THESIS ABSTRACT

HARRIS, AJ. Molecular and Morphological Inference of the Phylogeny, Origin, and Evolution of *Aesculus* L. (Sapindales). (Under the direction of Qiu Yun Xiang.)

Aesculus L. (Hippocastanaceae or Sapindaceae) is a Laurasian disjunct genus of 13-19 species in four endemic areas. The genus is a good model for biogoegraphic study due to a small number of species, a rich fossil record, and an intercontinental disjunct distribution of extant species. Therefore, further study of this genus may offer important insights into biogeographic history of modern plant biota in the northern hemisphere. Conflicting biogeographic hypotheses have been proposed for *Aesculus* in previous studies. The main goal of the study is to evaluate these alternative hypotheses using new and more data. I first reconstructed a more robust phylogeny of *Aesculus* by increasing taxon and character sampling and integrating fossils, morphology, and molecular data. Using the phylogeny as the basis, I performed biogeographic analysis and divergence time dating to determine the place and time of origin and migration pattern of the genus. This analysis provided insight into the timing of speciation and historical events responsible for modern distribution of the genus. DNA sequences of several molecular regions including *rps*16, *trn*HK, and *mat*K from chloroplast genome and the ITS of nuclear ribosomal gene and intron 1 of the *LEAFY* (*LFY*) gene from the nuclear genome. These were used for phylogenetic analyses in combination with morphological data. The results largely agree with previous molecular analyses preformed using less inter and intra-specific sampling and molecular data from two gene regions. The major difference was the placement of Sect. *Aesculus*. The phylogeny reconstructed in this study revealed three major lineages in *Aesculus* including an Asian clade consisting of all Asian constituents of Sect. *Calothyrsus*, a clade containing *A. californica* and Sect.

Macrothyrsus, and a clade consisting of Sect. *Aesculus*, Sect. *Pavia*, and Sect. *Parryana*. Section *Aesculus* was placed as sister to Sect. *Parryana* + Sect. *Pavia*, a relationship previously unreported. Relationships between these three major clades remained incompletely resolved, probably due to a rapid radiation of the genus in its early evolutionary history. For reconstruction of biogeography in cases of phylogenetic uncertainty, we developed a probabilistic approach with the method of dispersal - vicariance analysis (DIVA). In the analysis, 100 phylogenetic trees from Bayesian analysis were optimized for ancestral distribution with DIVA. The ancestral area of the root of a given lineage was determined as the most probable area(s) summed over the results from the 100 tree analyses. This new application of DIVA was shown to work with incompletely resolved phylogenies and to provide probability on estimated ancestral areas. Our results from DIVA, using a more traditional approach to this software, and divergence time dating using a Bayesian approach without a constant clock supported an Early Tertiary origin of *Aesculus* in Northeastern Asia and probably also western North America as an element of the high-latitude part of the boreotropical flora. Migrations to other areas occurred in both eastward and westward directions during the early Tertiary. This biogeographic history is substantially different from all previously proposed hypotheses, highlighting the importance of including fossils, particularly the newly discovered early Tertiary fossils of *A. hickeyi*. Incongruent positions of *A. californica* in the chloroplasts and nuclear DNA phylogenies supported the previous hypothesis that an ancestral polymorphism and subsequent lineage sorting of haplotypes had generated the discrepancy in the placements of *A. californica*.

MOLECULAR AND MORPHOLOGICAL INFERENCE OF THE PHYLOGENY, ORIGIN, AND EVOLUTION OF AESCULUS L. (SAPINDALES)

By

AJ HARRIS

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Plant Biology

Raleigh, North Carolina

2007

APPROVED BY:

_____	_____	_____
Dennis Werner	Candace Haigler	Deyu Xie

_____
Jenny Xiang
Chair of Advisory Committee

ACKNOWLEDGEMENTS

For her time and dedication spent in assisting me to prepare this work, for her continued confidence in my abilities, and for enthusiastic interest in my success and professional development throughout preparation of this work and pursuance of the MS degree, I would like to thank my advisor Dr. Jenny Xiang.

I would like to acknowledge all members of my committee, Dr. Xiang, Dr. Candace Haigler, Dr. Deyu Xie, and Dr. Dennis Werner, for their numerous contributions to this research and this document.

I am indebted to Dr. James Mickle (NCSU) for helpful discussions regarding interpretation and application of paleobotanical sources, to David Thomas (NCSU) for his instruction in lab techniques and use of software and to Dr. Brian Weigmann (NCSU) for his assistance with the technicalities of phylogenetic theory and application. I acknowledge Holly Forbes (UC Berkeley) for sending fresh leaf material and the Harvard University Herbarium for the loan of herbarium specimens. For his encouragement and insight in development of Chapter 2, I am indebted to Dr. Francois Lutzoni (Duke University). I also thank Sue Vitello for guiding me through the network of administrative requirements.

For financial support, I thank the NCSU Department of Plant Biology. Additionally, I am grateful for support for travel from the Deep Time and MORPH grants as well travel support from the NCSU Graduate School.

DEDICATION

The work herein is dedicated to four wonderful individuals:

To my mom, Nancy E. Harris, who told me that I was capable of doing anything;

To my dad, Joe K. Harris, who has taught me the value of dedication to work and the value of time for play;

To my partner, Andrew Dabbs, who has been patient while this work has been in progress and who has been both a constant cheerleader and an available shoulder to cry on;

And finally, to my grandmother, Rachel Easler, who I hope heard and understood when I said that everything I ever accomplish will be with her in mind.

BIOGRAPHY

AJ Harris was born in Raleigh, North Carolina to Joe and Nancy Harris on April 28, 1978 and was primarily raised on a hay and, later, a cattle farm in southern Alamance County, North Carolina. The author completed a high school equivalency test in 1995. In May of 2005 she graduated from North Carolina State University with a B.A. in Religious Studies and a minor in Botany. The author has been honored by induction into the honor societies Phi Beta Kappa and Phi Kappa Phi (2005). Ms. Harris currently resides with her partner, Andrew Dabbs, to whom she has been married for six years.

TABLE OF CONTENTS

List of Tables for Chapter 1 and 2

Chapter 1

Chapter 2

LIST OF FIGURES FOR CHAPTER 1 AND 2

Chapter 1

Chapter 2

LIST OF ABBREVIATIONS FOR CHAPTER 1 AND 2

ARC (software) – Ambiguous Regions Coding

CSA – Central and South America

DIVA (software) – DIspersal Vicariance Analysis

EA – East Asia

eNA – western North America

Eu – Europe

Gr – Greenland region, including the DeGreer and Thulean Landbridges

ITS – Internal Transcribed Spacer

MCMC – Markov Chain Monte Carlo

MP – maximum parsimony

MY – million years

MYA – million years ago

PAML (software) – Phylogenetic Analysis by Maximum Likelihood

PAUP* (software) – Phylogenetic Analysis Using Parsimony

PP – Bayesian posterior probabilities

P – probability

wNA – western North America

TAXONOMY AND SYSTEMATICS OF *AESCULUS* L.

Aesculus L. is a small genus of woody trees and shrubs. The genus is characterized by a showy terminal panicle of zygomorphic flowers, palmately compound leaves, 1-3 loculed capsule fruits that may be spiny warty or smooth and diffuse porous wood (Hardin 1957). The genus is of horticultural importance and has medicinal and folk uses. For example, saponaceous extracts from the seeds of some species show pharmacological promise as anti-inflammatory agents and for their anti-mutagenic and hypoglycemic activities (see Matsuda 1997, Wei 2004, Kimura et al. 2006). Fruits of *A. hippocastanum* L., the single extant European species of *Aesculus*, have a long history of ethnobotanical use in Europe. Ground as animal fodder, they were thought to enrich the quality of milk production in cattle and treat pulmonary disorder in horses (Barton and Castle 1887). In traditional Chinese medicine, the dried, ripe seeds of *A. chinensis* Bunge have been used to treat distension and abdominal pains (Wei et al. 2004). The eastern North American species of *Aesculus* have toxic chemical constituents, which have been reported to be neurotoxic in animals (Knight and Walter 2003). Some species of *Aesculus* also have socio-religious significance. *Aesculus chinensis* was treated by early Chinese Buddhists as a substitution for the Indian-grown sala tree (*Shorea robusta* Gaertn.), important in the religious biography of the Buddha (Burkhill 1946). *Aesculus indica* may have been used similarly by early Buddhists in the Kashmir region (Burkhill 1946).

Aesculus is one of the two genera traditionally classified in the Hippocastanaceae DC nom. con.; a ditypic family including *Aesculus* and *Billia* Peyr. A cladistic analysis of morphological data of Sapindales by Judd, Saunders and Donoghue (1994) including representative of Hippocastanaceae, Aceraceae, and Sapindaceae revealed that

Hippocastanaceae was derived from within the Sapindaceae and is most closely related to the monotypic Chinese genus *Handeliodendron* Rehder. Analyses of *mat*K, *rbcl,* and combined analysis of these markers by Harrington et al. (2005) supports Judd and colleagues' finding that *Handeliodendron* shares a more recent common ancestor with genera of Hippocastanaceae than with taxa of Sapindaceae. Molecular evidence from the chloroplast-encoded genes *rbcl* and *mat*K (Harrington et al. 2005) shows strong support for a clade including Sapindaceae, Aceraceae, and Hippocastanaceae, and the combined analysis of these markers show moderate to high support (70% bootstrap)for a clade consisting of *Billia* + *Aesculus* + *Handeliodendron* + *Aceraceae* derived within Sapindaceae (Harrington et al. 2005). In analyses of individual genes (Harrington et al. 2005), *rbcl* and *mat*K MP tree topologies supported a sister relationship of Hippocastanaceae including *Handeliodendron* + Aceraceae to Sapindaceae. The authors indicated that a broader circumscription of Sapindaceae would allow for inclusion of Aceraceae and Hippocastanaceae despite possible resolution of these taxa in a position basal to the rest of Sapindaceae. To reflect the findings from phylogenetic analyses, Hippocastanaceae is currently treated as a member of Sapindaceae by Stevens (2001 ff.).

The two genera most closely related to *Aesculus* are *Billia* and *Handeliodendron*, *Billia* has a Central and South American distribution (Hardin 1957a, b) while *Handeliodendron* is endemic in southern China. *Aesculus* is traditionally divided into five sections (Table 1). These are Sect. *Aesculus* (2 species) Sect. *Calothyrsus* (5-11 species), Sect. *Macrothyrsus* (1 species), Sect. *Parryana* (1 species), and Sect. *Pavia* (4 species). These sections were described in detail by Hardin (1957a, 1957b, 1960). In his monograph of the Hippocastanaceae, Hardin recognized 13 species of *Aesculus*. These

were *A. assamica* Griff., *A. californica* (Spach.) Nutt., *A. chinensis* Bunge, *A. indica* (Wall. ex Camb.) Hook, and *A. wilsonii* Rehder. in Sect. *Calothyrsus*; *A. glabra* Willd., *A. flava* Sol., *A. pavia* L., and *A. sylvatica* Bartram in Sect. *Pavia*; *A. hippocastanum* L. and *A. turbinata* Blume in Sect. *Aesculus*; *A. parryi* Gray in Sect. *Parryana*; and *A. parviflora* Walter in Sect. *Macrothyrsus*. Since the time of Hardin's publication, 6 East Asian *Aesculus* species have been described by the Chinese botanists Hu and Fang. They are *A. chuniana* Hu & Fang, *A. lansangensis* Hu & Fang, *A. megaphylla* Hu & Fang, *A. polyneura* Hu & Fang, *A. tsiangii* Hu & Fang, and *A. wangii* Hu. All of these species are putative constituents of Sect. *Calothyrsus*. The species status of these taxa has been questioned (Xiang et al. 1998, Turland and Xia 2005).

Recently, Turland and Xia (2005) reduced *A. wilsonii* to a variety of *A. chinensis*; *A. chinensis* var. *wilsonii* (Rehder.) Turland & Xia. This is consistent with reports that *A. chinensis* is found only in locations where cultivation, rather than natural occurrence, may be responsible for its presence (Hardin 1957b). The new treatment was to use *A. chinensis* var. *wilsonii* for the naturally occurring wild populations and *A. chinensis* var. *chinensis* to represent one or more cultivated forms. Turland and Xia (2005) also propose that *A. wangii* is *A. assamica*. These hypotheses remain to be tested.

Aesculus is a northern hemisphere or Laurasian disjunct genus (Wen 1999). *Aesculus* species occur in Europe, East Asia, western North America, and eastern North America. Section *Pavia* (including *A. glabra*, *A. flava*, *A. pavia*, and *A. sylvatica*), one of the three polytypic sections, occurs in eastern North America. Section *Aesculus* is disjunct in East Asia (*A. turbinata* in Japan) and Europe (*A. hippocastanum*). Section *Calothyrsus* includes *A. assamica*, *A. californica*, *A. chinensis*, *A. indica*, and *A. wilsonii*

Rehder. and is disjunct in East Asia and western North America. All species of this section occur in East Asia except *A. californica*, a western North American species. The treatment of *A. californica* in Sect. *Calothyrsus* is supported by morphological evidence (Hardin 1957a, Forest et al. 2001) and by sequence data from the chloroplast gene *mat*K (Xiang et al. 1998). However, phylogenetic reconstruction from the nuclear ribosomal marker ITS (Internal Transcribed Spacer) shows that *A. californica* is more closely related to the monotypic eastern North American section; Sect. *Macrothyrsus* (*A. parviflora*, endemic in the southeastern United States) (Xiang et al. 1998). Thus, the monophyly of Sect. *Calothyrsus* remains uncertain leaving the disjunct distribution in East Asia and western North America in question.

There remains uncertainty regarding the position of the traditional Hippocastanaceae within Sapindaceae, the relationships of taxa within *Aesculus*, and the species status of some *Aesculus* species. Systematic hypotheses presented above and the biogeographic history of *Aesculus* remain to be evaluated by further studies.

REFERENCES

Barton, B. and T. Castle. *The British Flora Medica: A history of the medicinal plants of Great Britton.* Chatto and Windus, Piccadilly: London. (1877)

Burkhill, J. On the dispersal of plants most intimate to Buddhism. *Journal of the Arnold Arboretum*. 27(4): 327-339 (1946)

Forest, et al. A morphological phylogenetic analysis of *Aesculus* L. and *Billia* Peyr. (Sapindaceae). *Canadian Journal of Botany*. 79: 154-169 (2001)

Hardin, J. A revision of the American Hippocastanaceae. *Brittonia*. 9(3): 145-171 (1957a)

Hardin, J. A revision of the American Hippocastanaceae II. *Brittonia*. 9(3): 173-195 (1957b)

Hardin, J. Studies in the Hippocastanaceae: Old World Species. *Brittonia*. 12(1): 26-38 (1960)

Harrington, M. et al. Phylogenetic inference in the Sapindales sensu Iato using plastid *mat*K and *rbcl* DNA sequences. *Systematic Botany*. 30(2): 366-382 (2005).

