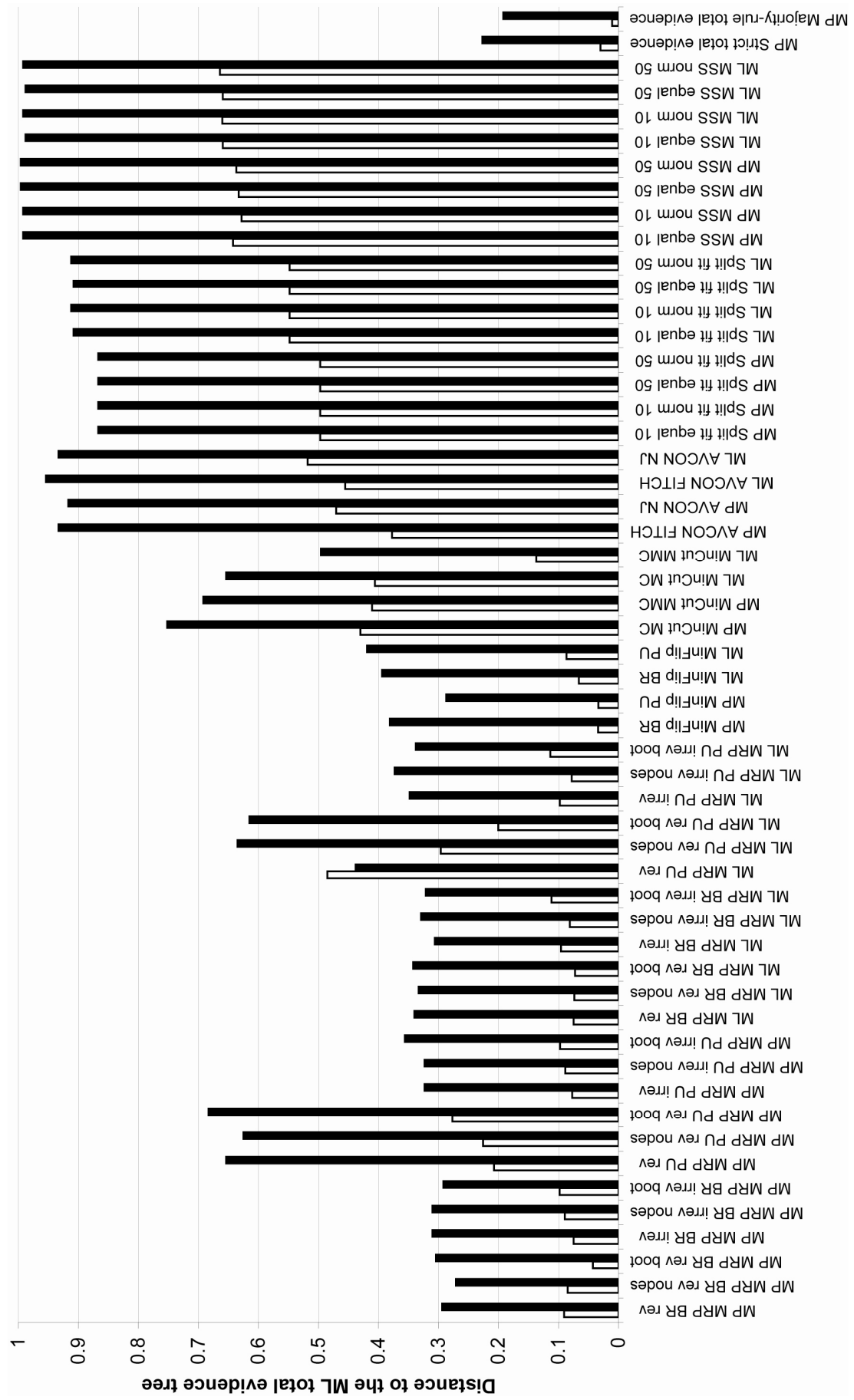


Figure 4

SUPPLEMENTARY MATERIAL

The best ML total evidence tree presented here is based on 243 specimens (240 specimens of Sapindaceae sensu lato representing more than 70% of the generic diversity and three outgroups) and supported by eight plastid and nuclear markers (Figs. S1-S6). Compared to Buerki et al. (2009), 90 ingroup specimens were produced corresponding to an increase of 14 genera (e.g., *Aesculus*, *Allophylus*, *Billia*, *Handeliodenron*). The topology is highly congruent (NPM distance of 0.037) with Buerki et al. (2009) and subdivided the family into three moderately to strongly-supported lineages and a fourth lineage only consisting of *Xanthoceras sorbifolia*, with the following relationships: (*Xanthoceras sorbifolia*, (clade A, (clade B, clade C))). Clade A corresponded to the former family Aceraceae (Buerki et al., 2009), whereas the inclusion of additional genera (*Aesculus*, *Billia* and *Handeliodendron*) in this analysis subdivided this clade into two groups (A-I; A-II) corresponding to subfamily Hippocastanoideae (Fig. S2). Clade B corresponded to subfamily Dodonaeoideae as described by Buerki et al. (2009) with the addition of *Diplokeleba* (Cupanieae; Sapindoideae) (Figs. S1-3). Clade C corresponded to subfamily Sapindoideae as defined by Buerki et al. (2009) and is divided into ten groups (Figs. S3-6). Buerki et al. (2009) recognized all the tribes (except the Paullinieae) as paraphyletic or polyphyletic. However, phylogenetic status of tribes Cossinieae and Koelreuterieae were not tested because only one genus per tribe was considered. The increase of sampling in this study revealed the paraphyly of Cossinieae and Koelreuterieae and confirmed the phylogenetic status of the rest of the tribes.

Figure S1. Best maximum likelihood total evidence tree of the Sapindaceae s.l. inferred from eight nuclear and plastid nucleotide sequences and 70% of the generic diversity. Bootstrap supports are indicated above branches. The revised infrafamilial classification based on molecular and morphological characters is indicated in grey (see Buerki et al., 2009 for more details). Abbreviations: Cossinieae (COS); Cupanieae (CUP); Dodonaeae (DOD); Doratoxyleae (DOR); Koelreuterieae (KOE); Harpullieae (HAR);

Lepisantheae (LEP); Melicocceae (MEL); Nephelieae (NEP); Paullinieae (PAU); Sapindeae (SAP); Schleichereae (SCH) and Thouinieae (THO).

Figure S2. Relationships within subfamilies Hippocastanoideae (clade A) and Dodonaeoideae (clade B). Bootstrap supports are indicated above branches. The revised infrafamilial classification based on molecular and morphological characters is indicated in grey. See Fig. S1 for abbreviations of tribes.

Figure S3. Relationships within subfamily Sapindoideae (clade C). Bootstrap supports are indicated above branches. The revised infrafamilial classification based on molecular and morphological characters is in grey. See Fig. S1 for abbreviations of tribes.

Figure S4. Phylogenetic relationships within the *Litchi* group (clade C-IV; see Fig. S3). Bootstrap supports are indicated above branches. See Fig. S1 for abbreviations of tribes.

Figure S5. Phylogenetic relationships within the *Cupania* group (clade C-VI; see Fig. S3). Bootstrap supports are indicated above branches. See Fig. S1 for abbreviations of tribes.

Figure S6. Phylogenetic relationships within the *Paullinia* group (clade C-X; see Fig. S3). Bootstrap supports are indicated above branches. See Fig. S1 for abbreviations of tribes.

Figure S7. NJ “tree of trees” based on pairwise NPM distance between the supertrees and total evidence trees. Abbreviations: black circles refer to MP trees, whereas white circles refer to ML trees. See in the text for definitions and abbreviations of supertree methods.