Judd, W., R. Sanders, and M. Donoghue. Angiosperm family pairs: preliminary phylogenetic analysis. *Havard Papers in Botany*. 1:1-15 (1994)

Kimura, H. et al. Identification of novel saponins from edible seeds of Japanese horse chestnut (*Aesculus turbinata* Blume) after treatment with wooden ashes and their nutraceutical activity. *Journal Of Pharmaceutical And Biomedical Analysis*. 41(5): 1657-1665 (2006).

Knight, A. and R. Walter. Plants affecting the nervous system. *A Guide to Plant Poisoning in Animals in North America*. Eds. A. Knight and R. Walter. Teton NewMedia: Jackson, WY. (2003).

Matsuda, H. Antiinflammatory effects of escins Ia, Ib, IIa, and IIb from horse chestnut, the seeds of *Aesculus* hippocastanum L. *Bioorganic and Medicinal Chemistry Letters*. 7(13): 1611-1616 (1997)

Turland, N. and N. Xia. A new combination in Chinese *Aesculus* (Hippocastanaceae). *Novon*. 15(4): 488-489 (2005)

Wei et al. Antiinflammatory triterpenoid saponins from the seeds of *Aesculus chinensis*. *Chemical and Pharmacological Bulletin*. 52(10): 1246-1248 (2004)

Wen, J. Evolution of eastern Asian and eastern North American Disjunct Distributions in flowering plants. *Annual Review of Ecology and Systematics*. 30: 421-455 (1999)

Xiang, Q.Y. et al. Origin and biogeography of *Aesculus* L. (Hippocastanaceae): A molecular phylogenetic perspective. *Evolution* 52(4):988-997 (1998)

CHAPTER 1

THE PHYLOGENY AND BIOGEOGRAPHY OF *AESCULUS* L. (SAPINDALES) INFERRED FROM MOLECULAR, MORPHOLOGICAL, AND FOSSIL EVIDENCE

INTRODUCTION

The classical interpretation of the discontinuous distribution of plants in the Northern Hemisphere is that a widespread mesophytic Tertiary flora, that was developed from an early Tertiary boreotropical flora (Wolfe 1975; Tiffney 1985a, 1985b), was disrupted by a series of geological events and climatic changes. These events include the rise of the Rocky Mountains, the disappearance of key landbridges, and repeated glaciation (see Wolfe 1975, Rosen 1978, Wen 1999, Tiffney and Manchester 2001, Donoghue and Smith 2004). Climatic changes caused by the Rocky Mountain rainshadow in central and western North America and spread of glaciers southward in Europe and North America resulted in extinctions and migrations, leaving remnant populations of the ancestral mesophytic flora isolated in two or more of the following areas: eastern North America; East Asia; Europe; and western North America. Among these, the most commonly observed and reported disjunction has been between East Asia and eastern North America. In many cases extant representatives of disjunct genera are absent altogether from western North America and Europe. Hence, this continental-scale taxon disjunction has commonly been called the East Asian-eastern North American disjunction due to its prominence; although disjunct patterns involving additional areas (like Europe and western North America) are also present (Wen 1999).

The origins of discontinuous distributions of plants in the northern Hemisphere has long been an interest of botanists, leading to many recent studies using a phylogenetic approach (Wen 1999; Donoghue and Moore 2003; Donoghue and Manos 2001; Xiang et al., 1998, 2000, 2005). Recently, Donoghue and Smith (2004) assembled the patterns of plant biogeographic histories in the northern Hemisphere based on meta-analysis of

multiple independent phylogenetic studies. They concluded that the eastern Asian-eastern North American disjunction is still the major pattern of intercontinental disjunction of plants in the northern hemisphere. Their analyses found a major "out of Asia" pattern of plant migration. However, they indicated that this conclusion is tentative due to several limiting factors of the study. The authors asserted that the selection of genera for phylogenetic and biogeographic studies may have biased conclusions, e.g., more genera with species occurring in eastern North America and eastern Asia were selected for phylogenetic studies, while fewer genera with species in Europe and western North America were selected for study, causing a bias toward the pattern of eastern Asia-eastern North American disjunction. Another limiting factor of the Donoghue and Smith study was that no fossils were included in the biogeographic or phylogenetic analyses of disjunct lineages, which could have biased the "out of Asia" migration pattern.

The *Aesculus* disjunction is part of a broader pattern of intercontinental disjunction in the northern hemisphere (Wen 1999, Donoghue and Smith 2004). A detailed biogeographic study of the genus would contribute to a better understanding of the origin and evolution of modern distribution of plants in the northern hemisphere. The genus is attractive for phylogenetic study for several reasons: (1) small number of species (13-19 species); (2) extant species distribution in four isolated Laurasian regions: East Asia (5-11 species), eastern North America (5 species), Europe (1 species), and western North America (2 species); and (3) a rich *Aesculus* fossil record, which allows for divergence time dating and for a more accurate reconstruction of the biogeographic history of the genus (place of origin, past distribution, and migration, etc.) Phylogenetic and biogeographic studies of *Aesculus* make a valuable contribution to the ongoing study

of the processes and patterns that characterize the Laurasian disjunction, and to the evaluation of the patterns detected in Donoghue and Smith (2004).

Previous studies of *Aesculus* have disagreed on the phylogeny, evolutionary origin, and biogeographic history of *Aesculus*. Hardin (1957a), based on his study of morphological characters and ground-plan mapping of 19 characters, proposed that *Aesculus* arose in Central or South America from a *Billia*-like ancestor. The linage moved northward in Tertiary times (or earlier). The northward migration resulted in divergence into an Appalachian lineage, now Sect. *Pavia*, while the other lineage moved up the west coast of North America and crossed the Bering Land Bridge into Asia and Europe. *A. parryi* and *A. californica*, western North American taxa, remain as relics of the northwestern migration of the genus. Section *Macrothyrsus* originated in the highlands of Mexico and migrated northward into the Appalachians and later to the coasts of Alabama and Georgia. Hardin's phylogeny is reproduced in Fig. 1a.

Xiang and colleagues (1998) conducted a phylogenetic study of the genus including *Handeliodendron*, *Billia* sp., and 12 of the 13 *Aesculus* species recognized by Hardin, and one of the six species described by Wang and Hu; *A. wangii*. The study reconstructed the phylogeny of *Aesculus* using DNA sequences of chloroplast gene *mat*K and nuclear ITS region (Internal Transcribed Spacer) of the ribosomal genes, as well as Hardin's morphological matrix extended to include the Old World taxa. The relationships among species suggested by the *mat*K and ITS data were largely congruent regarding the monophyly of sections except the placement of *A. californica*. The *mat*K data of Xiang et al. (1998) (Fig. 1b) suggested that *A. californica* was closely related to the Asian species, basal in the Sect. *Calothyrsus* clade, supporting Hardin's classification, while the

ITS data (Fig. 1c) suggested a closer relationship of *A. californica* with eastern North American taxa. Their analysis of combined data from ITS and morphology supported the ITS data and resulting tree topology in which *A. californica* was placed outside of Sect. *Calothyrsus*. This discrepancy regarding the relationship of *A. californica* was explained by plastid lineage sorting (Xiang et al. 1998). These researchers proposed a polymorphic *Aesculus* ancestor having plastid haplotypes A and B, with haplotype B lost in all Old World species and haplotype A lost in all New World species with the exception of *A. californica*. Using the ITS phylogeny and fossil calibrations, Xiang and colleagues derived a strict molecular clock to estimate divergence times of lineages. Ancestral character state mapping was used to reconstruct the biogeographic history of *Aesculus*. Based on the results, Xiang et al. (1998) proposed that *Aesculus* arose in the high latitudes of East Asia as part of the northern hemisphere boreotropical flora (Wolfe 1975). The genus diversified in northeast Asia from a pre-*Aesculus* ancestor. One lineage moved southward while another moved westward to Europe and eastward North America. The latter lineage was disrupted by glaciation in Eurasia in the Miocene, resulting in the disjunction of Sect. *Aesculus*. The North American part of this lineage diverged into two groups; the Sect. *Pavia* + Sect. *Parryana* and *A. californica* + Sect. *Macrothyrsus*. Eastern and western members of these groups were isolated from one another probably in the Oligocene by the Rocky Mountain rainshadow (Xiang et al. 1998).

Forest and collaborators (2001), conducted a phylogenetic study of *Aesculus* using morphological characters, which also included *Billia*, *Handeliodendron* and additional Sapindaceae outgroup taxa from the tribes Harpullieae, Koelreuterieae, and

Doratoxyleae. The 13 *Aesculus* species recognized by Hardin and both species of *Billia* were examined in this study. The resulting matrix consisted of 39 characters coded from fresh material, herbarium specimens, and literature. Results of their parsimony analysis (Fig. 1d) supported the placement of *A. californica* in Sect. *Calothyrsus* and in a basal position within that section, congruent with the *mat*K data of Xiang and colleagues (1998). Using Bremer's (1992) Ancestral Areas (AA) method for biogeographic analysis, Forest et al. (2001) proposed that *Aesculus* evolved in the high latitudes of North America as a component of the boreotropical flora. The *Aesculus* lineage spread across the Bering Land Bridge and occupied parts of North America and Eurasia. The Sect. *Parryana* lineage became isolated early from the remainder of the genus. Tertiary cooling and the discontinuity of the Bering Land Bridge separated the predominately Old World Sect. *Calothyrsus* + Sect. *Aesculus* from the New World Sect. *Pavia* + Sect. *Macrothyrsus* clade. *Aesculus californica* was isolated from the rest of Sect. *Calothyrsus* on the eastern side of the Bering Land Bridge. Section *Aesculus* and Sect. *Calothyrsus* diverged during the Tertiary and spread to Europe and Japan. Forest et al. (2001) also concluded that the xeric characters common to *A. californica* and *A. parryi* are a result of adaptations to Tertiary drying in western North America. Forest and colleagues did not consider the *Aesculus* fossil record when developing this scenario for the biogeographic history of the genus.

At the time of the Xiang et al. study, the oldest known *Aesculus* fossils were found in a Paleocene/Lower Eocene (65mya-54.8mya) bed in Anadirka, western Kamchatka in the Russian far East (Budantsev 1983, Manchester 2001). The biogeographic history inferred by Xiang et al. (1998) was congruent with the fossil

evidence available at that time. However, Manchester recently reported a Paleocene *Aesculus* fossil from western North America (2001). The discovery of this fossil species may have important implication on the early biogeographic history of the genus. Inclusion of *A. hickeyi* Manch. in the biogeographic reconstruction of *Aesculus*, for example, may result in a biogeographic history that differs significantly from that proposed by Xiang et al. (1998). Another limitation of the study of Xiang et al. (1998) is that the deep nodes of the phylogenies inferred by *mat*K and ITS data were weakly supported, leaving uncertainties about the conclusions. Therefore, it is necessary to reevaluate the phylogeny and biogeography of the genus with additional data from more genes and the new fossil data.

The primary goal of the present study is to evaluate the previous hypotheses regarding the placement of *A. californica* and the biogeographic history of *Aesculus*. Therefore specific objectives of this study are to (1) reconstruct a more robust phylogeny of *Aesculus* using more characters including morphology of living and fossil taxa and DNA sequences from five regions, *LEAFY* (*LFY*), *trn*HK, and *rps*16 in addition to *mat*K and ITS used in previous studies; (2) estimate lineage divergence time using the Bayesian MCMC (Markov Chain Monte Carlo) software Multidivtime (Thorne 2003), which allows rates of molecular evolution to be different among genes and lineages; and (3) reconstruct the biogeographic history of the genus using the phylogeny and fossil evidence, including the newly discovered *A. hickeyi*, and information about divergence times. Multiple accessions of *A. californica* were included in the study with the hope of offering further insight into the plastid lineage sorting hypothesis of Xiang et al. (1998).

MATERIALS AND METHODS

Amplification and sequencing of molecular markers

DNA for 15 of 19 putative species of *Aesculus* and the outgroup taxa *Billia* Peyr. sp. and *Handeliodendron bodinieri* (Levl) Rehd. was isolated using, variously, CTAB according to Xiang et al. (1998) and DNeasy Plant Mini kit (Qiagen, Inc., Valencia, CA). In some cases, DNA extractions from previous studies were available. All accessions, voucher information, and available material types are shown in Table 1. All 13 species recognized by Hardin (1957a, 1957b, 1960) in his monograph of the Hippocastanaceae were included. Material for two of the six recently described Chinese species was available and included in this study. Multiple accessions of each species were sampled where possible (Table 1).

The *rps*16 intron of the chloroplast genome was amplified by PCR using the primers reported by Oxelman (1997) and following the protocol of Weeks and Simpson (2004).

The chloroplast *mat*K gene was amplified according to the protocol of Modliszewski et al. (2006) but reagent volumes were modified as follows: 2µl template DNA, 1.5µl 10mg/ml BSA, 0.2µl *Taq* polymerase (Promega, Madison, WI.), 26.3µl deionized water.

The *trn*HK region of the plastid genome was isolated and amplified using a procedure modified from Demesure et al. (1995) using the *trn*H and *trn*K primers described by those authors. The following protocol was used for a 50µl reaction: 2µl template DNA, 5µl 10x Mg-free PCR buffer, 6µl $MgCl^2$, 8µl 2.5mM dNTPs, 2µl

10mg/ml BSA 1.5, 0.5µl of each 20mM primer, 0.25µl *Taq* polymerase, and 25.75µl deionized water. Internal sequencing primers for *trn*HK were developed in our lab as described by Modliszewski et al. (2004) (Table 2) and used for sequencing this region. Additional primers were also developed for this study. PCR cycles for amplification of *trn*HK were an initial cycle of 5 min at 94°C followed by 30 cycles of 40 sec at 94°C, 1 min at 65°C, and 2 min 30 sec at 72°C with a final extension of 1 min at 94°C, 1 min at 65°C and 5 min at 72°C.

The nuclear ITS, including the 5.8S gene, was amplified using ITS5 and ITS4 primers following Xiang et al. (1998) with a protocol for a 50µl reaction: 2µl template DNA, 5µl 10x Mg-free PCR buffer, 6µl $MgCl^2$, 8µl 2.5mM dNTPs, 1.5µl 10mg/ml BSA 1.5, 5µl DMSO, 0.5µl of each 20mM primer, 0.3µl *Taq* polymerase, and 21.2µl deionized water. PCR cycles for ITS were as follows: An initial cycle of 6 min at 94°C, 1 min at 48°C and 1min at 72°C followed by 30 cycles of 1 min at 94°C, 1 min at 48°C, and 2 min at 72°C and a final extension of 1 min at 94°C, 1 min at 48°C, and 5 min at 72°C.

A portion of the *LFY* gene, the *LFY* intron 1 was first amplified in this study in *A. californica*, accession #93-1203 from UC Berkeley Botanical Gardens, using degenerate primers developed in our lab from GenBank sequences. From the resulting sequence data from *A. californica*, additional *Aesculus* specific primers were developed. The first series of primers (Table 2) failed to amplify *LFY* in some taxa. For these taxa, series 2 primers successfully amplified *LFY* in all remaining taxa except for *A. parryi*, due to poor quality of DNA. The following PCR reaction protocol was used successfully with all primers: 2µl template DNA, 5µl 10x Mg-free PCR buffer, 6µl $MgCl^2$, 8µl 2.5mM dNTPs, 1.6µl

10mg/ml BSA 1.5, 0.8µl of each 20mM primer, 0.4µl *Taq* polymerase, and 25.4µl deionized water for a 50µl reaction. All primers and references are given in Table 2.

Plastid PCR products were purified following procedures reported in Fan and Xiang (2001) with the following modifications: 500µl of 75% and 95% ethanol, 40µl PEG, resuspended in 8-15µl deionized water. All *LFY* primers resulted in multiple bands, probably due to non-specific binding of primers. Bands of or near the expected length were extracted and purified using QIAquick Gel Extraction Kit (Qiagen, Inc., Valencia, CA). In cases in which more than one potential *LFY* band was extracted, resulting sequences were compared to existing sequences generated for the present study and tested by BLAST against GenBank sequences to determine which band sequence represented *LFY*. All sequencing reactions were carried out on a 3700 capillary DNA Sequencer (Applied Biosystems, Foster City, CA, USA) according to the standard protocol recommended by the company.

Sequences were aligned manually using MacClade 4.02 (Maddison and Maddison 2001) for *rps*16, ITS, and *LFY* and using ClustalX (Thompson et al. 1997) for *mat*K and *trn*HK followed by manual adjustment.

Constructing a morphological character matrix

The morphological data set used in this study (see Table 3 and Appendix 1) is a modification of the 39-character data set published by Forest et al. (2001). The following modifications were made: (1) combined the scoring of *Billia hippocastanum* Peyr. and *B. columbiana* Planchon & Linden into a single taxon, *Billia* sp.; (2) removed *A. glabra* var. *arguta* (Buckl.) Robinson and all outgroup taxa from Sapindaceae except

Handeliodendron from the data set; (3) corrected the leaf margin character state of *A. parryi* from entire to serrate (Manchester, pers. comm. 2007, Hardin 1957b); (4) included 3 additional characters; secondary vein curvature – camptodromic/craspedodromic, intermediate secondary veins – absent or obscure/prominent, and pollen mesocolp structure – striate/spinulose (see Appendix 1) and; (5) scored and included 5 *Aesculus* fossil species.

Fossils of Aesculus *considered in this study*

Fossil taxa of *Aesculus* were researched beginning with literature cross-references and use of data bases including the Plant Fossil Record database (IOP 1997), the Plantae: Fossil Record database (UCMP 1994a ff) and The Paleobiology Database (https://paleodb.org). References available in the data bases were obtained for this study. Fossils reported with questionable or indeterminable affinities to *Aesculus* were excluded from this study. The affinities were determined to be questionable based on descriptions, images, drawings, and discussion with paleobotanists (Erwin pers comm. 2006 at UCMP, Manchester pers. comm. 2007 at UF, Mickle pers comm. 2007 at NCSU). A summary of *Aesculus* fossils determined to be reliable is provided in Table 4. Additional description of these fossils is given in Appendix 2.

The following fossil taxa were considered reliable and were included in this study: *A. hankensii* Prakash & Barghoorn, *A. hickeyi* Manch., *A. indica* foss. (Puri 1945), *A. hippocastanum* foss (Szafer 1947, Szafer 1954, Mai and Walther 1988, and deLumely 1988) *A. longipedunculus* Schloemer-Jager, *A.* 'magnificum' (Budantsev 1983, Manchester 2001), *A. majus* (Nathorst) Tanai, and *A. miochinensis* Hu & Chaney. Fossil

morphology was coded based on descriptions and figures in literature, as well as on discussion with the author of *A. hickeyi* (Manchester, pers. comm 2007). Affinities of these fossils were determined in two ways, using phylogenetic analysis and by posterior mapping of the fossils onto the DNA + morphology (discussed elsewhere in text) tree topology. Posterior mapping was done using synapomorphies and affinities reported by the authors. This method was used when too few informative characters were available for phylogenetic analysis.

Fossils representing modern species, *A. indica* foss. (Puri 1945) and *A. hippocastanum* foss (Szafer 1947, Szafer 1954, Mai and Walther 1988, and deLumely 1988), were used only as constraints in divergence time analysis. *Aesculus hankensii* was placed using posterior mapping as sister to *A. pavia* (Sect. *Pavia*) and was used in biogeographic reconstruction and divergence time dating. This decision as well as discussion of excluded fossil materials is provided in Appendix 3. All other fossils (above) were included in phylogenetic analyses. Some nomenclatural and taxonomic notes on fossil 'Aesculus' taxa are given in Appendix 4.

Phylogenetic analysis

Phylogenetic analyses were conducted using parsimony and the Bayesian methods in PAUP* 4.0 (Swofford 2002) and MrBayes (Huelsenbeck and Ronquist 2001, Ronquist and Huelsenbeck 2003) respectively. All characters were treated as unordered. Gaps in DNA data were treated as missing. Alterative methods of coding gaps and gap rich regions were explored. These included use of simple gap coding (presence/absence)

and ARC (Ambiguous Regions Coding) software (Kauff, Miadlikowska, and Lutzoni 2003, Miadlikowska et al. 2003).

Parsimony analyses in PAUP* 4.0 were carried out using maximum parsimony criterion and the heuristic search option. Characters were unweighted and unordered. All default PAUP* settings were in effect except that random taxon addition of 1000 replicates was employed. For each marker, bootstrap analyses in PAUP* were carried out under the default settings for heuristic searches except that NREPS was set to 1,000, constant characters were excluded, and MAXTREES was set to 1000 per replicate with no increase. INCREASE = NO was invoked to avoid unmanageable file sizes resulting from some preliminary analyses in which bootstrap tree files were ≥ 1gigabyte. Consensus trees from parsimony analyses were generated using the strict consensus option. Trees from bootstrap replicates were compiled using the 50% majority rule option for generating a consensus tree.

For Bayesian analyses, ModelTest 3.0 (Posada & Crandall 1998) was used to determine the best model of evolution for each molecular marker. The following models were selected by the program under the Akaike Information Criterion: TVM + I + G for *mat*K, HKY + I + G for *trn*HK, HKY +G for *LFY*, TRN + I for ITS, and K81uf for *rps*16. Models were implemented in MrBayes v.3.1.2 using the PRSET and LSET commands. Two simultaneous, independent Markov chains were run for 22 million generations for each data set to check convergence. Three hot chains and a single cold chain were employed for each run to ensure mixing. Trees were sampled every 2000 generations. Burnin was set to 2.2 million generations or 1100 trees, but was observed using Tracer v.1.3 (Rambaut and Drummond 2005) to ensure adequate burnin. The stop rule, a rule

that ends the analysis when the average standard deviation of partition frequencies of all simultaneous runs reaches a user specified value, was set to 0.001 and was in effect for all data sets but was never implemented by the program. All runs completed the specified 22 million generations. The 19,800 remaining trees from 2 simultaneous, independent runs were summarized using a 50% majority rule consensus in PAUP*4.0.

As an intuitive way of evaluating conflict between results of parsimony and Bayesian analyses for each data set, we used a clade support criterion of ≥ 70% bootstrap support for parsimony trees and ≥ 90% Bayesian posterior probability (PP) for Bayesian consensus trees. Conflict was considered to exist between parsimony and Bayesian trees if this support of this level or greater was observed for conflicting nodes on the Bayesian and MP tree topologies. Using this criterion, no conflict was found between Bayesian and MP analysis for any data set. The same criterion was used to evaluate conflict between data sets, and no conflict was observed. This method of evaluating conflict ignores topological incongruence that is not supported by values of or greater than the conflict criterion.

Analyses of combined data sets were carried out using the same methods described above except where noted. The following combined data were analyzed: (1) combined cpDNA data (*mat*K, *trn*HK, and *rps*l6); (2) combined nuclear DNA data (ITS and *LFY*) (Bayesian analysis only with bootstrapping); (3) combined DNA data; (4) combined DNA data with gaps and gap rich regions excluded (Bayesian only); (5) combined DNA + morphology, excluding fossils; and (6) combined DNA + morphology with fossils included and DNA data coded as missing for fossils (Bayesian only). Accessions with DNA data from at least two regions were included in the combined

molecular analyses. For combined data including morphology, a representative individual of each species in the molecular matrix was selected for inclusion in the combined matrix.

A further MP analysis of the morphology matrix including fossils was carried out in PAUP* using the DNA + morphology consensus tree as a backbone topological constraint to place the fossils on the phylogenetic tree. This was performed due to the lack of resolution of fossil species from phylogenetic analyses of morphological data alone and combined analysis of DNA + morphology including fossils with no constraint. In this analysis with a backbone constraint, lack of informative characters and missing data in the matrix including fossils resulted in large number of equally parsimonious trees in each heuristic search and, thus, unmanageable file sizes even when a backbone constraint was employed. As an alternative, five heuristic searches were preformed using 1000 random taxon addition replicates and the commands MAXTREES=10,000 and INCREASE=NO invoked. A 50% majority rule consensus tree of these 50,000 trees was constructed to aid in determination of the most likely placement of fossils on the tree. Bootstrapping was done using the method described above for individual matrices.

Divergence time dating

Divergence time dating was performed using Multidivtime (Thorne 2003). In this analysis, we used the total evidence tree based on DNA + morphology of extant species. Node age constraints were set using fossil dates. Five age constraints were set. The fossil affinities were determined from: (1) the majority rule consensus of the 50,000 MP trees resulting from the analysis of the morphology matrix including fossils; (2) the MP

analysis of morphological data using the Bayesian DNA + morphology tree as the backbone constraint; and (3) evidence from paleobotanical literature (see Table 4). Node age was assumed to be at least (≥) as old as the oldest fossil diverging from that node. In other words, the age of fossils represent the minimum age of the stem lineage giving rise to the crown group consisting of the fossil species and its sister clade.

The Paleocene fossil, *A.* 'magnificum' (Budantsev 1983, Manchester 2001), representing the Asian clade (≥58 MYA), was used to establish the minimum age of divergence between the Asian clade and its sister clade. *Aesculus* longipedunculus (from the lower Eocene), the oldest fossil representing the clade including Sect. *Aesculus*, Sect. *Pavia*, and Sect. *Parryana* was used as the minimum age for the divergence of this clade and its sister. The sister of this clade and of the Asian clade were not resolved due to a deep node trichotomy above the root of *Aesculus* including the Asian clade (Fig. 2), Sect. *Aesculus* + Sect. *Pavia* + Sect. *Parryana*, and Sect. *Macrothyrsus* + *A. californica*. Therefore, the older of these two fossils, *A.* 'magnificum' (Budantsev 1983, Manchester 2001) (Paleocene ~58MYA) and *A. longipedunculus* (Early Eocene ~54.8MYA), was used to constrain the root of *Aesculus* (Fig. 2). The Paleocene fossil *A. hickeyi*, placed as sister to Sect. *Aesculus* in phylogenetic analysis, was used to set the minimum age for divergence of Sect. *Aesculus* and its sister, the clade containing Sect. *Pavia* + Sect. *Parryana*. This age constraint is set at the lower boundary of the Paleocene (≥ 54.8MYA), slightly younger than the age constraining the root of *Aesculus*. The Upper Miocene fossil, *A. hankensii* constrains the divergence of Sect. *Pavia* from Sect. *Parryana* to a time ≥ 5.3MYA Based on the presence of *A. indica* fossils in the Pliocene, *A. indica* must have diverged from *A. polyneura* at least 1.8 MY*A.* Pliocene fossil

remains of the extant species *A. hippocastanum* allowed for age constraint of ≥1.8MYA on the node defining that species divergence from *A. turbinata*. Figure 2 shows all of the age constraints on nodes.

The prior root age of *Aesculus* was set at 58 million years, or at the Paleocene-Eocene boundary. The "bigtime" value, or maximum root age, was set at 150MY, a Late Jurassic date debatably consistent with the earliest known angiosperm fossil, *Archaefructus liaoningensis* (Soltis et al. 2005) and predating the divergence core-eudicots (Chaw et al. 2004) to which Aesculus belongs. The branch length prior (rtrate) was set using an average of the branch lengths calculated for each molecular marker in PAML. This value was compared to branch length calculated from a random sample of 10 Bayesian trees from the combined analysis of all molecular markers excluding gap coding. These values differed only beyond 3 significant digits right of the decimal, and, thus, we considered the PAML estimate to be reasonable. Mean branch length was divided by 65MY, the upper bound of the Paleocene, resulting 0.0015264 substitutions per site per million years. The rate standard deviation was set as equal to the mean branch length at 0.0015264. The time standard deviation was determined using data from a preliminary run in which the bigtime value was set at 250MYA. The largest posterior standard deviation from that run, 12.8, was used as a prior in subsequent runs. The Markov chain was set to run with a burnin of 2,200,000 generations, sampling every 200 generations thereafter with 99,000 total samples. This equates to a total run of 22,000,000 generations. An initial run using these settings but without sequence data was performed to determine the contribution of the data to the resulting estimations and to determine

whether our priors were reasonable. The analysis with sequence data was run twice to ensure that resulting values did not differ substantially.

Biogeographic reconstruction

Biogeographic reconstruction was done using the Bayesian trees and resulting consensus and a probabilistic approach to applying DIVA (Ronquist 1999). In this method, the probability of each of the three possible arrangements of the three major lineages of *Aesculus* from the consensus topology, including fossils (Fig. 3) placed using phylogeny and manual placement, were determined. Since the 50% majority rule Bayesian consensus tree shows a trichotomy among the three major lineages (Fig. 3), the three possible alternative arrangements of the three linages were examined. The frequencies (or probabilities) of the three alternative tree topologies were considered in the calculation of the final probability of ancestral areas for each node based on results from all three trees. The arrangement with the highest probability was calculated by constructing a majority rule tree using PAUP* and allowing compatible arrangements with probability less than 50%. Section *Macrothyrsus* + *A. californica* and Sect. *Parryana* + Sect. *Pavia* + Sect. *Aesculus* were resolved as sisters, supported by 48% PP. The probability of the other two arrangements were estimated as a function of their frequency of their occurrence observed in a random sample of 100 Bayesian trees. The topology showing Sect. *Calothyrsus* and Sect. *Parryana* + Sect. *Pavia* + Sect. *Aesculus* as sisters was observed in 32% of the topologies. The topology showing Sect. *Calothyrsus* as sister to Sect. *Macrothyrsus* + *A. californica* was observed in 12% of the sampled trees. These relationships are discussed in greater detail below (see Results).