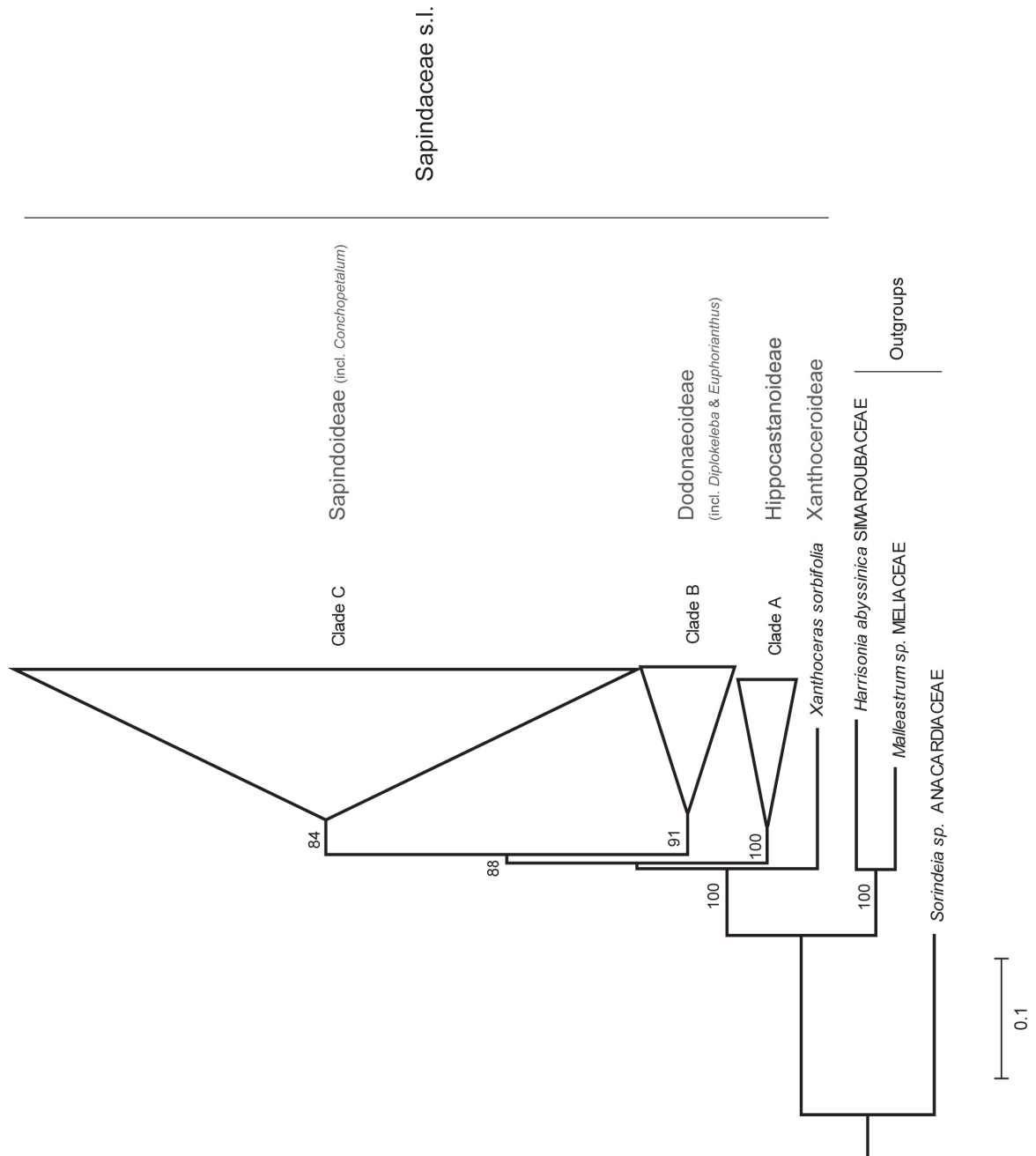


Figure S1

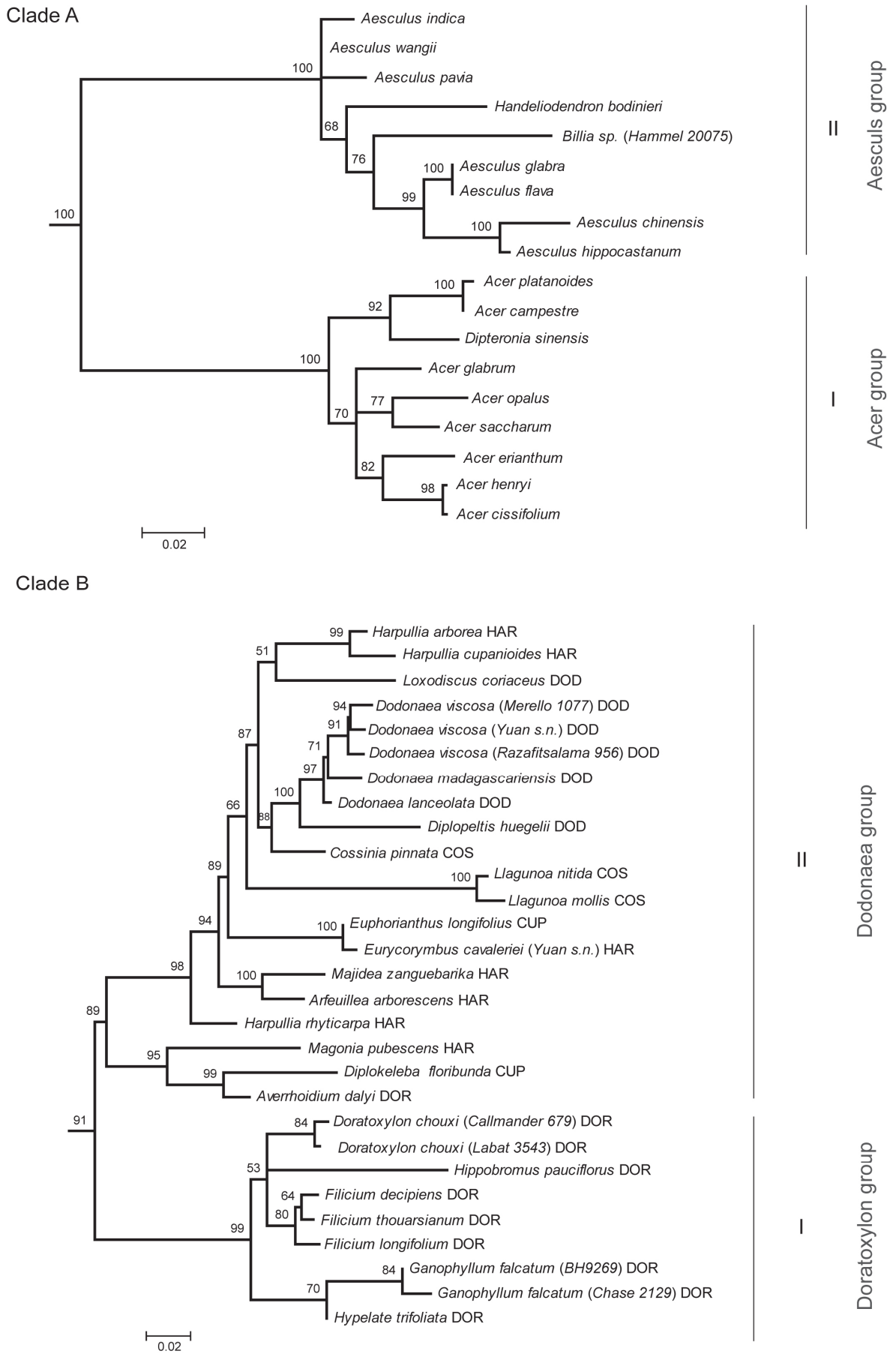


Figure S2

Clade C

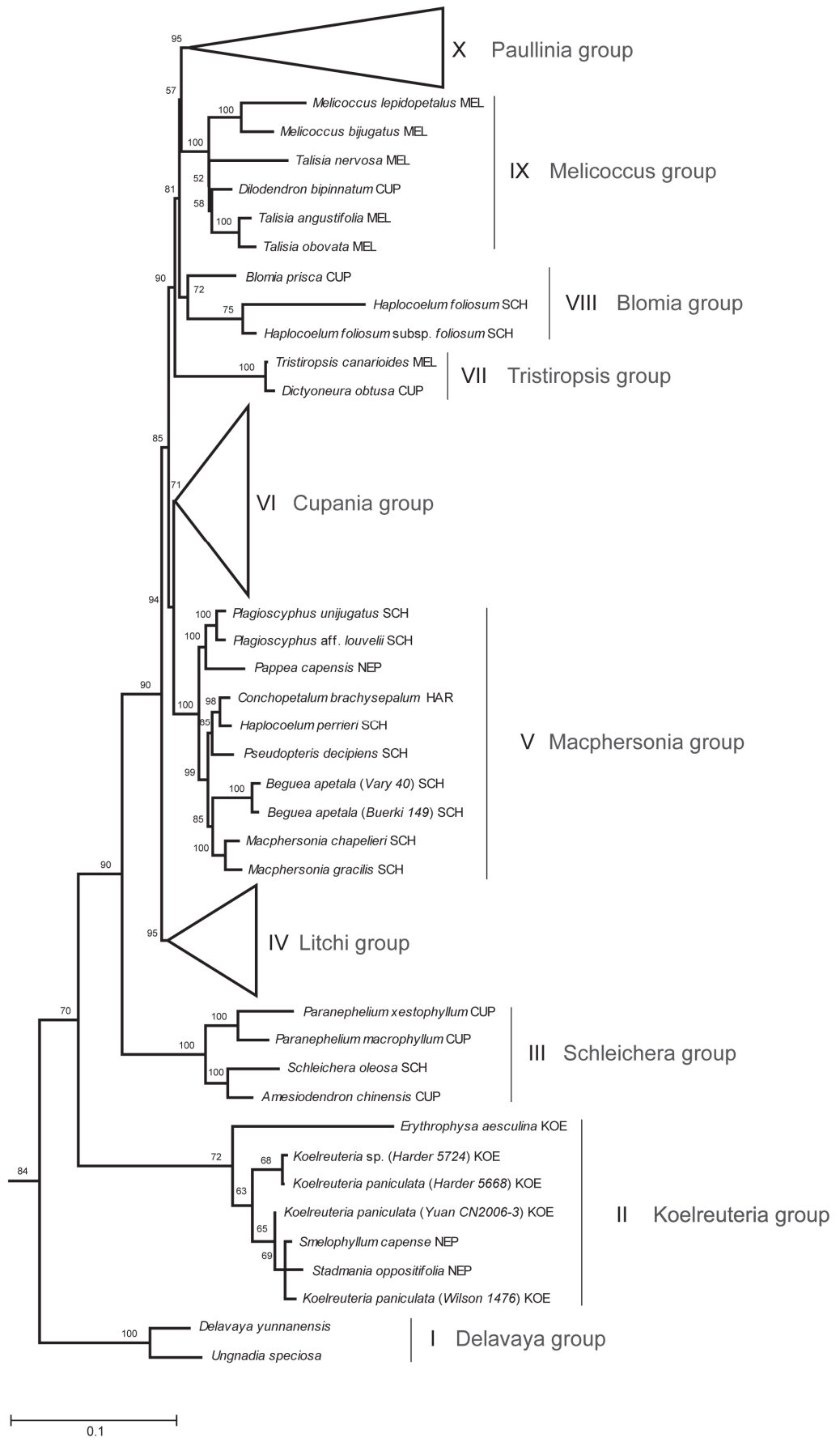


Figure S3

Clade C-IV

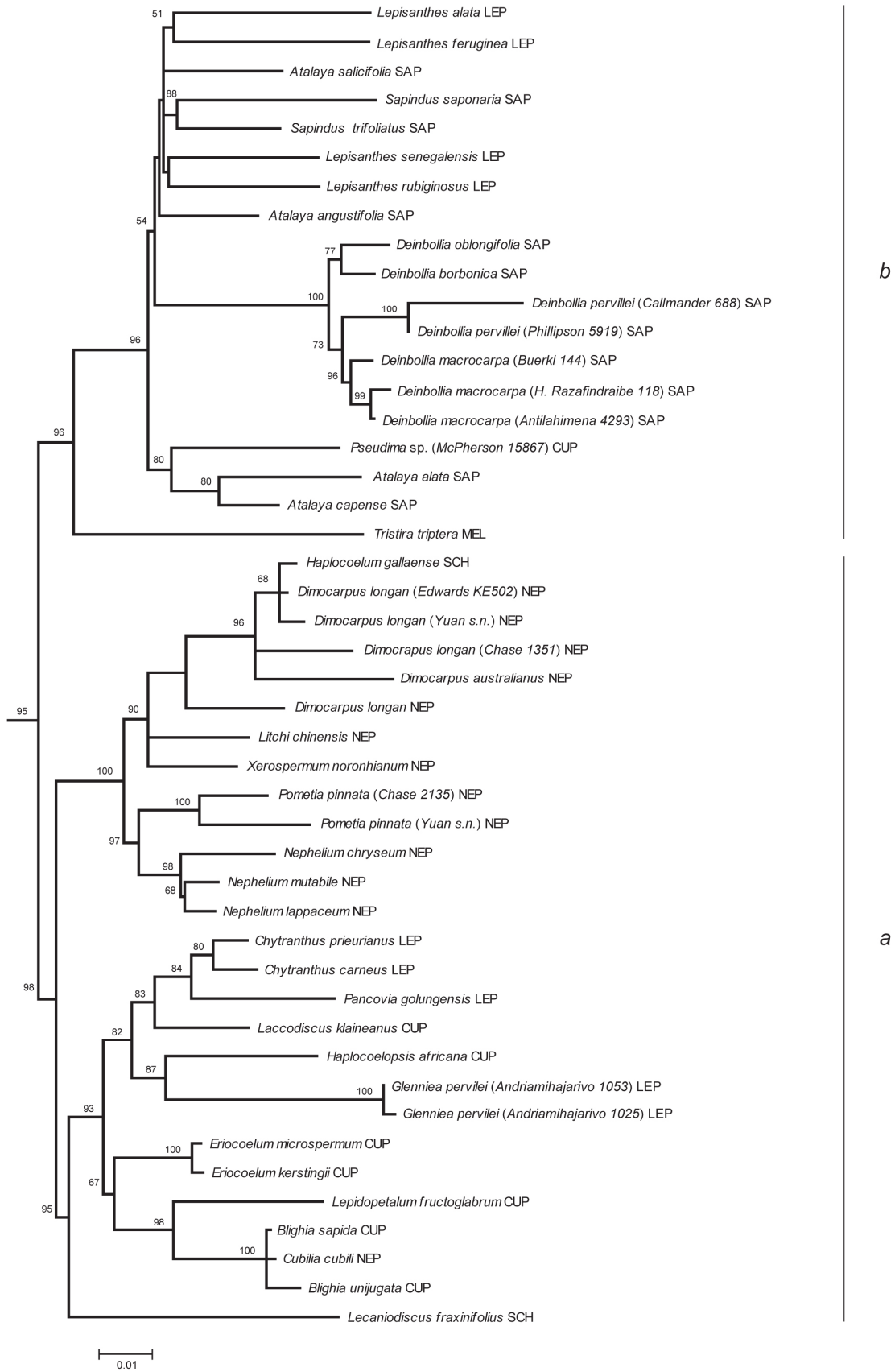


Figure S4

Clade C-VI

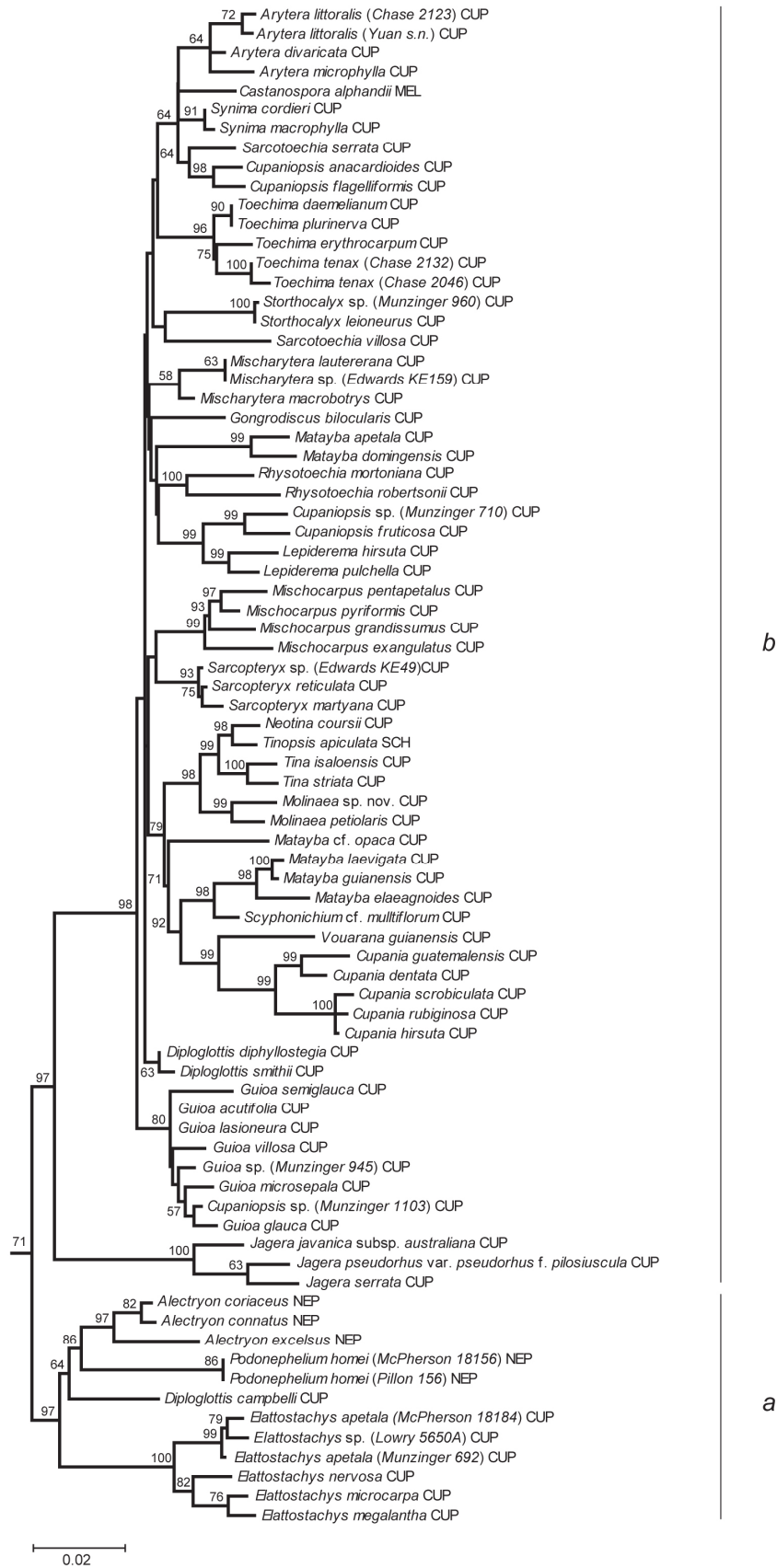


Figure S5

Clade C-X

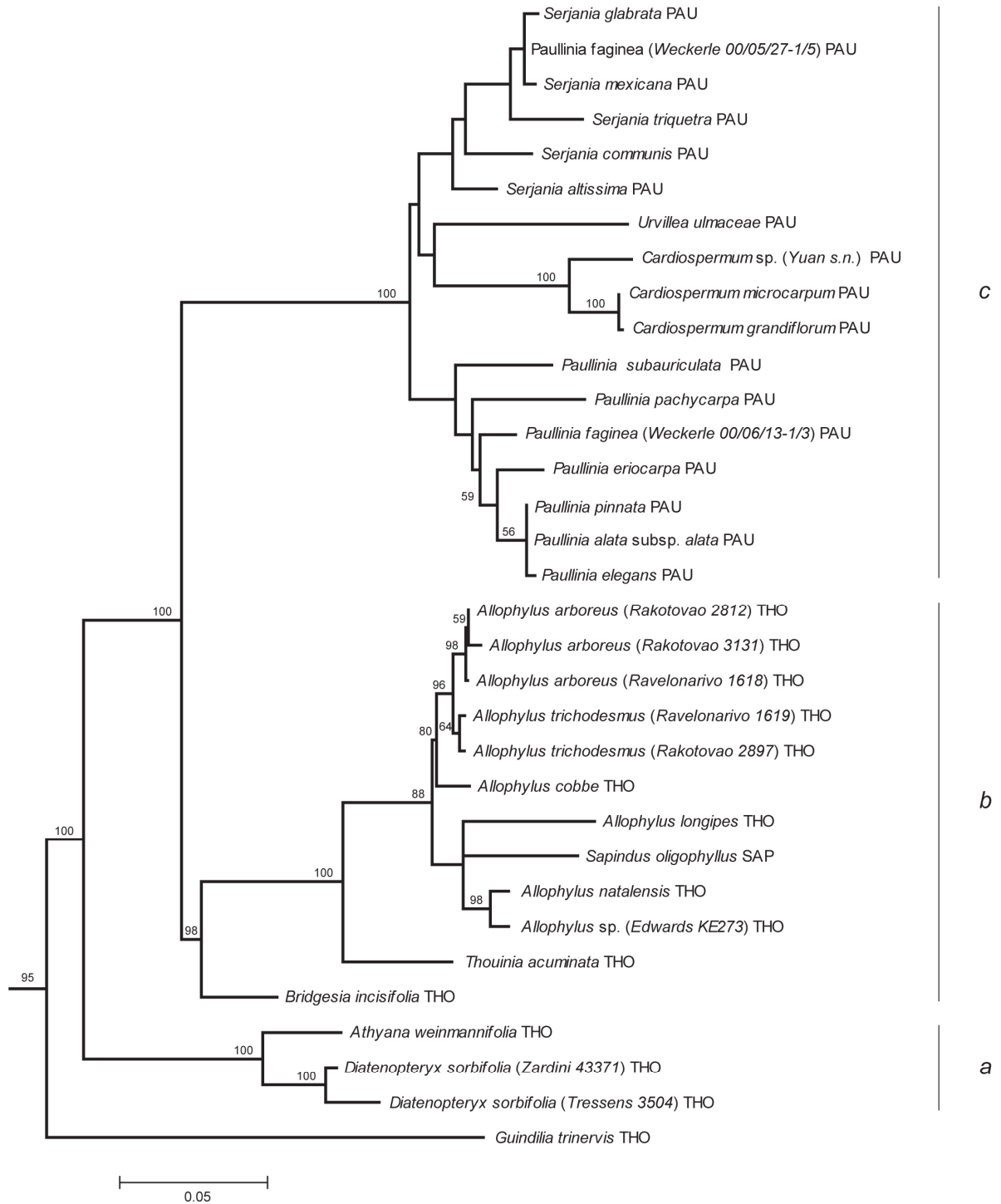


Figure S6

Conclusion & Perspectives

The main piece of work of this study was the inference of a phylogenetic framework for the soapberry family (Sapindaceae). This study was the first step in the understanding of relationships within this highly diversified family and allowed the recognition of monophyletic units at the base of a new classification. Once relationships between lineages were established, the phylogenetic hypothesis provided a robust framework for the investigation of the evolution of the family at a large scale of time and space. When extended to encompass more than 70% of the generic diversity (represented by >240 samples), our