Ancestral areas for all three topologies described above were optimized in DIVA using maxareas=2 and the printrecs function. Resulting optimizations were used in a probability calculation for all nodes, including nodes for which the sister group remains unspecified (*x*, see Chapter 2) in the 50% majority rule consensus topology; i.e. the sister of each of the three major lineages. The probability of an ancestral distribution in the set of optimal scenarios optimized for each node by DIVA was calculated as $Y_A = P_{C1}(R/O) + P_{C2}(R/O) + P_{C3}(R/O)$; where Y_A is the probability of ancestral area A at node Y, P_{Cn} is the probability of each of the three alternative topologies, O is the total number of optimal scenarios for each tree topology, and R is the number of times area A occurs in O.

Taxa were scored as being present or absent in six areas: Europe (Eu), East Asia (EA), eastern North America (eNA), western North America (wNA), the Greenland region including other areas of the Thulean and deGreer Landbridges (Gr), and Central and South America (CSA). The EA fossil species *A.* 'magnificum' (Budantsev 1983, Manchester 2001), *A. magus*, and *A. miochinensis*, were treated as a group representing Asian clade fossils in DIVA analysis (Fig. 3, see also the discussion of these fossils in Appendix 2). The fossil taxa *A. longipedunculus* (Gr), *A. hankensii* (eNA), and *A. hickeyi* (wNA) were also included in the biogeographic reconstruction.

Ronquist (1999) noted that at deep nodes, DIVA tends to predict widespread ancestors, since nearer to the root of the tree, optimizations rely heavily on the distributions of outgroups, i.e., connection of the tree to the rest of the tree of life. This effect was minimized in our method by using the maxareas=2 option. To further increase the accuracy of DIVA optimizations, especially for the ancestral area of last shared

ancestor of *Aesculus* and *Billia*, we added Aceraceae as a further outgroup to the root of all three trees optimized in DIVA (see Thesis Introduction). When using a taxon higher than the species level in DIVA, it is more accurate to use the ancestral area of the root of that taxon rather than using the distribution areas of all constituent species. The root of Aceraceae was determined to be wNA and EA according to Boulter et al. (1996). These authors showed that Aceraceae evolved in the high latitudes of the pacific on land areas that became part of the Bering Land Bridge connecting EA and wNA.

RESULTS

Sequence analysis

Amplified portions of the *mat*K, *rps*16, and ITS regions were similar in length for all species (average: 1812bp, 812bp, and 597bp respectively). For *trn*HK, the average sequence length for all taxa and individuals, excluding *Handeliodendron* and *A. tsiangii*, was ~1727bp. Large deletions, virtually identical in location, were observed in *Handeliodendron* and *A. tsiangii* of ~437bp and ~434bp respectively, or ~445bp measured over both taxa. Deletion length was measured with gaps open in the sequence alignment of other samples. Two adjacent microsatillites were observed in the amplified portion of the *LFY* intron 1 (Table 5). An AT repeat ranged from ~18 repeated pairs in *Billia* 1 to virtually absent in *A. parviflora* 2. Compared to the AT repeat, the adjacent CAT repeat was more interrupted by point mutations in all taxa in which it was present. The CAT was absent in *Billia* 1. It was the longest in all *A. californica* individuals; ~10

units. In *A. tsiangii*, a ~459bp repeat was observed following a 16bp 3' flanking region of the adjacent microsatillites.

The matrix combining all chloroplast and nuclear DNA was 6489 characters and consisted of 30 individuals representing 16 *Aesculus* species, *Billia*, and *Handeliodendron*. This matrix included all individuals for which at least DNA sequences from two markers are available although some missing data are present at the 5' or 3' ends of some sequences. The matrix of total molecular data combined with the 42-character morphological matrix, containing 16 species of *Aesculus* as well as *Billia* and *Handeliodendron*, selected representative individuals from each species for which all or nearly all molecular markers had been sequenced. This matrix included 6447 characters. Twelve of 42 characters in the morphology matrix were constant or uninformative. Additional sequence information including average number of bases and % parsimony informative characters for each gene region is given in Table 6.

Phylogenetic analysis

Analyses of individual data sets: Parsimony and Bayesian trees resulting from analyses of each data set were not in conflict, using the criteria described above. Using the same criterion, conflict was not observed between data sets. Incongruent topologies were observed but these received posterior probability values and bootstrap support below the 70% bootstrap and 90% PP criterion.

Analysis of morphological data resulted in 30 MP trees with a strict consensus topology differing from the one reported by Forest et al. (2001) (Fig. 1d and Fig. 4). We found a topology showing Sect. *Macrothyrsus* was resolved as sister to the rest of

Aesculus, *A. californica* was placed as sister to the Asian clade, and Sect. *Parryana*, *A. glabra* , (*A. Pavia* + *A. sylvatica* + *A. flava*) and Sect. *Aesculus*, were resolved as a polytomy clade, sister to the Asian clade + *A. californica*. Only the monophyly of Sect. *Aesculus* and the clade including *A. pavia*, *A. flava*, *A. sylvatica* received bootstrap support ≥ 70% (Fig. 4).

Combined cpDNA (*trn*HK, *rsp*16, and *mat*K): These analyses consisted of 30 individuals with all species of *Aesculus* (included in this study) represented by at least one individual and with both outgroup taxa represented. Results of combined analysis of cpDNA data strongly support the monophyly of *Aesculus* (99% bootstrap,100% PP), the monophyly of all *A. californica* individuals (100% bootstrap, 100% PP), the monophyly of Sect. *Pavia* (100% bootstrap), the monophyly of an Asian clade (93% bootstrap, 100% PP), and for the monophyly of Sect. *Aesculus* (100% bootstrap, 100% PP). A previously reported sister relationship between Sect. *Pavia* and Sect. *Parryana* (Xiang et al. 1998) was highly supported by these data (96% bootstrap, 100% PP) (Fig. 5). The Bayesian consensus tree has a bifurcating *Aesculus* root node containing the Asian clade + *A. californica* (traditional Sect. *Calothyrsus*) and a clade of Sect. *Parryana* + Sect. *Aesculus* + Sect. *Pavia* + Sect. *Macrothyrsus*. The traditional Sect. *Calothyrsus*, including *A. californica*, was supported by 85% PP, with the *A. californica* lineage being the earliest diverging lineage within the group (93% bootstrap, 100% PP). A sister relationship between Sect. *Pavia* + Sect. *Parryana* and Sect. *Aesculus* was supported with 99% PP. Section *Macrothyrsus* was suggested to be sister to Sect. *Aesculus*, Sect. *Pavia*, and Sect. *Parryana* with moderate support (83% PP). The MP tree topology (not shown) was not in conflict with these results, but poorly resolved the ancestral node of *Aesculus*, showing a

polytomy including 5 lineages: Sect. *Aesculus*, Sect. *Pavia* + Sect. *Parryana*, Sect. *Macrothyrsus*, *A. californica*, and the Asian clade.

Combined nuclear DNA (ITS and *LFY*): Results of analyses of combined nuclear DNA data are highly congruent with those found in cpDNA analyses except that *A. californica* was resolved as sister to Sect. *Macrothyrsus* with moderate support (84% PP, 65% bootstrap) (Fig. 6) rather than in Sect. *Calothyrsus* as shown in the cpDNA tree (Fig. 5). However, most relationships revealed in the analysis of ITS and *LFY* data are more weakly supported except for the monophyly of Sect. *Aesculus* (97% PP) and Sect. *Pavia* (98% PP). Weak support (68% PP and no bootstrap support ≥ 50%) was shown for the monophyly of *Aesculus*. The Asian clade was supported with 89% PP. A sister relationship between Sect. *Parryana* + Sect. *Pavia* and Sect. *Aesculus* was supported with 49% PP, while a sister relationship between Sect. *Parryana* + Sect. *Pavia* + Sect. *Aesculus* and *A. californica* + Sect. *Macrothyrsus* was supported by only 28% of the 19,800 sampled Bayesian trees.

Combined DNA (cpDNA and nuclear DNA): These analyses included 30 taxa, including at least one individual of each species sampled and both outgroup taxa. Parsimony and Bayesian methods of analysis did not result in conflicting topologies. The combined DNA tree topology was similar to the nuclear DNA tree topology, except that *A. californica* was not resolved as sister to Sect. *Macrothyrsus*, but as sister to Sect. *Parryana*, Sect. *Pavia* + Sect. *Aesculus* (Fig. 7) with low support (≤ 50% bootstrap, 84%PP). Another difference was that Sect. *Macrothyrsus* was resolved as sister to the rest of *Aesculus* with weak support (54%PP). In general, relationships between sections were weakly supported except for the sister relationship between Sect. *Pavia* and Sect.

Parryana. Analysis of the total DNA data also reveals strong support for the monophyly of *Aesculus* (100% bootstrap, 100% PP) (Fig. 7).

The Bayesian analysis of the combined DNA, excluding gaps and gap rich regions, resulted in a tree topology (not shown) with poor resolution (≤ 54% PP) of relationships between traditional sections, the *A. californica* lineage, and the Asian clade except that a relationship between Sect. *Parryana* and Sect. *Pavia* was supported in 100% of the sampled Bayesian trees.

Combined DNA + morphology, excluding fossils: The matrix analyzed here included 1 representative individual from all 16 of the 19 extant *Aesculus* species sampled in this study as well as sequence and morphological data for *Billia* sp. and *Handeliodendron*. Analysis of these data strongly support the monophyly of *Aesculus* (100% bootstrap, 100% PP) (Fig. 8a,b) and resolves *Aesculus* as a clade of three major constituent lineages; the Asian clade (100% bootstrap, 100% PP), Sect. *Macrothyrsus* + *A. californica* (60% bootstrap, 96% PP), and Sect. *Parryana* + Sect. *Pavia* + Sect. *Aesculus* (92% PP). Section *Pavia* was resolved as sister to Sect. *Parryana* (70% bootstrap, 100% PP).

Phylogenetic analyses including fossils and fossil placement: Only 2 of the 5 fossils included in this study appear to have sufficient characters, informative and scorable from literature, to be resolved with any confidence in the phylogenetic analysis; *A. longipedunculus* and *A. hickeyi*. The analysis of the morphology matrix, including fossils, never accomplished more than a single taxon addition replicate before the MAXTREE limit was reached. A brief examination of the 10,000 stored trees from each run seems to suggest that 5 different islands were explored in each of the five analyses. It

is reasonable to assume that additional runs would have indicated additional islands. Furthermore, by limiting the number of runs to 5, tree space was only minimally explored. A majority rule consensus of the of the 50,000 MP shows an Asian clade including *Aesculus* 'magnificum' Budantsev, *A. miochinensis* and *A. majus*, was supported by 35% of the 50,000 trees (Fig. 9). This is significantly different (when $\alpha = 0.05$) from the expected value of 20% if we assumed that each run explored a different island and that topological variations within each island were minimal due to the specified maximum number of trees. Also statistically significant was the placement of *A. hickeyi* in a monophyletic group with Sect. *Aesculus* (91%). *Aesculus longipedunculus* occurred 87% of the time in a group including Sect. *Pavia*, Sect. *Parryana*, Sec. *Aesculus*, and *A. hickeyi*. Within that group, 89% of the data supports a monophyletic relationship between Sect. *Pavia*, Sect. *Parryana*, Sect. *Aesculus*, and *A. hickeyi*, suggesting that *A. longipedunculus* is sister to the rest of the group. Bootstrapping supported the placement of *A. longipedunculus* (32%) and *A.* hickeyi (42%) described above. Other fossils received no bootstrap support > 9% (Fig. 9).

The Bayesian analysis of DNA + morphology including fossils (not shown, PP support shown in Fig. 9) yielded 39% support for the monophyly of *Aesculus*. As in the MP analysis, a clade including *A. longipedunculus*, Sect. *Pavia*, Sect. *Parryana*, and (Sect. *Aesculus* + *A. hickeyi*) was reconstructed with a posterior probability of 13%. Within this group, *A. longipedunculus* was resolved as sister to the remaining taxa (24%). The Asian clade was resolved with 10% PP with *Aesculus majus* sister to the rest of the clade (11%). *Aesculus miochinensis* was placed as sister to the rest of *Aesculus* with 12% PP.

Divergence time dating

Reasonable values for actual divergence times of nodes fell within the posterior 95% CIs when Multidivtime was run without sequence data. Values were considered reasonable based on the prior work of Xiang et al. (1998), and other prior information from the *Aesculus* fossil record and the hypothesized divergence time of angiosperms (see Soltis et al. 2005). There was little variation observed between the posterior distribution of divergence times and 95% CIs and those resulting from the prior run with no sequence data. This indicates a heavy reliance of our Multidivtime estimations on our priors. This can probably be explained by low sequence divergence in the molecular markers selected for this study (see Table 6).

Results of the analysis using Multidivtime are summarized in Fig. 10, showing estimated divergence times and 95% CIs on the tree topology. The age of the root of the crown groupwas estimated to be Cretaceous to mid-Paleocene, ~62.36MY. Diversification of the Asian clade began in the Eocene or ~45.09MYA. Of the extant lineages, *A. assamica* was the first to diverge, followed by the *A. indica* + *A. polyneura* lineage at ~38.46MYA, also in the Eocene. *Aesculus polyneura* was shown to have diverged from *A. indica* in the Early Oligocene or ~31.96MYA. The most recent speciation event within the Asian clade occurred ~15.87MYA in the Miocene when *A. tsiangii* diverged from *A. chinensis*. The divergence between Sect. *Aesculus* and its sister consisting of Sect. *Pavia* and Sect. *Parryana* was estimated to be in the Paleocene; ~61.07MYA. *Aesculus hippocastanum* and *A. turbinate* of Sect. *Aesculus* diverged from a common ancestor in the Oligocene or ~37.06MYA. Section *Parryana*, represented by

A. parryi from western North America, diverged from the polytypic eastern North American Sect. *Pavia* in the Eocene or ~49.04MYA. Divergence of species within Sect. *Pavia* began at ~31.31MYA, in the Early Oligocene, when *A. flava* diverged from the stem lineage. The most recent speciation event within Sect. *Pavia*, the divergence of the *A. sylvatica* and *A. glabra*, occurred ~17.50MYA in the Miocene. The monotypic Sect. *Macrothyrsus* was shown to have diverged from its sister, *A. californica*, in the Paleocene, ~61.15MYA.

Biogeographic reconstruction

The ancestral distributions at each node of *Aesculus* with the highest probabilities estimated with DIVA are shown in Figure 11. The ancestor of *Aesculus* and *Billia* are suggested to have a wide spread range including East Asia and Central and South America prior to the divergence of *Aesculus* (BF, P=0.80) (Fig. 11). The last shared ancestor of all *Aesculus* was distributed in of East Asia and western North America (P=0.63). The Asian clade and an unspecified sister (*x*), perhaps the rest of *Aesculus* (supported in 48% of the data, see above in Materials and Methods), had an East Asian-western North American range. The last shared ancestor of this polytypic group had an East Asian distribution. All further diversification in this group was within East Asia (BD, P=1.00).

The ancestor of the large clade including Sect. *Aesculus*, Sect. *Parryana*, and Sect. *Pavia*, as well as the fossil taxa *A. longipedunculus*, *A. hickeyi*, and *A. hankensii*, had western Noorth American range (D, P=0.51) before diverging from its sister (*A. californica* + Sect. *Macrothyrsus* or the Asian clade). *Aesculus longipedunculus* diverged

first within this group from an ancestor found in western North America and in the Greenland region (DE, P=0.73). The Greenland region lineage, represented by the fossil taxon *A. longipedunculus*, became extinct. Section *Aesculus*, including *A. hickeyi*, diverged from a shared ancestor with the Sect. *Parryana* + Sect. *Pavia* lineage in western North America (D, P=0.81). Section *Pavia* had an eastern North American ancestor prior to the diversification of that group. The presence of *A. hankensii* in western North America is explained by dispersal and subsequent isolation from the eastern North American ancestor shared with *A. pavia* (C, P=1.0). Section *Aesculus* spread to eastern Asia before the divergence of *A. hickeyi* from the rest of the lineage (BD, P=0.52). The presence of *A. hippocastanum* in Europe is explained by dispersal into Europe (A on Fig. 11), followed by a vicariance event, which isolated the *A. hippocastanum* and *A. turbinata* lineages.

The last shared ancestor of the lineage Sect. *Macrothyrsus* + *A. californica* and its unspecified sister was distributed in western North America (D, P=0.56). This ancestor is most likely (48%, see above, Materials and Methods) to have been shared with the clade including Sect. *Aesculus*, Sect. *Parryana*, and Sect. *Pavia*, as well as the fossil taxa *A. longipedunculus*, and *A. hickeyi*, which was shown to have the same ancestral range (wNA). Before divergence of the Sect. *Macrothyrsus* and *A. californica* lineages, a shared ancestor had spread to eastern North America (CD, P=1.00). The range of this ancestor was disrupted, resulting in isolation and speciation and explaining the disjunction of *A. californica* in western North America and Sect. *Macrothyrsus* in eastern North America.

Other areas suggested for each node all had probabilities significantly lower than the areas described above. An exception was found for the range of the ancestor of Sect. *Aesculus* and *A. hickeyi*. The ancestral distribution with the highest probability is eastern North America-western North America (BD, P=0.515). An ancestral distribution in Europe and western North America is almost equally likely (AD, 0.405). However, lack of a direct terrestrial connection between western North America and Europe may allow us to put less weight on this alternative scenario.

DISCUSSION

The phylogeny of Aesculus

The relationships resolved by the total evidence and nuclear DNA data were mostly congruent with those found in Xiang et al (1998), and some are congruent with those in Forest et al. (2001) (Figs.1c, 1d, 8). The new finding of the present study is the sister relationship between Sect. *Aesculus* and Sect. *Parryana* + Sect. *Pavia*, which has not been recognized in any previous studies. This relationship was resolved by all data partitions (Figs. 6-9) and by total evidence (92% PP).

Despite the increase of data and taxa in the present study, relationships between the three major lineages (Sect. Pavia + Sect. Parryana) + Sect. *Aesculus*, Sect. *Macrothyrsus* + *A. californica, and* Sect. *Calothyrsus* except *A. californica* remained unconvincingly solved (Fig. 8). This suggests a likely rapid radiation of *Aesculus* in the early evolutionary history of the genus. This phenomenon results in short branches at base of the phylogeny, which, in this case, cannot be resolved with the data available.

Alternatively, the lack of resolution among the three major lineages may be attributed to potentially conflicting phylogenetic signals in the cpDNA and nuclear DNA data, particularly with respect to the placement of *A. californica* and Sect. *Macrothyrsus*. Conflicting relationships were resolved for these taxa with moderate support (Fig. 5, 6).

Plastid lineage sorting and the placement of A. *californica*

Our results agree with the finding of a conflicting relationship of *A. californica* observed in the cpDNA and nuclear DNA trees by Xiang et al. (1998) (Figs. 1c, 1d, 5, 6), although this conflict received support lower than our conflict criterion (≥ 70% bootstrap, ≥ 90% PP). In the combined cpDNA data analysis (Fig. 5), *Aesculus californica* was united with the Asian species of Sect. *Calothyrsus* (85% PP). These results are further supported by the individual analyses of the *mat*K and *trn*HK plastid markers (not shown), in which *A. californica* was shown to share a more recent common ancestor with the traditional Sect. *Calothyrsus* than with other species of *Aesculus*. In the nuclear DNA tree topology (Fig. 6), *A. californica* was united with *A. parviflora*, the single species of Sect. *Macrothyrsus* from southeastern US, and this group is sister to the clade consisting of Sect. *Parryana*, Sect. *Pavia*, and Sect. *Aesculus*. The sister group relationship between *A. californica* and Sect. *Macrothyrsus* was also strongly supported further by the combined total DNA + morphological data (Fig. 8; 96% PP). This incongruence can be explained by the lineage sorting hypothesis proposed by Xiang et al. (1998); i.e., by an ancestral polymorphism and subsequent differential loss of cpDNA haplotypes in different lineages. Chloroplast DNA polymorphisms have been found elsewhere in *Aesculus*; reported by Modliszeski et al. (2006) in Sect. *Pavia*. Due to the uncertain placement of

the *A. californica* + Sect. *Macrothyrsus* clade within *Aesculus* (Figs. 5, 6, 7, 8), tracing the history of lineage sorting of the polymorphic ancestral cpDNA across the tree topology is ambiguous. It is likely that the polymorphism dated back to the ancestor of the genus, at least as early as the Paleocene (~62.36MYA), and was present in *Aesculus* at least as late at the divergence of Sect. *Macrothyrsus* from A. californica (~61.15MYA) in the mid-Paleocene. The ancestor of the genus possessed two types of cpDNA (AB) (Fig. 12). An ancestor of the Asian species of Sect. *Calothyrsus* and *A. californica* lost the B type, while Sect. *Macrothyrsus* (*A. parviflora*) and the ancestor of (Sect. *Pavia* + Sect. *Parryana*) + Sect. *Aesculus* lost the A type. This scenario, based on the present cpDNA and nuclear phylogeny (Figs. 5, 6, 12), is similar to that proposed in Xiang et al. (1998) except we show that Sect. *Aesculus* has the B haplotype (Fig. 12).

Evolutionary and biogeographic history of Aesculus

The estimated divergence time for the root of *Aesculus* (~62.36MYA) indicates a Paleocene origin of *Aesculus* that is similar to estimated divergence time based on the strict ITS clock (Xiang et al. 1998). During this time, the latest Cretaceous and early Paleocene, the floristic interchange across the Northern Hemisphere appears to have been more limited than in the earlier Cretaceous (Budantsev 1992). Two floras, one covering much of EA and wNA and the other covering much of Europe and eNA persisted in relative isolation from one another (Budantsev 1992). These floras were later united across the Northern Hemisphere due loss of geographic barriers including the Turgai Strait (isolating Europe from Asia through the early Oligocene) and the inter-continental seaway (persisting until the Paleocene) (Budantsev 1992, Tiffney and Manchester 2001).

Extant *Aesculus* species exhibit characters more commonly associated with mesophytic species, such as distinct growth rings and deciduousness in some species, while *Handeliodendron* and *Billia* exhibit characters more commonly associated with thermophytic species, including indistinct or lacking growth rings (*Handeliodendron* and *Billia*) (Kamer 1939, Klaassen et al. 1995), an evergreen habit (Record and Hess 1943, Forest et al. 2001), and fibrous sheathing of leaf veins (*Billia* and unreported in *Handeliodendron*) (Hardin 1957a). While the thermophytic characters in *Billia* and *Handeliodendron* could have evolved later in those lineages, we propose that *Aesculus* diverged from *Handeliodendron* and *Billia*, with the two related taxa being part of a Late Cretaceous thermophilic flora (see: Budantsev 1992). Results from DIVA show that the ancestor of *Aesculus* and *Billia* was widespread in CSA and EA. It is probable that the ancestor of *Aesculus* and *Billia* evolved in EA and migrated east and southward from EA to CSA. Extinction in the northern hemisphere resulted in the geographic isolation and the divergence of *Aesculus* and *Billia*. Presence of an *Aesculus-Billia* ancestor in EA is consistent with the presence of the extant Handeliodendron lineage in that region, an ancestral distribution of Aceraceae in what later became the Bering Land Bridge (EA-wNA) (Boulter 1996), and with the presence of one of the oldest *Aesculus* fossils in this region, the Beringian (northeastern Russia, EA) fossil, *A.* 'magnificum' (Budantsev 1992, Manchester 2001).

The last shared ancestor of *Aesculus* was inferred to occupy a range includin EA and wNA. Based on the occurance of early fossils of *Aesculus* in EA and wNA (*A.* 'magnificum' from Kamchatka in northeastern Russia and *A. hickeyi* from wNA), the ancestral distribution of *Aesculus* was likely spanning across the Beringian Land Bridge.

Presence of the Paleocene fossil, *A. hickeyi* in wNA in an area outside of Beringia, may indicate that the distribution of the earliest *Aesculus* taxon was not limited in EA and wNA to Beringia. The deep node polytomy in the phylogeny of *Aesculus* supporting an early rapid radiation of the genus is consistent with a pattern of floristic radiation from the Beringian region discussed by Budantsev (1992). This author showed that disappearance of geographic barriers (e.g., the Turgai Strait in Eurasia and the epicontinental seaway in North America) allowed a Beringian temperate flora to spread and diversify rapidly across northern Laurasia.

The data indicate that from its ancestral distribution the genus became widespread across Laurasia by at least the Eocene. Fossil evidence shows that, by at least the Early Eocene (~54.8 MYA), *Aesculus* was present in EA, wNA, and Gr (Fig. 10, 11), specifically in Spitsbergen (Schlomner-Jager 1904, Golovneva 2000). Spitsbergen, though politically associated with Europe, is a Scandinavian island and was accessible to boreotropical and mesophytic floristic elements from North America via Greenland and the DeGeer land route and from northern Asia via the Artic Ocean coast during the Paleocene (Golovneva 2000). Golovneva (2000) infers from floristic connections between West Greenland and the British Isles that the Thulean Land Bridge (part of the narrowing North Atlantic Land Bridge), linking Spitsbergen to mainland Europe via the western coast of Greenland, became an important migration route for the boreotropical and early mesophytic floristic elements in the early Eocene. In *Aesculus*, our data from DIVA and divergence time analysis a spread of the ancestor of the clade Sect. *Aesculus* (including *A. hickeyi*)+ Sect. *Pavia* + Sect. *Parryana* + *A. longipedunculus* from wNA to Gr (Fig. 11) in the Paleocene (~62.36-61.07, Fig. 10). During this time, migration from

wNA to Gr via the North Atlantic Land Bridge or the Beringian Land Bridge was possible. The arrival of *Aesculus* on the European mainland, represented by the modern species *A. hippocastanum*, was inferred as the result of dispersal from EA sometime between the Paleocene (~61.07MYA) and the Oligocene (~37.06MYA) (Fig. 10, 11). This dispersal probably occurred after the Turgai Strait retreated in the early Oligocene (Tiffney and Manchester 2001). This period is characterized by the invasion of Europe by Asian floristic elements (Tiffney and Manchester 2001).The presence of *Aesculus* in eNA (*A. parviflora* and Sect. *Pavia*) was shown to rbe the result of a dispersal from wNA between the Paleocene and the Eocene (~61.07-49.04MYA, Fig. 10, 11). This was possible because the epicontinental seaway dividing North America began its retreat in the Paleocene (Tiffney and Manchester 2001). During the early to mid-Eocene, the central United States was dominated by temperate rainforest, including evergreen and deciduous dicots (Rosen 1978). As dry conditions developing from the rise of the Rocky Mountains in the central North America since the late Eocene, these forests shifted to the East (Rosen 1978). Imposition of the Rocky Mountain rainshadow and other changes resulting from Rocky Mountain orthography isolated the Section *Parryana* (wNA) and Section *Pavia* (eNA) lineages. The extant Section *Parryana* species, *A. parryi*, exhibits xeric characters (Forest et al. 2001), probably developed early in that lineage allowing for its survival in the increasingly dry western United States. A back dispersal from eNA to wNA was suggested within Sect. *Pavia* during the Miocene (Fig. 10, 11). This probably occurred via islands of forests or through a high latitude corridor (Tiffney and Manchester 2001). Section *Macrothyrsus* became isolated from the wNA A. californica during the Paleocene (~61.15MYA, Fig. 10), probably as the result of isolation due to

local climatic changes associated with the retreat of the epicontinental seaway (Budantsev 1992, Tiffney and Manchester 2001). This isolation was followed by differential habitat adaptations; xeric adaptations in *A. californica* (Forest et al. 2001) in the dryer west and mesic adaptations of *A. parviflora* in the wetter east.

The ancestral area of the Asian clade was inferred to be EA with high probability (B, P=1.0) (Fig. 11). As the climate cooled in the high latitudes of East Asia (Budantsev 1992), the East Asian *Aesculus* lineage retreated south and west to warmer areas. The core lineage diverged, with one lineage surviving in southwest China and the Himalayas (*A. assamica*, *A. indica*, and *A. polyneura*)and one surviving in southern and central China (*A. tsiangii*, *A. chinensis*, *A. wilsonii*, and *A. wangii*). The southwestern lineage, *A. assamica* and *A. indica* + *A. polyneura*, diversified further as dryer and cooler conditions imposed barriers across central and western East Asia. The southern/central lineage (*A. chinensis*, *A. tsiangii*, *A. wilsonii*, and *A. wangii*) diversified further beginning in ~ 32.15MYA (Early Oligocene). Although exact placement of *A. majus* within the Asian clade remains uncertain, this taxon represents the known eastern limit of the clade (additional information in Appendix 2). The *A. majus* lineage probably emerged from within the southern/central lineage and moved eastward into what is now Japan. *Aesculus miochinensis*, known from Shatung Province in China, may also represent an extinct lineage derived from within the southern/central lineage.