data set confirmed intricate relationships at the subfamilial and tribal levels (and even contested the monophyly of several genera; e.g., *Acer*, *Cupaniopsis*, *Guioa*; Fig. 1). For instance, the paraphyly of subfamily Dodonaeoideae was confirmed by the inclusion of one genus previously included in subfamily Sapindoideae (*Diplokeleba*; Cupanieae; Fig. 1C), and the inclusion of additional genera of tribes Cossinieae and Koelreuterieae revealed the paraphyletic status of both tribes. In order to provide a broad and accurate framework for further investigations, the best maximum likelihood phylogenetic tree for Sapindaceae based on an extended data set is presented hereafter (Figs. 1-3; see Supplementary material Table 1 for a survey of the sampling). As highlighted by Buerki et al. (2009), the phylogeny of Sapindaceae is in general better supported by biogeography than by any morphological characters. This statement might explain the paraphyletic status of the disjunct tribe Cossinieae, distributed between the Mascarenes, Australia (*Cossinia*) and South America (*Llagunoa*) (Fig. 1C). On the other hand, the addition of new genera and species confirmed the monophyletic status of the previously recognized Aceraceae and Hippocastanaceae (currently recognized as a subfamily within Sapindaceae; Fig. 1A-B).

Conclusion & Perspectives

Sapindoideae (clade C)

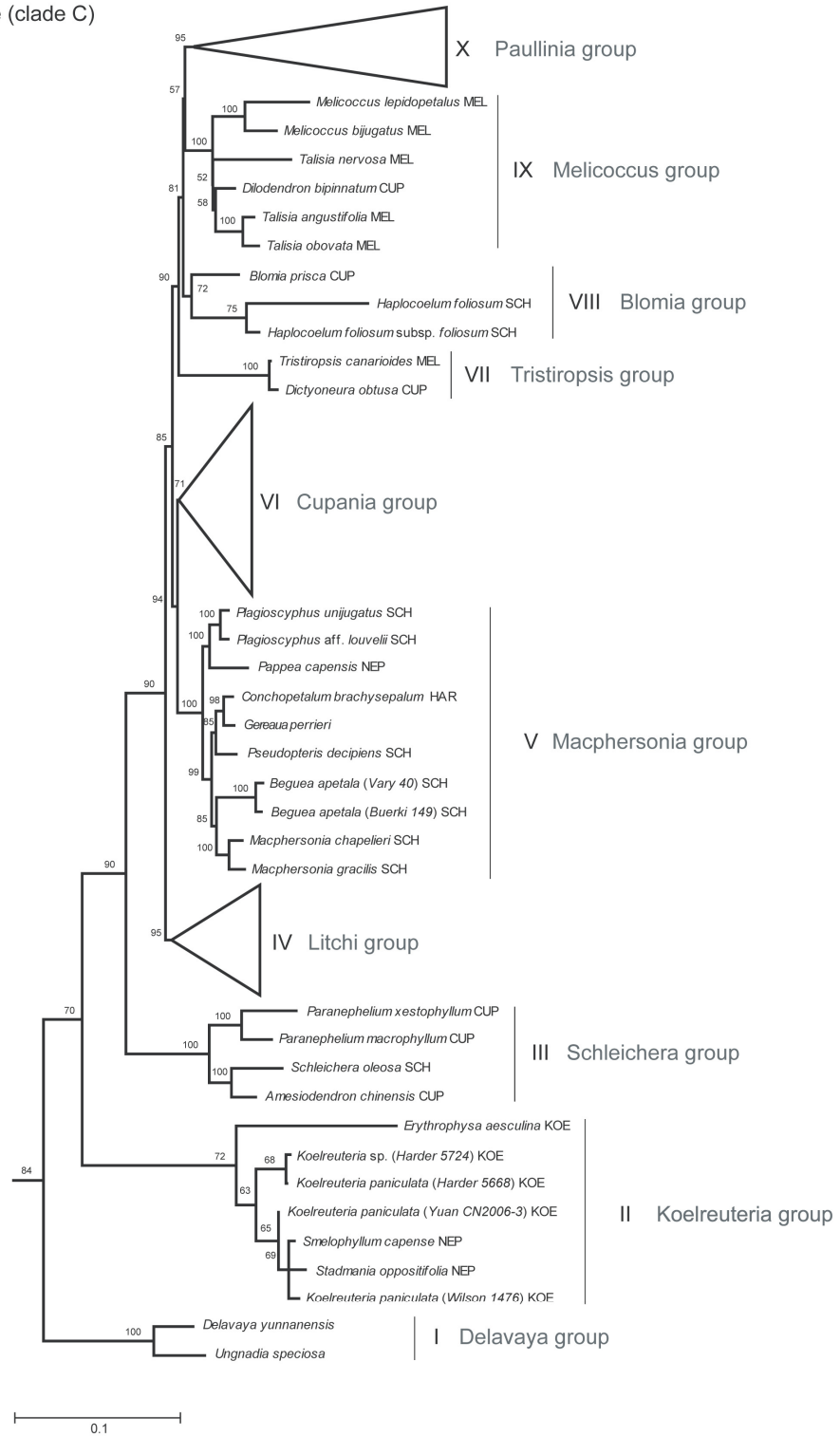


Figure 2. See legend above



Figure 3. See legend above

Figure 1. A. Best maximum likelihood phylogenetic tree for Sapindaceae. B. Relationships within subfamilies Hippocastanoideae (B) and Dodonaeoideae (C). Bootstrap supports are indicated above branches. See Supplementary Table 1 and Buerki et al. (2009) for more details on sampling and abbreviations of tribes.

Figure 2. Relationships within subfamily Sapindoideae (clade C). Bootstrap supports are indicated above branches. See Supplementary Table 1 and Buerki et al. (2009) for more details on sampling and abbreviations of tribes.