The biogeographic history of *Aesculus* inferred from the present study including fossils was generally congruent with previous hypotheses (Xiang et al. 1998, Forest et al. 2001), which postulated that *Aesculus* arose during the transition from Cretaceous to Tertiary (~ 64 MYA) as part of the boreotropical flora (Xiang et al. 1998, Golovneva

2000, Forest et al. 2001) described by Wolfe (1975). Our results similary suggest a high latitude origin of Aesculus spanning the Bering Land Bridge and including EA and wNA. In the late Cretaceous, during which angiosperms are reported to have radiated (Wolfe 1975, Crane et al. 1995), the floras of wNA and EA were largely isolated from those of eNA and Europe. During the early Tertiary these floras merged across the Northern Hemisphere by way of the Bering and North Atlantic Landbridges (see: Wen 1999), the arctic coastline, other hypothesized land routes (Golovneva 2000, Tiffney and Manchester 2001), and the Tethys Seaway (Tiffney 1985). This flora covered a land area as southerly as 50°N (Wolfe 1975) in some localities during the Eocene. Later, when climatic cooling occurred, the boreotropical flora developed into to a more-or-less continuous mesophytic flora that ringed the high northern latitudes and existed without major geographic or climatic disruption during the late Eocene through the early Oligocene (Tiffney and Manchester 2001). The continuous mesophytic flora, precursor to the modern flora, was interrupted during the later Tertiary and Quaternary (Tiffney 1985).

A possible rapid radiation of *Aesculus* from the Beringian region in the Paleocene with subsequent intercontinental migrations is inferred for *Aesculus* in this study. This radiation out of Beringia is consistent with patterns floristic evolution and spread observed for other taxa originating in that area (Budantsev 1992). The current disjunct distribution of Aesculus is a result of migrations followed by geographic isolation and extinctions at different geologic times.

Implications for future studies of Laurasian disjunct taxa

Our results show the importance of wNA in the early biogeographic history of *Aesculus*, despite the comparatively small number of extant *Aesculus* species in those areas. The inference of an early presence of *Aesculus*, the ancestor of *Aesculus*, and the Sect. *Pavia* + Sect. *Parryana* + Sect. *Aesculus* clade in wNA is largely attributed to the presence of the *A. hickeyi* fossil in wNA. This indicates the importance of including fossils from outside of modern distribution areas in biogeographic analysis. Even with the broader disjunct distribution of extant *Aesculus* species in Laurasian regions (i.e.; in addition to EA and wNA), an "out-of-Asia" migration pattern may have been inferred as in Xiang et al. (1998) if the study had included extant taxa only. However, with the inclusion of fossils, we detected an "out-of-Asia-wNA" migration for *Aesculus*. Inclusion of fossils in biogeographic analysis of Laurasian disjunct taxa is essential for determining the global patterns of the the Laurasian disjunction (Donoghue and Smith 2004).

Caveats

In this study, divergence time estimations rely heavily on priors set in the program Multidivtime. This, along with wide CIs such as those defining the divergence of *A. tsiangii*, *A. chinensis*, *A. wilsonii*, and *A. wangii* in the Asian clade and among species in Sect. *Pavia* (Fig. 10), is indicative of low sequence variation and negligible branch length (Thorne, pers. comm.) (Fig. 8b). Results, especially those with wide CIs, should be interpreted with caution. Additional sequence data with higher interspecies variability may alter estimations. Further, new data supporting placement of the oldest *Aesculus* fossils, *A. hickeyi* and *A.* 'magnificum' (Budantsev 1983, Manchester 2001), elsewhere

on the tree topology and, hence, their use for age constraint of different nodes than were constrained in this study, may slightly alter the divergence time dating results shown here.

It also should be noted that it is unknown whether *LFY* is truly single-copy in *Aesculus* and whether ITS has completeted concerted evolution. The potential for lack of concerted evolution among the homologues of ITS is well documented (examples: Sang 2002, Doyle et al. 2004). Inadvertently isolating and amplifying non-homologous copies of ITS and/or *LFY* can result in noise and in errors in phylogenetic reconstruction. However, congruence of our ITS phylogeny with that of Xiang et al. (1998, Fig. 1c) and with the *LFY* phylogeny, suggests that concerns related to these issues should be minimal.

CONCLUSIONS

Our total evidence phylogeny (DNA + morphology, excluding fossils) shows strong support for 2 of the three polytypic sections of *Aesculus*: Sect. *Pavia* and Sect. *Aesculus*. The third polytypic section, Sect. *Calothyrsus*, was shown to be monophyletic if *A. californica* is excluded as reported by Xiang et al. (1998). *Aesculus californica*, traditionally treated as a constituent of Sect. *Calothyrsus*, shares a more recent common ancestor with Sect. *Macrothyrsus*, a monotypic section with an eNA distribution. Results here indicate that Section *Aesculus* is sister to a clade consisting of Sect. *Parryana* + Sect. *Pavia*, a finding new to this study. The phylogeny presented here shows an early

divergence of *Aesculus* into three major lineages; an Asian clade (all constituents of Sect. Calothyrsus excluding A. *californica*), Sect. *Parryana* + Sect. *Pavia* + Sect. *Aesculus*, and A. *californica* + Sect. *Macrothyrsus*. Relationships between these lineages remain unresolved, suggesting a possible early rapid radiation iof *Aesculus* in the Paleocene. Intercontinental migrations primarily involved the Bering Land Bridge in the early Tertiary. These early migrations are "out-of-Asia-and-wNA." Later migrations may rely on the North Atlantic Land Bridge or the arctic coastline. Our results also agree with a plastid lineage sorting hypothesis proposed by Xiang et al. (1998) to explain the different placements of *A. californica* in the chloroplast and nuclear DNA tree topologies.

REFERENCES

Axlerod, D. Mio-Pliocene floras from West-Central Nevada. *University of California Publications in Geological Sciences*. University of California Press, Berkeley and Los Angeles (1956)

Baum, D. and M. Donoghue. Choosing a species concept. *Systematic Botany* 20(4): 560-573 (1995)

Blaxter, M. The promise of DNA taxonomy. *Philosophical Transactions of the Royal Society in Biology*. 359: 669-679 (2004)

Bombardelli, E., Pet al. *Aesculus hippocastanum* L. *Fitoterapia*. 67(6): 483-511 (1996)

Bower, *A.* Delimitation of Phylogenetic Species with DNA Sequences: A Critique of Davis and Nixon's Population Aggregation Analysis**.** *Systematic Biology***.** 48(1): 199-213 (1999).

Bremer, K. Ancestral Areas: A cladistic reinterpretation of the center of origin concept. *Systematic Biology*. 41(4): 436-445 (1992)

Budantsev, L.Y. History of the Artic flora of the early Cenophytic epoch. Nauka, Leningrad. 156pp. (1983) (In Russian, cited and discussed in Manchester 2001).

Budantsev, L.Y. Early stages of formation and dispersal of the temperate flora in the boreal region. *Botanical Review*. 58(1): 1-48 (1992)

Buijsen, J. et al. A phylogenetic analysis of Harpullia (Sapindaceae) with notes on historical biogeography. *Systematic Botany*. 28(1): 106-117 (2003)

Burkhill, J. On the dispersal of plants most intimate to Buddhism. *Journal of the Arnold Arboretum*. 27(4): 327-339 (1946)

Burnham, R. and A. Graham. The history of neotropical vegetation: New developments and status. *Annals of the Missouri Botanical Gardens.* 86(2): 546-589 (1999)

Chaw et al. Dating the monocot–dicot divergence and the origin of core eudicots using whole chloroplast genomes. *Journal of Molecular Evolution*. 58(4): 424-441 (2004)

Crane et al. The origin and early diversification of angiosperms. *Nature*. 347: 27-33 (1995)

Davis, J. and K. Nixon. Populations, genetic variations, and the delimitation of phylogenetic species. *Systematic Biology*. 41(4): 421-435 (1992)

Dawson, G. Note of the plants collected by GM Dawson, from the lignite tertiary deposits near the forty-ninth parallel. Report on the geology and resources in the vicinity of the forty-ninth parallel, from the Lake of the Woods to the Rocky MountainsBritish North American Boundary Commission. Ed. GM Dawson. p. 227-231 (1875)

de Lumley, H. La stratigraphie du remplissage de la Grotte du Vallonnet. *L'Anthropologie* 92(2):407-428 (1988)

dePamphilis, C. and R. Wyatt. Hybridization and introgression in buckeyes (*Aesculus*: Hippocastanaceae): a review of the evidence and a hypothesis to explain long distance gene flow. *Systematic Botany*. 14(4): 593-611 (1989)

Desmure et al. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology*. 4: 129-131 (1995).

Donoghue, M. and B. Moore. Towards an integrative historical biogeography. *Integrative and Comparative Biology*. 43(2): 261-270 (2003)

Donoghue, M. and S. Smith. Patterns in the assembly of temperate forests around the Northern Hemisphere. *Philosophical Transactions of the Royal Society in Biological Sciences*. 359(1450): 1633-1644 (2004)

Doyle et al. Evolution of the perennial soybean polyploid complex (Glycine subgenus Glycine: a study of contrasts. *Biological Journal of the Linnean Society.* 82: 583-597 (2004)

Engelhardt, H. Tertiarflora des Jesuitengrabens bie Kundratitz in Nordbohmen. *Nova Acta* (1885) (In German).

Fan, C. and Q-Y Xiang. Phylogenetic relationships within *Cornus* (Cornaceae) based on 26S rDNA sequences. *American Journal of Botany.* 88(6): 1131-1138 (2001)

Forest, et al. A morphological phylogenetic analysis of *Aesculus* L. and *Billia* Peyr. (Sapindaceae). *Canadian Journal of Botany*. 79: 154-169 (2001)

Golovneva, L. Early Paleogene Floras of Spitzbergen and North Atlantic Floristic Exchange. *Acta Universitatis Carolinae Geologica*. 44(1): 39-50. (2000)

Hardin, J. A revision of the American Hippocastanaceae. *Brittonia*. 9(3): 145-171 (1957a)

Hardin, J. A revision of the American Hippocastanaceae II. *Brittonia*. 9(3): 173-195 (1957b)

Hardin, J. Studies in the Hippocastanaceae: Old World Species. *Brittonia*. 12(1): 26-38 (1960)

Hu, H.H. and R. Chaney. A Miocene flora from Shatung Province, China. *Contributions to Paleobotany: Carnegie Institute of Washington Publication no. 507*. (1940)

Hollick, A. The Tertiary floras of Alaska. *U.S. Geological Survey*. no. 182 (1936)

Horikoshi, E. Opening of the Sea of Japan and the Koroko deposit formation. *Mineralium Deposita.* 12(2): 140-145 (1990)

International Organization of Paleobotany (IOP). Plant Fossil Record v.2.2 (PFR2.2) Online at http://www.biologie.uni-hamburg.de/b-online/library/iopaleo/pfr.htm. (1997).

Jeong, E. et al. Fossil woods from Janggi Group (Early Miocene) in Pohang Basin, Korea. *Journal of Plant Research*. 117: 183-189 (2004)

Kamer, P. The woods of *Billia*, Cashalia, Henoonia, and Juliania. *Tropical Woods*. 58: 1-2 (1939).

Kauff, F., J. Miadlikowska, and F. Lutzoni. ARC: a program for Ambiguous Regions Coding. Available from the authors http://www.lutzonilab.net/pages/download/ shtml (2003).

Kearney, M. and J. Clarke. Problems due to missing data in phylogenetic analyses including fossils: A critical review. *Journal of Vertebrate Paleontology.* 23(2): 263-274 (2003)

Keigwin, L. Pliocene closing of the Isthmus of Panama, based on biostratigraphic evidence from nearby Pacific Ocean and Caribbean Sea cores. *Geology*. 6(10): 630-634 (1978)

Klaassen, R. et al. Wood anatomy of trees and shrubs from China: VII. Sapindaceae. *IAWA Journal*. 16(2): 191-215 (1995)

Knowlton, F. Fossil plants from Kukak Bay. *Harriman Alaska Expedition* vol. 4. US Geological Survey (1904).

Lorenz et al. The problems and promise of DNA barcodes for species diagnosis of primate biomaterials. *Philosophical transactions of the Royal Society in Biology*. 360: 1879-1878 (2005)

McIver, E. The paleoenvironment of Tyrannosaurus rex from southwestern Saskatchewan, Canada. Can. J. Earth Sci. 39: 207-221 (2002)

Maddison, D. and W. Maddison. MacClade v.4.02. Sinauer Associates, Inc. Publishers. (2001)

Mai, D. and H. Walther. Die pliozaenen Floren von Thueringen, Deutsche Demokratische Republik. *Quartaerpalaeontologie* 7:55-297 (1988)

Manchester, S. Leaves and fruits of *Aesculus* (Sapindales) from the Paleocene of North America. *International Journal of Plant Science.* 162(4):985-988 (2001).

Miadlikowska, et al. New approach to an old problem: Incorporating signal from gap-rich regions of ITS and rDNA large subunit into phylogenetic analysis to resolve the Peltigera species complex. *Mycologia*. 95(6): 1181-1205 (2003).

Modliszewski, J. et al. Ancestral chloroplast polymorphism and historical secondary contact in a broad hybrid zone of *Aesculus* (Sapindaceae). *American Journal of Botany* 93(3): 377-388 (2006).

Nixon, K and Q. Wheeler. Extinction and the origin of species. Ed. M. Novacek and Q. Wheeler. *Extinction and Phylogeny*. Columbia University Press, New York. pp 119-143 (1992)

Oxelman, B, M. Liden, and D. Berglund. Chloroplast rps16 intron phylogeny of the tribe *Sileneae* (Caryophyllaceae). *Plant Systematics and Evolution.* 206: 393-410 (1997)

Ozaki, K. Late Miocene and Pliocene floras in central Honshu, Japan. *Bulletin of the Kanagawa Prefectural Museum of Natural Science*. special issue (1991).

Paleobiology Database, The. Online at https://paleodb.org.

Posada D and Crandall K*A.* Modeltest: testing the model of DNA substitution. Bioinformatics 14 (9):817-818 (1998).

Prakash, U. and E. Barghoorn. Miocene fossil woods from the Columbia Basalts of Central Washington, II. 42: 165-203 (1961)

Puri, G.S. Some fossil leaflets of *Aesculus indica* Colebr. from the Karewa Beds at Laredura and Ningal Nullah, Pir Panjal, Kashmir. *Journal of the Indian Botanical Society*. 24 (1945)

Rambaut, *A.* and *A.* Drummond. Tracer v.1.3. MCMC trace file analyzer program available from the authors at http://evolve.zoo.ox.ac.uk/software.html.

Record, S. and Robery Hess. *Timbers of the New World*. Yale University Press: New Haven. (1943).

Ronquist, F. DIVA version 1.1. Computer program and manual available by anonymous FTP from Uppsala University: ftp.uu.se or ftp.systbot.uu.se (1996)

Rosen, D. Vicariant patterns and historical explanation in biogeography. *Systematic Zoology*. 27(2): 159-188 (1978)

Soltis et al. *Phylogeny and Evolution of Angiosperms.* Sinaur Associates, Inc, Sunderland Massachusets. pp85-86 (2005)

Schloemer-Jager, *A.* Alttertiare pflanzen aus flozen der bragger-halbinsel Spitzbergens. Paleontographica Abt. B. 1904: 39-103. (In German)

Stevens, P. F. *Angiosperm Phylogeny Website*. Version 7. Available at http://www.mobot.org/MOBOT/research/APweb/ (2001 ff.)

Suzuki, M. and K. Terada. Fossil wood flora from the lower Yanagida Formation, Noto Peninsula, Central Japan. *IAWA* 17(4): 365-392 (1996)

Swofford, D. L. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Vers. 4.0b10. Sinauer Associates, Sunderland, M*A.* (2000 ff.)

Szafer, W. *The Pliocene Flora of Kroscienko in Poland*. 213 p., 15 pl (1947)

Szafer, W. Pliocene flora from the vicinity of Czorstzyn (West Carpathians) and its relationship to the Pleistocene. *Polish Geological Institute* 111:1-238 (1954)

Tanai, T. The fossil vegetation from the coalified basin of Nishitagawa, Prefecture of Yamagata, Japan. *The Japanese Journal of Geology and Geography*. 22: 119-135 (French). (1952)

Thompson, J. et al. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*. 24: 4876-4882 (1997)

Tiffney, H. The Eocene North Atlantic Land Bridge: its importance in tertiary and modern phytogeography of the Northern hemisphere. *Journal of the Arnold Arboretum*. 66(2): 243-273 (1985a)

Tiffney, H. Perspectives on the origin of the floristic similarity between eastern Asia and eastern North America. *Journal of the Arnold Arboretum*. 66(1): 73-94 (1985b)

Tiffney, B. and S. Manchester. The use of geological and paleontological evidence in evaluating plant phylogeographic hypotheses in the northern hemisphere tertiary. *International Journal of Plant Sciences.* 162: s3-S17 (2001)

Turland, N. and N. Xia. A new combination in Chinese *Aesculus* (Hippocastanaceae). *Novon*. 15(4): 488-489 (2005)

Thorne, D. and H. Kishino. Multidivtime. Available at http://statgen.ncsu.edu/thorne/multidivtime.html (2003)

Sang, T. Utility of low copy nuclear gene sequences in plant phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology.* 37(3): 121-147 (2002)

Suzuki, M. and K. Terada. Fossil wood flora from the lower Miocene Yanagida Formation, Noto Peninsula, Central Japan. *IAWA Journal*. 17(4): 365-392 (1996)

University of California Museum of Paleontology (UCMP). Plantae: Fossil Record. Online at http://www.ucmp.berkeley.edu/plants/plantaefr.html (1994a ff).

University of California Museum of Paleontology (UCMP). Tour of geologic time. *Online Exhibits.* Online at http://www.ucmp.berkeley.edu/exhibits/geologictime.php (1994b ff).

Weeks, *A.* and B. Simpson. Molecular genetic evidence for interspecific hybridization among endemic Hispaniolan *Bursera* (Burseraceae). *American Journal of Botany*. 91(6): 976-984 (2004)

Wen, J. Evolution of eastern Asian and eastern North American Disjunct Distributions in flowering plants. *Annual Review of Ecology and Systematics*. 30: 421-455 (1999)

Wharton et al. *The Jade Garden: New and Notable Plants from Asia*. Timber Press: Portland. (2005).

Wolfe, J. Some aspects of plant geography of the northern hemisphere during the late Cretaceous and Tertiary. *Annals of the Missouri Botanical Gardens*. 62(2): 264-279 (1975).

Xiang, Q.Y. et al. Origin and biogeography of *Aesculus* L. (Hippocastanaceae): A molecular phylogenetic perspective. *Evolution* 52(4): 988-997 (1998)

Xiang, Q.Y. et al. Timing of the eastern Asian-eastern North American floristic disjunction: molecular clock corroborates paleontological estimates. *Molecular Phylogenetics and Evolution*. 15(3): 462-472 (2000)

Xiang, Q.Y. et al. Phylogeny, biogeography, and molecular dating of cornelian cherries (Cornus, Cornaceae): tracking Tertiary plant migration. *Evolution*. 59(8): 1685-1700 (2005)

Xiang, Q.Y. et al. Species level phylogeny of the genus Cornus (Cornaceae) based on molecular and morphological evidence – implications for taxonomy and Tertiary intercontinental migration. *Taxon*. 55(1): 9-30 (2006)

Table 1: Taxon sampling: Reference numbers refer to the number by which individuals of the same species will be distinguished in figures, tables, and in the text.

Taxon	Voucher/Accession	Material available	Regions amplified/Sequence data available	Reference number
Ingroup				
Section *Aesculus*				
A. hippocastanum L.	Kew 00-69.11289-263	extracted DNA	*trn* HK, *rps* 16, *mat* K, ITS, *LFY*	-
A. turbinata Blume	J. Wen sene. non	sequence data	*mat* K, ITS	1
A. turbinata Blume	D.J. Crawford , 411	extracted DNA	*trn* HK, *rps* 16, *mat* K, ITS, *LFY*	2
A. turbinata Blume	JC Raulston Arboretum 950016	fresh leaves	*trn* HK, *rps* 16, *mat* K, ITS, *LFY*	3
Section *Calothyrsus* (traditional)				
A. assamica Griff.	Mongolia Expedition 10039	herbarium material	*trn* HK, *rps* 16, ITS, *LFY*	-
A. californica (Spach.) Nutt.	D.J. Crawford, 406	extracted DNA	*trn* HK, *rps* 16, *mat* K, ITS, *LFY*	1
A. californica (Spach.) Nutt.	T.M. Hardig, 2795	extracted DNA	*trn* HK, *rps* 16, *mat* K, ITS, *LFY*	2
A. californica (Spach.) Nutt.	J.C. Raulston arboretum 950413	fresh leaves	*trn* HK, *rps* 16, *mat* K, ITS, *LFY*	3
A. californica (Spach.) Nutt.	UC Berkeley 93.1203	frozen leaves	*trn* HK, *rps* 16, *mat* K, ITS, *LFY*	4
A. californica (Spach.) Nutt.	UC Berkeley 93.1116	frozen leaves	*trn* HK, *rps* 16, *mat* K, ITS, *LFY*	5

Table 1 cont.

A. chinensis Bunge	Q.Y. Xiang 305	extracted DNA	*trn* HK, *rps* 16, ITS, *LFY*	1
A. chinensis Bunge	Q.Y. Xiang 04-C88	frozen leaves	*trn* HK, *rps* 16, *mat* K, ITS, *LFY*	2
A. chinensis Bunge	Q.Y. Xiang 06-12	frozen leaves	*trn* HK, *rps* 16, ITS	3
A. chinensis Bunge	Q.Y. Xiang 06-17	frozen leaves	*trn* HK, *rps* 16, *LFY*	4
A. indica (Camb.) Hook	Q.Y. Xiang 301	extracted DNA	*trn* HK, *rps* 16, *mat* K, ITS, *LFY*	1
A. indica (Camb.) Hook	J.C. Raulston arboretum 001405	frozen leaves	*trn* HK, *rps* 16, *mat* K, ITS, *LFY*	2
A. polyneura Hu & Fang	Q.Y. Xiang 02-255	frozen leaves	*trn* HK, *rps* 16, *mat* K, ITS, *LFY*	-
A. tsiangii Hu & Fang	Q.Y. Xiang 04-C37	frozen leaves	*trn* HK, *rps* 16, *mat* K, ITS, *LFY*	-
A. wilsonii Rehder.	Q.Y. Xiang 02-105	frozen leaves	*trn* HK, *rps* 16, ITS, *LFY*	1
A. wilsonii Rehder.	Q.Y. Xiang 04-C9	frozen leaves	*trn* HK, *rps* 16, *mat* K, ITS, *LFY*	2
A. wangii Hu	Q.Y. Xiang 303	extracted DNA	*trn* HK, *rps* 16, *mat* K, ITS, *LFY*	-
Section *Macrothyrsus*				
A. parviflora Walter	Kew 00069-10442-265	sequence data	*mat* K, ITS	1

Table 1 cont.

A. parviflora Walter	J.C. Raulston Arboretum sene non.	fresh leaves	*trn* HK, *rps* 16, *mat* K, ITS, *LFY*	2
Section *Pavia*				
A. glabra Willd.	D.J. Crawford 413	extracted DNA	*trn* HK, *rps* 16, *mat* K, ITS, *LFY*	1
A. glabra Willd.	C.W. DePamphilis 931	sequence data	*mat* K, ITS	2
A. flava Sol.	D.J. Crawford 408	sequence data	*mat* K, ITS	1
A. flava Sol.	C.W. DePamphilis F-MI-4	sequence data	*mat* K	2
A. flava Sol.	Q.Y. Xiang 98-150	extracted DNA	*trn* HK, *rps* 16, *mat* K, ITS, *LFY*	3
A. flava Sol.	Q.Y. Xiang 01-190	extracted DNA	*mat* K, ITS, *LFY*	4
A. pavia L.	D.J. Crawford 404	sequence data	*mat* K	1
A. pavia L.	C.W. DePamphilis	sequence data	*mat* K, ITS	2
A. pavia L.	Q.Y. Xiang 01-54	extracted DNA	*trn* HK, *rps* 16, *mat* K, ITS, *LFY*	3
A. pavia L.	Q.Y. Xiang 98-135	extracted DNA	*rps* 16, *mat* K, ITS	4
A. sylvatica Bart.	C.W. DePamphilis 50	sequence data	*mat* K	1

Table 1 cont.

A. sylvatica Bart.	C.W. DePamphilis S-Ga-2	sequence data	*mat* K, ITS	2
A. sylvatica Bart.	Q.Y. Xiang 01-251	extracted DNA	*trn* HK, *mat* K, ITS, *LFY*	3
A. sylvatica Bart.	Q.Y. Xiang 98-110	extracted DNA	*trn* HK, *rps* 16, *mat* K, ITS, *LFY*	4
Section *Parryana*				
A. parryi Gray	RSABG BC-17230	sequence data	*mat* K, ITS	1
A. parryi Gray	Epling 1936 sene non.	herbarium material	*trn* HK, *rps* 16, *mat* K, ITS	2
Outgroup				
Handeliodendron bodinieri (Levl.) Rehd.	Q.Y. Xiang 302	extracted DNA	*trn* HK, *rps* 16, *mat* K, ITS, *LFY*	-
Billia Peyr sp.	B. Hammel 20075	sequence data	*mat* K, ITS	1
Billia Peyr sp.	Q.Y. Xiang 02-12	extracted DNA	*trn* HK, *rps* 16, *mat* K, ITS, *LFY*	2

Table 2: Base composition of amplification and sequencing primers (5'-3') - ǂ The reverse compliments of the trn2F and the 829F sequences were used as reverse primers; [f] Forward primer; [r] Reverse primer; * A degenerate primer was also designed using this sequence: TTT CAT CGT CCT C**RW** TCT CA; [2] Series 2 LFY primers (see text, Materials and Methods)

Region	Primer name	Primer Sequence	Source
matK			
	1F	ACT GTA TCG CAC TAT GTA TCA	Modliszewski et al. 2006
	300F	GGG ATT TGC AGT CAT TGT GG	Xiang Lab (unpub.)
	500R	TGG ACR GGR TRG GGT ATT AG	Xiang Lab (unpub.)
	3F	AAG ATG CCT CTT CTT TGC AT	Modliszewski et al. 2006
	799F	TTC TGG ACT CCT TCT TGA GCA	This Study
	920R	GCC AGA ATR SAT TTT CCT TG	Xiang Lab (unpub.)
	1288F	TAT TAT CGA CCG GTT TGT GC	This Study
	1416R	CGC GCA CAG TAC TTT TGT GT	This Study
	3R	GAT CCG CTG TGA TAA TGA GA	Modliszewski et al. 2006
	1R	GAA CTA GTC GGA TGG AGT AG	Modliszewski et al. 2006
trnHK			
	trn H[f]	ACG GGA ATT GAA CCC GCG CA	Desmure et al. 1995
	trn H2F[ǂ]	ACT CGT ATA CAC GAA GAT CG	Modliszewski et al. 2006
	trn HK-F497	ACG TTC GTG CAT AAC TTC CA	Xiang Lab (unpub.)
	514R	CCG TGC TAA CCT TGG TAT GG	Xiang Lab (unpub.)
	829F[ǂ]	GGC CAA GCA GCT AGA AAG AA	This Study
	1231F	ATG CAA CAG CAA TCC AAG G	This Study
	trn 2R	TGA ACC CGT TTC TGG ATC TC	Modliszewski et al. 2006
	trn 3F	CTT ATA GCC CCG TGT CAA CC	Modliszewski et al. 2006
	1779R	CAC GGG GCT ATA AGT CAT GTT	This Study
	trn K[r]	CCG ACT AGT TCC GGG TTC GA	Desmure et al. 1995
rps16			
	rps F	GTG GTA GAA AGC AAC GTG CGA CTT	Oxelman et al. 1997
	rps R2	TCG GGA TCG AAC ATC AAT TGC AAC	Oxelman et al. 1997

Table 2 cont.