Figure 3. Relationships within Litchi (A), Cupania (B) and Paullinia (C) groups. Bootstrap supports are indicated above branches. See Supplementary Table 1 and Buerki et al. (2009) for more details on sampling and abbreviations of tribes.

Several morphological and biogeographic synapomorphies (e.g., simple, opposite leaves and a temperate distribution) support the monophyly of the two formerly recognized families.

Their inclusion within Sapindaceae was mainly motivated by the recognition of the Chinese and Korean monotypic genus *Xanthoceras* as the most basal lineage within the family (APGII, 2003). This species was described as a Sapindaceae; however, Radlkofer (1933) commented almost a century ago on its dubious position (see Buerki et al., 2009 for more details). Within subfamily Sapindoideae, the ten groups defined in Buerki et al. (2009) were recovered and slightly expanded (Figs. 2, 3; Table 1). For instance, the previously identified Chinese *Delavaya* group was broadened to include the Texan and Floridian monotypic genus *Ungnadia*; the definition of the *Koelreuteria* group was extended by the African and Malagasy *Erythrophysa*, *Stadmania* and the South African *Smelophyllum*, and the pantropical genus *Allophylus* was included in the *Paullinia* group (Figs. 2, 3; Table 1).

Although Sapindaceae constitute a suitable case study for the investigation of large scale evolutionary processes (see below), its current broad circumscription has some collateral effects such as impending the establishment of regional and worldwide taxonomic

treatments. These collateral effects are mainly caused by i) the lack of synapomorphies supporting the family and ii) the recognition, in chapter 1, of the plasticity of fruit characters used to circumscribe taxonomic entities within Sapindaceae. Many researchers investigated micro- and macromorphological characters to identify synapomorphies [e.g., macromorphology and palynology (Müller and Leenhouts, 1976); wood anatomy (Klaassen, 1999)]; however no synapomorphies supporting the family were recognized, a fact that reflects the incredible diversity expressed by this family (see the introduction for more details). A possible strategy to solve this problem might be to (i) establish a new monotypic family based on *Xanthoceras*, (ii) resurrect Aceraceae and Hippocastanaceae, and (iii) consider Dodonaeoideae and Sapindoideae as comprising a single family, Sapindaceae. Through such an approach, each of the four families would be supported by morphological and molecular characters. Once adopted, the circumscription of tribes might then be conducted by investigating clades restricted to specific areas. For instance, the *Macphersonia* group is mainly restricted to Madagascar and exhibits a high range of variability in morphological characters (see chapter 4). This group presents a potentially intriguing opportunity to investigate evolution and diversification in the southwest Indian Ocean, incorporating data on a wide range of characters (e.g., chromosome numbers, anatomy, biochemistry and floral development) along with a focus on molecular phylogenetics using broader sampling to identify synapomorphies for the group and to clarify its position within a revised classification system for Sapindaceae.

When investigating large scale evolutionary processes I pointed out that the Asian and Australian region (more precisely from the Indochinese peninsula to New Guinea, with high diversity encountered in South-Eastern Asia) has played a major role during the evolution of Sapindaceae. Although the family is likely to have originated in temperate China sometime in the Late Cretaceous, geological and climatic conditions encountered in the Asian and Australian region (especially Southeast Asia) triggered their diversification once they crossed the tropical border. This diversification might also have been accelerated by the abrupt climate change during the Eocene-Oligocene boundary (see chapter 3 for more

details). Interestingly, similar biogeographic patterns with an origin in South-Eastern Asia or Australia have been recovered for other angiosperm families such as Cucurbitaceae (Schaefer et al., 2009) and Melastomataceae (Renner and Meyer, 2001). This pattern can also be paralleled to the evidence that most basal lineages of angiosperms are restricted to the Australian and South East Asian region (e.g., Amborellaceae, Hydatellaceae, Austrobaileyaceae, Trimeniaceae and Himantandraceae; Fig. 4) and the most ancient fossil known to date was found in China (Friis et al., 2006). The role played by this area in driving the diversification of angiosperms across time can be discussed in a biogeographic framework.



Figure 4. Pattern of distribution of five of the seven most primitive lineages of angiosperms (ANA clade;

APGII, 2003):
 Amborellaceae
 (black filled line),
 Hydatellaceae
 (narrow black
 dashed line),
 Austrobaileyaceae
 (grey filled line),
 Illiciaceae (large
 black dashed line)

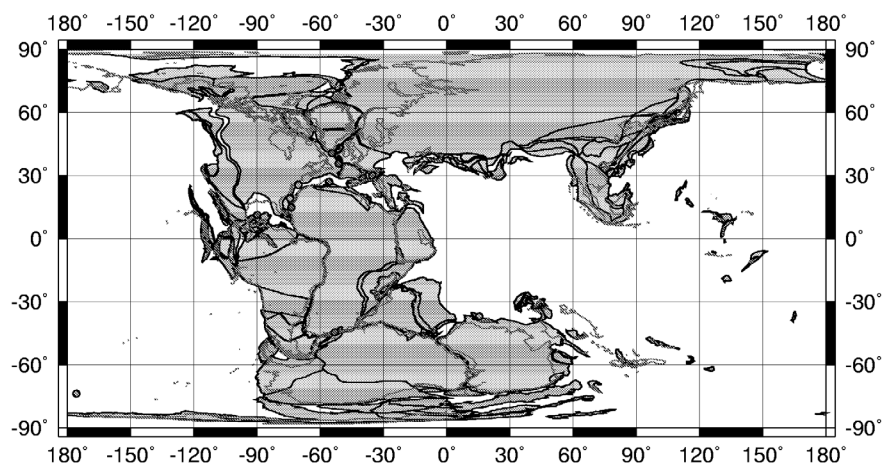


Figure 5. Configuration of continents during the Early Jurassic (ca. 170 Mya). Based on www.odsni.de.

and Tremeriaceae (narrow grey dashed line). The two remaining families (Cabombaceae and Nymphaeaceae) are worldwide.

Recently conducted time analyses estimated the apparition of early angiosperm lineages sometime between the Triassic and Jurassic (Bell et al., 2005). During this period, Gondwana was in place and several islands connected its northern part (Australia) to Laurasia (Indochina) (Fig. 5). Such lines of evidence therefore might indicate a putative spatial origin of the angiosperms somewhere between Australia and southern Asia. By considering such a hypothesis, we might explain that the lack of evidence for the recognition of the most recent common ancestor of the angiosperms (Frohlich and Chase, 2007) might be caused by the disappearance of the islands (meaning that the great majority of the most basal lineages of angiosperms have gone extinct in the mean time; Fig. 5). In contrast, these now disappeared islands might have played the role of a natural laboratory for lineage divergence and could have promoted the diversification of angiosperms, across the Cretaceous and the Tertiary. As shown in this study, such process in this area might at least have played a role in the soapberry family.

Perspectives

This study must be complemented by several further investigations :

1. Infer the evolution of a large panel of morphological character based on the extended data set presented in chapter 6. This should be done by applying Bayesian statistical methods (Ronquist, 2004) that enable the study of character evolution while accounting for both phylogenetic and mapping uncertainty. Such an approach might help to provide a new taxonomy for the family but also to understand the spatial evolution of Sapindaceae related to e.g. dispersal abilities through time.
2. At the beginning of this study, molecular methods were planned to be used as tools to unravel taxonomic entities in Madagascar. However, the high level of paraphyly and polyphyly discovered prevented achieving this goal. Additional fieldwork coupled with

Bayesian analysis of character evolution might provide effective guidelines (especially within the clade comprising *Molinaea*, *Tina*, and *Tinopsis*).

3. In order to fulfil the first two points, floral and fruit ontogeny should be investigated. A careful look at flower and fruit development might provide evidence to recognize putative synapomorphies supporting molecular clades. For instance, the ontogeny of the aril in Sapindaceae is not well known. Adema et al. (1994) suggested that the funiculus (the tissue commonly at the origin of the aril) might be absent in Sapindaceae and that consequently the different types of arils may have different origins. Analysing the aril ontogeny in taxa representative of both molecular and morphological diversity should allow a better understanding of this trait, which likely represents a key-character in dispersal abilities of this family.
4. Bayesian divergence time estimations (e.g., BEAST; Drummond and Rambaut, 2007) and diversification rates analyses (e.g., LASER; Rabosky, 2006) should be performed to confirm the observed higher diversification of Sapindaceae during the Eocene-Oligocene boundary. However, BEAST analyses are particularly difficult to set and its MCMC algorithm is less efficient than the one implemented in MrBayes (Ronquist and Huelsenbeck, 2003). In this study, BEAST analyses were attempted but they failed to reach convergence mainly due to the complexity of the soapberry data set (>150 specimens and eight markers; data not shown).
5. Further taxonomic and molecular investigations should be done on African Sapindaceae to confirm if the observed low species diversity (compared to South America, Australia and South-East Asia) is caused by massive extinction masses, for instance in the Eocene-Oligocene boundary, or simply by underestimating its species diversity. Currently, treatments of Sapindaceae are not completed for many African countries.

References

- Adema, F., Leenhouts, P.W., van Welzen, P.C., 1994. Sapindaceae. Flora Malesiana Ser. I 11, 419–768.

- Angiosperm Phylogeny Group (APG II), 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APGII. *Bot. J. Linn. Soc.* 141, 399–436.
- Bell, C.D., Soltis, D.E., Soltis, P.S., 2005. The age of the angiosperms: a molecular timescale without a clock. *Evolution* 59, 1245—1258.
- Buerki, S., Forest, F., Acevedo-Rodríguez, P., Callmander, M.W., Nylander, J.A.A., Harrington, M., Sanmartín, I., Küpfer, P., Alvarez, N., 2009. Plastid and nuclear DNA markers reveal intricate relationships at subfamilial and tribal levels in the soapberry family (Sapindaceae). *Mol. Phylogenet. Evol.* 51, 238—258.
- Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7, 214.
- Friis, E.M., Pedersen, K.R., Crane, P.R., 2006. Cretaceous angiosperm flowers: innovation and evolution in plant reproduction. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 232, 251—293.
- Frohlich, M.W., Chase, M.W., 2007. After a dozen years of progress the origin of angiosperms is still a great mystery. *Nature* 450, 1184—1189.
- Klaassen, R., 1999. Wood anatomy of the Sapindaceae. *Int. Assoc. Wood. Anat. J.* 2 (suppl), 1–214.
- Müller, J., Leenhouts, P. W., 1976. A general survey of pollen types in Sapindaceae in relation to taxonomy. In: Ferguson, I.K., Müller, J. (Eds.), *The evolutionary significance of the exine*. London, Academic Press, pp. 407–445.
- Rabosky, D.L., 2006. LASER: a maximum likelihood toolkit for detecting temporal shifts in diversification rates from molecular phylogenies. *Evol. Bioinform. Online* 2, 257—260.
- Radlkofer, L., 1933. Sapindaceae. In: Engler, A. (Ed.), *Das Pflanzenreich IV*, 165 (Heft 98a-h). Leipzig, Verlag von Wilhelm Engelmann.
- Renner, S.S., Meyer, K., 2001. Melastomeae come full circle: biogeographic reconstruction and molecular clock dating. *Evolution* 55, 1315—1324.

Conclusion & Perspectives

Ronquist, F., 2004. Bayesian inference of character evolution. *Trends Ecol. Evol.* 19, 475—481.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572—1574.

Schaefer, H., Heibl, C., Renner, S.S. 2009. Gourds afloat: a dated phylogeny reveals an Asian origin of the gourd family (Cucurbitaceae) and numerous oversea dispersal events. *Proc. R. Soc. Lond. B.* 276, 843—851.

Wikström, N., Savolainen, S., Chase, M. W., 2001. Evolution of the angiosperms: calibrating the family tree. *Proc. R. Soc. Lond. B.* 268, 2211—2220.

Supplementary Material

Table 1. Sampling survey of the soapberry family covered by this study. Generic phylogenetic status are indicated based on Buerki et al. (2009) and chapter six (*).

Abbreviations: M: monophyletic; P: para-/polyphletic, -: generic status not assessed due to either the monotypic status of the genus or poor sampling (less than two taxa).

Family: Sapindaceae Jussieu: 104 genera / 1753 sp.

Phylogenetic status

Subfamily: Xanthoceroideae Thorne & Reveal: 1 genus / 1 sp.

Xanthoceras Bunge -

Subfamily: Hippocastanoideae Burnett: 5 genera / 129 sp.

Aesculus group: 3 genera / 16 sp.

Aesculus L. P

Billia Peyr.* -

Handeliodendron Rehder* -

Acer group: 2 genera / 113 sp.

Acer L. P

Dipteronia Oliver -

Subfamily: Dodonaeoideae Burnett: 18 genera / 127 sp.

Doratoxylon group: 5 genera / 12 sp.

Doratoxylon Thou. ex. Hook. f. M

Filicium Thw ex. Hook. f. M

Ganophyllum Blume M

Hippobromus Ecklon & Zeyher* -

Hypelate P. Browne* -

Dodonaea group: 13 genera / 115 sp.

Arfeuillea Pierre ex. Radlk. -

Averrhoidium Baillon -

Cossinia Comm. ex. Lam.* -

Diplokeleba N.E. Br.* -

Diplopeltis Endl. -

Dodonaea Miller M

Euphorianthus Radlk. -

Eurycorymbus Handel-Mazzetti -

Harpullia Roxb.* P

Llagunoa Ruiz & Pavon M

Loxodiscus Hook. f. -

Magonia A. St. Hil.* -

Majidea J. Kirk ex. Oliver -

Subfamily: Sapindoideae Burnett: 80 genera / 1496 sp.

Delavaya group: 2 genera / 2 sp.

Delavaya Franchet -

Ungradia Endl.* -

Phylogenetic status

Subfamily: Sapindoideae Burnett: 80 genera / 1496 sp.

Koelreuteria group: 4 genera / 19 sp.

<i>Erythrophysa</i> E. Mey ex. Arnott*	-
<i>Koelreuteria</i> Laxmann	P
<i>Smelophyllum</i> Radlk.*	-
<i>Stadmania</i> Lam.*	-

Schleichera group: 3 genera / 6 sp.

<i>Amesiodendron</i> Hu	-
<i>Paranephelium</i> Miq.	M
<i>Schleichera</i> Willd	-

Litchi group: 22 genera / 206 sp.

<i>Atalaya</i> Blume	P
<i>Blighia</i> Koenig	P
<i>Chytranthus</i> Hook. f.	M
<i>Cubilia</i> Blume	-
<i>Deinbollia</i> Schumach. & Thonn.	M
<i>Dimocarpus</i> Lour.	P
<i>Eriocoelum</i> Hook. f.	M
<i>Glennia</i> Hook. f.	M
<i>Haplocoelopsis</i> F.G. Davies	-
<i>Haplocoelum</i> p.p. (<i>H. gallense</i>)*	P
<i>Laccodiscus</i> Radlk.	-
<i>Lecaniodiscus</i> Planch. ex. Benth.	-
<i>Lepidopetalum</i> Blume	-
<i>Lepisanthes</i> Blume	P
<i>Litchi</i> Sonn.	-
<i>Nephelium</i> L.	M
<i>Pancovia</i> Willd.	-
<i>Pometia</i> Forst. & Forst.	M
<i>Pseudima</i> Radlk.	-
<i>Sapindus</i> L.*	P
<i>Tristira</i> Radlk.*	-
<i>Xerospermum</i> Blume*	-

Macphersonia group: 7 genera / 25 sp.

<i>Beguea</i> Capuron	M
<i>Conchopetalum</i> Radlk.	-
<i>Gereaaua</i> Buerki & Callm.	-
<i>Macphersonia</i> Blume	M
<i>Pappea</i> Eckl. & Zeyh.	-
<i>Plagioscyphus</i> Radlk.	M
<i>Pseudopteris</i> Baill.*	-

Cupania group: 26 genera / 428 sp.

<i>Alectryon</i> Gaertn.	M
<i>Arytera</i> Blume	M
<i>Castanospora</i> F. Muell.*	-
<i>Cupania</i> L.	M
<i>Cupaniopsis</i> Radlk.	P
<i>Diploglottis</i> Hook. f.	P
<i>Elattostachys</i> (Blume) Radlk.	M

	Phylogenetic status
Subfamily: Sapindoideae Burnett: 80 genera / 1496 sp.	
<i>Gongrodiscus</i> Radlk.	-
<i>Guioa</i> Cav.	P
<i>Jagera</i> Blume	M
<i>Lepiderema</i> Radlk.	M
<i>Matayba</i> Aublet.	P
<i>Mischarytera</i> (Radlk.) H. Turner	M
<i>Mischocarpus</i> Blume	M
<i>Molinaea</i> Comm. ex. Juss.	M
<i>Neotina</i> Capuron	-
<i>Podonophelium</i> Baillon	M
<i>Rhysotoechia</i> Radlk.	M
<i>Sarcopteryx</i> Radlk.	M
<i>Sarcotoechia</i> Radlk.	P
<i>Scyphonichium</i> Radlk.*	-
<i>Synimia</i> Radlk.	M
<i>Tina</i> Roem. & Schult.	M
<i>Tinopsis</i> Radlk.	-
<i>Toechima</i> Radlk.	M
<i>Vouarana</i> Aublet.	-
Tristiropsis group: 2 genera / 6 sp.	
<i>Dictyoneura</i> Blume	-
<i>Tristiropsis</i> Radlk.	-
Blomia group: 2 genera / 3 sp.	
<i>Blomia</i> Miranda	
<i>Haplocoelum</i> (<i>H. inopleum</i> + <i>H. foliosum</i>)	P
Melicoccus group: 3 genera / 63 sp.	
<i>Dilodendron</i> Radlk.*	-
<i>Melicoccus</i> P. Browne	M
<i>Talisia</i> Aublet.	P
Paullinia group: 11 genera / 738 sp.	
<i>Allophylus</i> L.*	P
<i>Athyana</i> (Giseb.) Radlk.	-
<i>Bridgesia</i> Bertero ex. Cambess.	-
<i>Cardiospermum</i> L.	M
<i>Diatenopteryx</i> Radlk.	M
<i>Guindilia</i> Hook & Arn.*	-
<i>Paullinia</i> L.	P
<i>Sapindus</i> p.p. (<i>S. oligophyllus</i>)	P
<i>Serjania</i> Miller	P
<i>Thouinia</i> Poit.	-
<i>Urvillea</i> Kunth	-

**APPENDIX I: OVERVIEW OF THE SUPERTREE METHODS USED IN THIS
STUDY**

APPENDIX: OVERVIEW OF THE SUPERTREE METHODS USED IN THIS STUDY

MRP approach This supertree reconstruction was developed independently by Baum (1992) and Ragan (1992). MRP uses additive binary coding (Farris et al., 1973) to represent the hierarchical structure of source trees as a series of matrix elements (Baum and Ragan, 1993). Each node (i.e., component *sensu* Wilkinson, 1994) on each source tree is represented by a binary matrix element (hereafter named matrix representation; MR). Terminal taxa delimited by each node are scored as 1 and all other taxa as 0. Missing taxa from an individual source tree are then coded as missing values. Trees are rooted either by an all-zero outgroup taxon (Ragan, 1992; Purvis, 1995) or by using an outgroup taxon common to all source trees. Regarding MR coding, Purvis (1995) argued that the elements derived from the source trees are lacking independence and hence added redundant information to the MR. He proposed to remove this apparent redundancy by coding only taxa among sister clades as 0, whereas all the others and the missing taxa are scored as missing values. Subsequently, Ronquist (1996) argued that the bias underlined by Purvis was not resolved by changing the scoring of the MR and proposed to remove it by weighting trees according to their relative size. He considered that the differences in the amount of information comprised in each contributing source tree could be removed by inversely weighting each tree according to the number of its nodes. Moreover, he proposed additional weighting schemes based on nodes support (i.e., bootstrap or decay index), which would considerably improve the analyses (Ronquist, 1996). Later on, Bininda-Emonds and Bryant (1998) considered that the use of parsimony algorithms allowing reversals (Fitch or Wagner parsimony) entailed that clades in the supertree might be partially supported by zeros in the MR. This would imply that those specific clades that were supported in some components of the source trees were artefacts. To avoid those artificial relationships they suggested the use of irreversible parsimony (Camin-Sokal parsimony; Camin and Sokal, 1965) in MRP analyses.

MinFlip approach This supertree method adapted by Eulenstein et al. (2004) to analyse large data sets uses the MR as a starting point. The flip supertree method is an

error-correction approach that seeks to find a minimum number of changes (called “flips”; i.e., from 0 to 1 or vice versa) to turn the supertree binary matrix into the MR of the input trees. The resulting supertrees are those corresponding to matrices in which the numbers of flips compared to the input trees has been minimized.

Split fit approach This method analyses the MR according to “compatibility”, a criterion where all the splits (i.e., nodes or components) from any proposed supertree are compared to those of the source trees (Rodrigo, 1993; Creevey and McInerney, 2005). The supertree that maximizes the number of shared splits with the source trees is selected.

Average consensus approach This method uses the least-squares criterion. First, it combines the information from the input trees by calculating the topological distance (path-length) between each taxa for all the input trees and subsequently combines the resulting matrices into an average path-length matrix. Missing topological distances (arising when the input trees are not sharing the same samples) are estimated from the available distances using ultrametric estimates. Second, the average path-length matrix is analysed using the least-squares criterion in order to obtain a consensus solution. The tree obtained through the AVCON is a solution that minimizes the sum of squared distances between the consensus and the source trees (Lapointe and Cucumel, 1997). This approach is the only supertree method that uses and provides branch length information when available. According to Creevey and McInerney (2005) the average path-length matrix might also be analysed using the neighbour-joining algorithm.

Most similar supertree approach This method also uses MR of equal branch length path-length distances. It compares each source tree individually to the supertree by subtracting the distance matrix derived from the supertree from that derived from the source tree (i.e., the distance matrix is calculated for both trees by counting the number of nodes between each taxa). The optimal supertree is the one that minimizes the weighted sum of the absolute differences in path-length between the supertree and the input trees (Creevey et al., 2004). The comparison of a supertree to an input tree is only possible if they share the same samples. Pruning the unshared taxa within the supertree circumvent this limitation.

MinCut approach This direct supertree method is derived from the Aho et al. (1981) algorithm which was developed for computing sciences. This algorithm relies on network theory and returns a graph displaying relationships among compatible trees sharing the same number of taxa. The Aho algorithm was adapted to the supertree problem by Semple and Steel (2000), in order to combine partially overlapping and conflicting input trees into an Adams consensus tree (for more information see Wilkinson et al., 2005). The Modified MinCut (MMC) method was developed by Page (2002) and differs from MC in ensuring that non contradicted relationships in source trees are present in the output trees.

References

Aho, A. V., Y. Sagiv, T. G. Szymanski, and J. D. Ullman. 1981. Inferring a tree from lowest common ancestors with an application to the optimization of relational expressions. *SIAM J. Comput.* 10:405-421.

Baum, B. R. 1992. Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. *Taxon* 41:3-10.

Baum, B. R., and M. A. Ragan. 1993. Reply to A. G. Rodrigo's "A comment on Baum's method for combining phylogenetic trees". *Taxon* 42:637-640.

Bininda-Emonds, O. R. P., and H. N. Bryant. 1998. Properties of matrix representation with parsimony analyses. *Syst. Biol.* 47:497-508.

Camin, J. H., and R. R. Sokal. 1965. A Method for deducing branching sequences in phylogeny. *Evolution* 19:311-326.

Creevey, C. J., and J. O. McInerney. 2005. Clann: investigating phylogenetic information through supertree analyses. *Bioinformatics* 21:390-392.

Creevey, C. J., D. A. Fitzpatrick, G. K. Philip, R. J. Kinsella, M. J. O'Connell, M. M. Pentony, S. A. Travers, M. Wilkinson, and J. O. McInerney. 2004. Does a tree-like phylogeny only exist at the tips in the prokaryotes? *Proc. R. Soc. B* 271:2552–2558.

Eulenstein, O., D. H. Chen, J. G. Burleigh, D. Fernandez-Baca, and M. J. Sanderson. 2004. Performance of flip supertree construction with a heuristic algorithm. *Syst. Biol.* 53:299-308.

Farris, J. S. 1973. Comparing the shapes of taxonomic trees. *Syst. Zool.* 22:50-54.

Lapointe, F. J., and G. Cucumel. 1997. The average consensus procedure: Combination of weighted trees containing identical or overlapping sets of taxa. *Syst. Biol.* 46:306-312.

Page, R. D. M. 2002. Modified mincut supertrees. *Lecture Notes in Computer Science* 2452:537-551.

Purvis, A. 1995. A modification to Baum and Ragan's method for combining phylogenetic trees. *Syst. Biol.* 44:251-255.

Ragan, M. A. 1992. Phylogenetic inference based on matrix representation of trees. *Mol. Phylogenet. Evol.* 1:53-58.

Rodrigo, A. G. 1993. A Comment on Baum method for combining phylogenetic trees. *Taxon* 42:631-636.

Ronquist, F. 1996. Matrix representation of trees, redundancy, and weighting. *Syst. Biol.* 45:247-253.

Appendix I: Supertree methods

Semple, C., and M. Steel. 2000. A supertree method for rooted trees. *Discrete Appl. Math.* 105:147-158.

Wilkinson, M., J. A. Cotton, C. Creevey, O. Eulenstein, S. R. Harris, F. J. Lapointe, C. Levasseur, J. O. McInerney, D. Pisani, and J. L. Thorley. 2005. The shape of supertrees to come: Tree shape related properties of fourteen supertree methods. *Syst. Biol.* 54:419-431.

Wilkinson, M. 1994. Common cladistic information and its consensus representation: Reduced Adams and reduced cladistic consensus trees and profiles. *Syst. Biol.* 43:343-368.

Appendix II: Curriculum Vitae

Sven Buerki

Date of Birth: 08.12.80 (Neuchâtel, Switzerland)

Nationality: Swiss

Marital status: Single

Current position: Assistant

Lab. Evolutionary Botany

Institute of Biology

University of Neuchâtel

11, rue Emile-Argand

2000 Neuchâtel

Switzerland

Phone number:

+41 (0)32 7182338

+41 (0)79 6574261

E-mail: sven.buerki@unine.ch



EDUCATION & RESEARCH

2000

Grade de bachelier ès science (type C), Lycée Blaise-Cendrars, Chaux-de-Fonds, Switzerland.

2000-2004

Undergraduate studies in Biology at the University of Neuchâtel (Switzerland).

2004

Course of tropical Botany, University of Geneva (Switzerland), under the direction of Prof Rodolphe-Edouard Spichiger.

2005

Master in evolutionary Ecology and Ethology from the University of Neuchâtel, with high distinction: "Biogeography and systematics of the genus *Gouania* (Rhamnaceae) in the western Indian Ocean"; advisor: Prof Philippe Küpfer and co-advisors: Pete B. Phillipson and Martin W. Callmander (Missouri Botanical Garden, USA).

Nov. 2005-

PhD position at the laboratory of Evolutionary Botany (University of Neuchâtel): "Worldwide biogeography and systematics of Sapindaceae: a molecular and taxonomic survey involving large data sets and novel methodological approaches": advisor: Prof Philippe Küpfer and co-advisor: Dr Nadir Alvarez.

INTERNSHIPS

Oct. – Nov. 2003

Internship in Madagascar: contribution to the checklist of vascular plants of the national park of Masoala; under the direction of Prof Philippe Küpfer and Dr Martin W. Callmander.

June-July 2004

Internship in the Conservatoire et Jardin botanique de la ville de Genève: determination of Malagasy collections; under the direction of Dr Laurent Gautier.

April-May 2005

Internship in the Muséum d'Histoire Naturelle de Paris: training in taxonomy; under the direction of Pete P. Phillipson from the Missouri Botanical Garden (USA) (based at the Muséum d'Histoire Naturelle de Paris for the "Catalogue of the vascular plants of Madagascar" project).

April-June 2008

Internship in the Jodrell Laboratory (Royal Botanic Gardens, Kew, London, U.K.): training in phylogenetic methods; under the direction of Dr Félix Forest.

Feb. 2009

Internship in the Real Jardín Botánico: training and developing parametric biogeographic methods; under the direction of Prof Isabel Sanmartín.

EMPLOYMENT

2003-2005: Younger assistant at the laboratory of Evolutionary Botany (University of Neuchâtel, Switzerland).

Since Nov. 2005: PhD student at the laboratory of Evolutionary Botany (University of Neuchâtel, Switzerland).

GRANT SUPPORT FOR RESEARCH

2004: Fieldwork grant (Matthey-Dupraz/Wüthrich Fund): 2500 CHF

2008: SYNTHESYS grant to visit the Real Jardín Botánico (CSIC) in Madrid: 2000 Euros

STUDENTS SUPERVISION

2006-2007: Co-supervision of two Master's thesis students in Neuchâtel (**Julien Burnier & Aurélien Labhardt**) on the evolution of the ploidy level in the two alpine plant species *Ranunculus kupeferi* and *Bupleurum ranunculoides*.

TEACHING EXPERIENCE

2005-: Practices on *Classics in Biology* at the University of Neuchâtel (Graduate level, 4 hours/week).

2008-2009: Lectures on *Systematics and Phylogenetic methods* at the University of Neuchâtel (Switzerland) (Undergraduate level, 6 hours/year).

AWARDS

2005: Prize Louis Paris for the best Master thesis

2005: Prize Jean Landry for the best average (5.56/6)

FIELD EXPERIENCE

2003-2006: four fieldworks (3 months each) in Madagascar for PhD studies.

2005: one fieldwork (2 weeks) in Italy and Croatia to collect species of *Aegilops* with Dr. Nils Arrigo.

EXTERNAL REFEREE

Journal: Molecular Ecology

LANGUAGES

Mother tongue: French

Fluent in English

Good knowledge in German

HOBBIES

Botany
Trekking
Photography
Basket-ball

PUBLICATIONS IN INTERNATIONAL PEER-REVIEWED JOURNALS

1. Laivao OM., Callmander MW, **Buerki S.** 2006. Sur les *Pandanus* (Pandanaeae) à stigmatés saillants de la Côte Est de Madagascar. *Adansonia, sér. 3*, 28: 267-285.
2. Laivao OM, Callmander MW, **Buerki S.** 2007. Révision de *Pandanus* sect. *Foullioya* à Madagascar. *Adansonia sér. 3*, 29(1): 39-57.
3. Callmander MW, Phillipson PB, **Buerki S.** 2008. Révision du genre *Bathiorhamnus* Capuron (Rhamnaceae) endémique de Madagascar. *Adansonia sér. 3*, 30: 151-170.
4. Callmander MW, **Buerki S.**, Wohlhauser S. 2008. A new threatened species of Pandanaeae from northwestern Madagascar: *Pandanus sermolliana*. *Novon*, 18: 421-424.
5. **Buerki S.**, Forest F, Acevedo-Rodríguez P, Callmander MW, Nylander JAA, Harrington M, Sanmartín I, Küpfer P, Alvarez N. 2009. Plastid and nuclear DNA markers reveal intricate relationships at subfamilial and tribal levels in the soapberry family (Sapindaceae). *Molecular Phylogenetics and Evolution*, 51: 238-258.
6. **Buerki S.**, Callmander MW, Schüpfer F, Ravokatra M, Küpfer P, Alvarez N. 2009. Malagasy *Dracaena* Vand. ex L. (Ruscaceae): an investigation of discrepancies between morphological features and spatial genetic structure at a small evolutionary scale. *Plant Systematics and Evolution*, 280: 15-28.
7. Kissling J, **Buerki S.**, Mansion G. *Klackenbergia* (Gentianaceae - Exaceae): a new endemic genus to Madagascar. *Taxon*, in press.

8. Burnier J¶, **Buerki S**¶, Arrigo N, Küpfer P, Alvarez N. Genetic structure and evolution of Alpine polyploid complexes: *Ranunculus kuepferi* (Ranunculaceae) as a case study. *Molecular Ecology*, in press. ¶ These authors have contributed equally and are considered joint first authors.
9. Callmender MW, Rakotoavao C, Razafitsalama J, Phillipson PB, **Buerki S**, Hong-Wa C, Rakotoarivelo N, Andriambololonera S, Koopman MM, Derooin T, Johnson DM, Andriamandranto R, Solo S, Labat J-N, Lowry II PP. New species from two unknown and highly threatened mountainous areas in north-western Madagascar: the Galoka and Kalabenono massifs. *Candollea*, in press.
10. **Buerki S**, Phillipson PB, Callmender, MW. A taxonomic revision of the genus *Gouania* Jacq. (Rhamnaceae) in Madagascar and the islands of the western Indian Ocean. **Accepted in *the Annals of the Missouri Botanical Garden*.**
11. **Buerki S**, Phillipson PB, Lowry II PP, Callmender MW. Molecular phylogenetics and morphological evidence support recognition of *Gereaua*, a new endemic genus of Sapindaceae from Madagascar. **Accepted in *Systematic Botany* pending minor revisions.**

SUBMITTED PAPERS

12. **Buerki S**, Forest F, Alvarez N. Comparative Performance of Supertree Algorithms in Large Data Sets using the Soapberry Family (Sapindaceae) as a case study. **In revision at *Systematic Biology*.**
13. **Buerki S**, Callmender MW, Lowry II PP, Phillipson PB. A synoptic revision of the genus *Lepisanthes* Blume (Sapindaceae) in Madagascar. **In revision at *Adansonia*.**

PAPERS IN PREPARATION

14. **Buerki S**, Forest F, Alvarez N, Nylander JAA, Arrigo N, Sanmartín I. New insights on parametric biogeography based on the worldwide soapberry family (Sapindaceae). **To be submitted to *Systematic Biology*.**
15. **Buerki S**, Alvarez N, Sanmartín I, Forest F. Abiotic factors as driving force to promote speciation in Sapindaceae. **To be submitted to *PNAS*.**
16. Arrigo N, Felber F, **Buerki S**, Alvarez N, David J, Guadagnuolo R. Ancient and recent evolutionary history of *Aegilops geniculata* Roth, a Mediterranean wild relative to the wheat. **To be submitted to *Molecular Ecology*.**
17. **Buerki S**, Callmender MW, Phillipson PB. Synoptic revision of the genus *Molinaea* Comm ex Juss. (Sapindaceae) in Madagascar. **To be submitted to *Adansonia*.**

PARTICIPATION IN SCIENTIFIC SYMPOSIA & SEMINARS

Conferences

Buerki S., Callmender MW, Küpfer P (2007) Phylogenetic analyses of the Malagasy Cupanieae reveal a high degree of polyphyly at the generic level. 18th AETFAT congress. February 2007. Yaoundé, Cameroon.

Buerki S, Sanmartin I, Alvarez N (2007) Comparative performance of the Supertree algorithms in large data sets: the case of the soapberry family (Sapindaceae). November 2007. Real Jardín Botánico, Madrid, Spain.

Buerki S, Sanmartin I, Alvarez N (2008) Untangling the systematics of the soapberry family (Sapindaceae): supertrees, biogeography and character evolution. January 2008. Royal Botanic Gardens, Kew, London, United Kingdom.

Buerki S, Forest F, Acevedo-Rodriguez P, Callmander MW, Nylander JAA, Küpfer P., Alvarez N (2008) Towards a new classification for the soapberry family (Sapindaceae). June 2008. Conservatoire et Jardins botanique de la ville de Genève, Geneva, Switzerland.

Vary L., **Buerki S** (2008). Breeding system evolution in the Malagasy Sapindaceae. Botany without borders. 26-30 July 2008. University of British Columbia, Vancouver, Canada.

Buerki S, Forest F, Alvarez N, Nylander JAA, Arrigo N, Sanmartín I. (2009) Out of Asia and into the tropics: phylogeny and biogeography of the Sapindaceae. 15th of April 2009. SYNTHESIS final report. Real Jardín Botánico, Madrid. Spain.

Posters

Buerki S, Sanmartin I, Callmander MW, Alvarez N (2008) Systematics and biogeography of the soapberry family (Sapindaceae). NCCR International Conference on Plant Species Concepts and Evolution (in honour of the retirement of Philippe Küpfer). 30 January–1 February 2008. University of Neuchâtel, Neuchâtel, Switzerland.

Buerki S, Callmander MW, Schüpfer F, Ravokatra M, Küpfer P, Alvarez N (2008) Solving the inconsistency between morphology and phylogeny in the Malagasy *Dracaena* Vand. ex. L. (Ruscaceae). NCCR International Conference on Plant Species Concepts and Evolution (in honour of the retirement of Philippe Küpfer). 30 January–1 February 2008. University of Neuchâtel, Neuchâtel, Switzerland.