ITS			
	ITS5 **f**	GGA AGG AGA AGT CGT AAC AAG G	Xiang et al. 1998
	ITS2 **r**	GCT GCG TTC TTC ATC GAT GC	Xiang et al. 1998
	ITS3 **f**	GCA TCG ATG AAG AAC GCA GC	Xiang et al. 1998
	ITS4 **r**	TCC TCC GCT TAT TGA TAT GC	Xiang et al. 1998
LFY			
	LFY 1F	TTC WCG GCS AGY TTR TTC AAG TGG G	Degenerate (our lab)
	LFY 2R	TAA TCA AGR CCR TTC TTY TTS CCAA C	Degenerate (our lab)
	Aes*LFY* 1F	GAT TTG GGT GGG CTG GAG	This Study
	Aes*LFY* 1R*	TTT CAT CGT CCT CGT TCT CA	This Study
	Aes 2 *LFY* F^2	TGG ACA TGA AAG ACG AGG AG	This Study
	Aes 2 *LFY* R^2	CTG CCG CCT CTT TCT CTT G	This Study
	Aes 2 *LFY* $2F^2$	GGT CAT CGT TAT TCT CGS ACT C	This Study
	Aes 2 *LFY* $2R^2$	GCT AAG CTC AAG CAG TCA ACC	This Study

Table 3: Matrix of 42 morphological characters used in phylogenetic analysis, modified from Forest et al. (2001) - Showing the matrix of 42 characters, modified from Forest et al. (2001) used in phylogenetic analysis.Modifications are included in text (see: Materials and Methods). Characters 1-39 are presented in the same order as they were presented in Forest et al. Characters 40-42 are listed and described in Appendix 1. Taxa for which scored morphological charactes are not available: *A. polyneura; A. tsiangii; and A. wangii* . Aesculus 'magnificum' refers to the fossils described by Budantsev (1983). [f] = fossil species, / = polymorphic

Species	**Characters**																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
***Aesculus* Ingroup**																			
Aesculus assamica	1	1	0	0	0	1	1	0	0	1	0	1	0	1	1	2	0	0	1
A. chinensis	1	1	0	0	0	1	1	0	0	1	0	1	0	1	1	2	0	0	1
A. wilsonii	1	1	0	0	0	1	1	0	0	1	0	1	0	1	1	2	0	0	1
A. indica	1	1	0	0	0	1	1	0	0	1	0	1	0	1	1	2	0	0	1
A. parryi	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1
A. californica	0	1	0	0	0	1	1	0	0	0	0	1	0	0	1	2	0	0	1
A. parviflora	0	1	0	0	0	1	1	0	0	0	0	1	0	0	1	2	0	0	0
A. turbinata	1	1	0	0	0	0	1	1	0	0	1	1	0	1	0	2	0	0	1
A. hippocastanum	1	1	0	0	0	0	1	1	0	0	1	1	0	1	0	2	0	0	1
A. sylvatica	1	0	0	0	0	1	1	0	1	0	1	1	0	0	1	2	0	0	1
A. pavia	1	0	0	0	0	1	1	0	1	0	1	1	0	0	1	2	0	0	1
A. glabra	1	0	0	0	0	1	1	0	1	0	0	1	0	0	0	2	0	0	1
A. flava	1	0	0	0	0	1	1	0	1	0	1	1	0	0	1	2	0	0	1
A. hickeyi [f]	?	?	?	0	?	0	1	?	?	?	0	?	?	?	?	?	?	?	?
A. longipedunculus [f]	?	?	?	0	?	1	1	?	?	?	?	?	?	?	?	?	?	?	?
A. miochinensis [f]	?	?	?	?	?	1	?	?	?	?	?	?	?	?	?	?	?	?	?
A. majus [f]	?	?	?	?	?	?	1	?	?	?	?	?	?	?	?	?	?	?	?
A. 'magnificum'[f]	?	?	?	?	?	1	1	?	?	?	?	?	?	?	?	?	?	?	?
Outgroup																			
Handeliodendron	0	?	0	0	0	1	0	0	0	0	?	0	0	0	0	1	0	0	0
Billia sp.	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0

Species	**Characters**																		
	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
***Aesculus* Ingroup**																			
Aesculus assamica	1	1	0	0	0	0	0	0	0	1	1	0	0	0	1	2	0	0	2
A. chinensis	1	1	0	0	0	0	0	0	0	1	1	0	0	1	1	2	0	0	2
A. wilsonii	1	1	0	0	0	0	0	0	0	1	1	0	0	1	1	2	0	0	2
A. indica	1	1	0	0	0	0	0	0	0	1	1	0	1	1	1	2	0	0	2
A. parryi	1	1	0	1	0	0	1	0	0	0	1	0	1	1	1	0	0	1	2
A. californica	1	1	0	0	0	0	0	0	0	1	1	0	0	1	1	2	0	?	2

Table 3 cont.

A. parviflora	0	1	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	?	2
A. turbinata	1	1	1	0	2	0	0	0	1	1	1	0	1	1	1	2	0	0	2
A. hippocastanum	1	1	1	0	2	0	0	0	1	1	1	0	1	1	1	2	0	1	2
A. sylvatica	2	1	1	0	0	1	0	1	0	0	1	0	0	0	1	0/1	0	0	2
A. pavia	2	1	1	0	0	1	1	1	0	0	1	0	1	0	1	0/1	0	0	2
A. glabra	2	1	1	0	1	0	0	1	0	1	1	0	1	1	1	0	0	1	2
A. flava	2	1	1	0	0	1	1	1	0	0	1	0	0	0	1	0/1	0	0	2
A. hickeyi [f]	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	1	?	1	1
A. longipedunculus [f]	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
A. miochinensis [f]	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
A. majus [f]	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
A. 'magnificum'[f]	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Outgroup																			
Handeliodendron	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	?	0	0	1
Billia sp.	1	1	1	1	2	0	0	0	0	1	1	0	0	?	?	?	0	0	2

Species	**Characters**			
***Aesculus* Ingroup**	39	40	41	42
Aesculus assamica	?	0	1	1
A. chinensis	2	0	1	0
A. wilsonii	2	0	1	0
A. indica	?	0	1	0
A. parryi	1	1	0	0
A. californica	1	0	1	0
A. parviflora	1	0	1	0
A. turbinata	2	1	0	0
A. hippocastanum	2	1	0	0
A. sylvatica	1	1	0	0
A. pavia	1	1	0	0
A. glabra	1	1	0	0
A. flava	1	1	0	0
A. hickeyi [f]	2	1	?	0
A. longipedunculus [f]	?	1	1	?
A. miochinensis [f]	?	0	?	?
A. majus [f]	?	0	?	?
A. 'magnificum'[f]	?	?	?	?
Outgroup				
Handeliodendron	?	?	?	?
Billia sp.	?	0	1	?

Table 4: *Aesculus* fossils included in this study: The references are cited in the Chapter 1 reference list. When more than one reference is listed, the prologue may be found in the first reference in the list. 1 - Prakash and Barghoorn 1961, 2 - Manchester 2001, 3 - Szafer 1947, 4 - Szafer 1954, 5 - Mai and Walther 1988, 6 - de Lumley 1988, 7 - Puri 1945, 8 - Schloemer-Jager 1958, 9 - Golovneva 2000, 10 - Budantsev 1983, 11 - Tanai 1952, 12 - Hu and Chaney 1940

Epithet	Laurasian Range	Material	Age Range	References
A. hankensii Prakash & Barghoorn	western North America	wood	Miocene-Miocene	1
A. hickeyi Manch.	western North America	leaves, pollen, fruits	Paleocene-Paleocene	2
A. hippocastanum L.	mainland Europe	seed/fruit	Pliestocene-Pliocene	3, 4, 5, 6
A. indica (Camb.) Hook	East Asia	leaflets	Pliestocene	7
A. longipedunculus Schloemer-Jager	Greenland Region	leaves	Lower Eocene-Lower Eocene	8, 9
A. 'magnificum' det. Budantsev	East Asia	leaves	Paleocene-Early Eocene	10, 2
A. majus (Nathorst) Tanai	East Asia	leaflets	Upper Oligocene-Lower Miocene	11
A. miochinensis Hu & Chaney	East Asia	leaflets	Miocene-Miocene	12

Table 5: Adjacent microsatellites and flanking regions in *Aesculus LFY* sequence - AT and CAT repeats are boxed. The ~459bp repeat in *A. tsiangii* immediately follows the 3' end of the sequence shown below. Ambiguous bases are shown using standard IUPAC characters. In *A. flava*, the remainder of the sequence downstream of the microsattellite region was unreadable. However, the number of ambiguous bases (Ns) returned in the chromatogram and accompanying sequence data indicates that this sequence is comparable in length to *A. flava*, and, therefore, is not thought to include the same bp repeat found in *A. tsiangii*.

A. californica 3	TTGAGCTTAGCTTCTATG-TATATATATATA----------GACACA----CATACATACATAC----AATATAGAACTTTTGT
A. californica 5	TTGAGCTTAGCTTCTATG-TATATATATATATATA-------CACACA----CATACATACATAC----AATATAGAACTTTTGT
A. californica 4	TTGAGCTTAGCTTCTATG-TATATATATATATATA-------CACA------CATACATACATAC----AATATAGAACTTTTGT
A. californica 1	TTGAGCTTAGCTTCTATG-TATATATATGTATA---------CACACA----CATACATACATAC----AATATAGAACTTTTGT
A. californica 2	TTGAGCTTAGCTTCTATG-TATATATATATA----------CACACACA--CATACATACATAC----AATATAGAACTTTTGT
A. parviflora 2	TTGAGCT----------ATATA----------CACACA----CATACATACATAT----AATATAAAACTTTTGT
A. hippocastanum	TTGAGCTTAGCTT-------ATATATA----------CATACATAT--------AATATAAAACTTTTGT
A. turbinata 2	TTGAGCTTAGCTT-------ATATA----------TACATGCATAT----AATATAAAACTTTTGT
A. assamica	TTGAGCTTAGCTT-------ATATATATA----------CATAT----------AATATTAGACTTTTGT
A. indica 1	TTGAGCTTAGCTT-------ATATATATA----------CATAT----------AATATAAAACTTTTGT
A. indica 2	TTGAGCTTAGCTT-------ATATATATA----------CATAT----------AATATAAAACTTTTGT
A. wangii	TT----TTAGCTT-------ATATATATA----------CATAT----------AATATAAAACTTTTGT
A. polyneura	??????????????????TATATATATA----------CATAT----------AATATAAAACTTTTGT
A. wilsonii 3	TTGAGCTTAGCTT-------ATATATATA----------CATAT----------AATATAAAACTTTTGT
A. wilsonii 2	TTGAGCTTAGCTT-------ATATATATA----------CATAT----------AATATAAAACTTTTGT
A. tsiangii	TTGAGCTTAGCTT-------ATATATATATA----------CATAT----------AATATAAAACTTTTGT
A. chinensis 4	TTGAGCTTAGCTT-------ATATATATATA----------CATAT----------AATATAAAACTTTTGT
A. chinensis 1	TTGAGCTTAGCTT-------ATATATATA----------CATAT----------AATATAAAACTTTTGT
A. chinensis 2	TTGAGCTTAGCTT-------ATATATATATA----------CATAT----------AATATAAAACTTTTGT
A. pavia 3	TTGAGCTTAGCTTC-----TATATATATA----------CACGCACGCACACAT----------AATATAAAACTTTTGT
A. flava 4	TTGAGCTTAGCTTC-----TATATATATA----------CACGCACGCACACAT----------AATATAAAACTTTTGT
A. flava 3	TTGAGCTTAGCTTC-----TATATATATA----------CACGCACGCACACAT----------AATATAAAA????????
A. sylvatica 3	TTGAGCTTAGCTTC-----TATATATATA----------CACGCACGCACACAT----------AATATAAAACTTTTGT
A. sylvatica 4	TTGAGCTTAGCTTC-----TATATATATA----------CACGCACGCACACAT----------AATATAAAAMYTTTG?
A. glabra 1	TTGAGCTTAGCTTC-----TATATATATA----------CACGCGCACACACAT----------AATATAAAACTTTTGT
Billia sp. 1	TTAAGCTTAGCTTCCATAATATATATATATATATATATATATATAT----------CTTTTGT
Handeliodendron	TTGAGCTTAGCTTC-----TATATATATA----------CACATACATTCATACATAAATAT----AATATAAAACTTTTGT

Table 6: Sequence information for the amplified portions of *rps* 16, *trn* HK, *mat* K, ITS, and *LFY* - Percent informative characters was calculated from sequence data with gaps treated as missing. [a] = Excluding A. *tsiangii* (see text), [b] = Excluding *Handeliodendron* and *A. tsiangii* (see text)

Gene	Average # of bp	% informative characters
rps 16	812	2.6%
trn HK	1727[a]	2.1%
mat K	1812	5.4%
ITS	597	15.9%
LFY	715[b]	4.5%

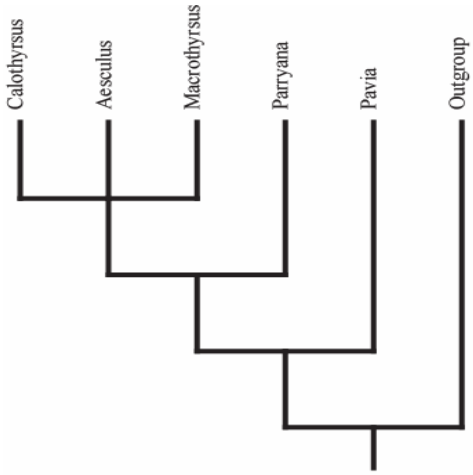

Fig. 1a: Hardin's phylogeny (1957a).

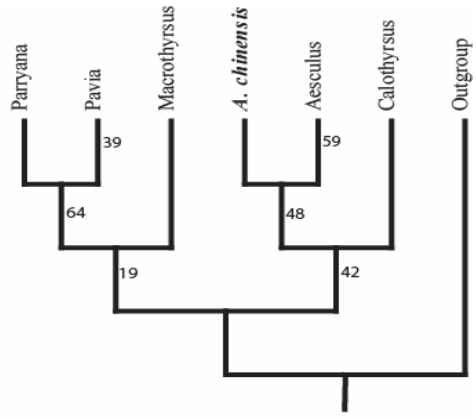

Fig. 1b: One of 80 *mat*K MP trees from Xiang et. al (1998) with bootstrap support.

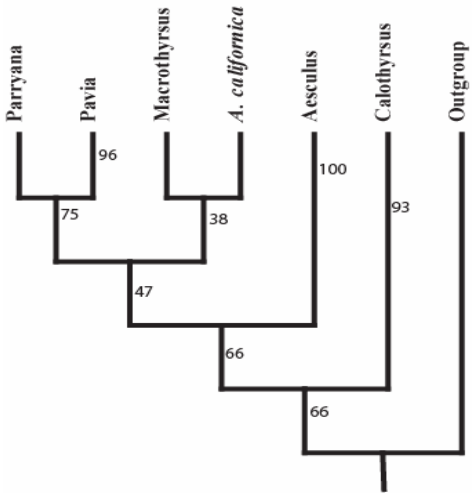

Fig. 1c: ITS MP tree from Xiang et. al (1998) with bootstrap support.

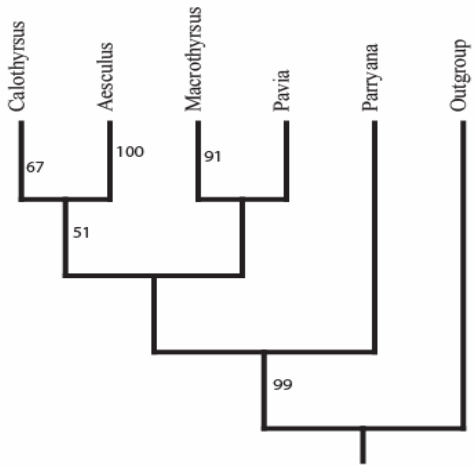

Fig. 1d: One of 2 MP trees from Forest et al. (2001), shown with bootstrap support.

Fig. 1a- 1d: Summary trees showing previously published phylogenies of *Aesculus* – Labels are section names, except where italics are used indicating species placed outside of their traditional section affiliations. Outgroups are as described by the authors (see References).

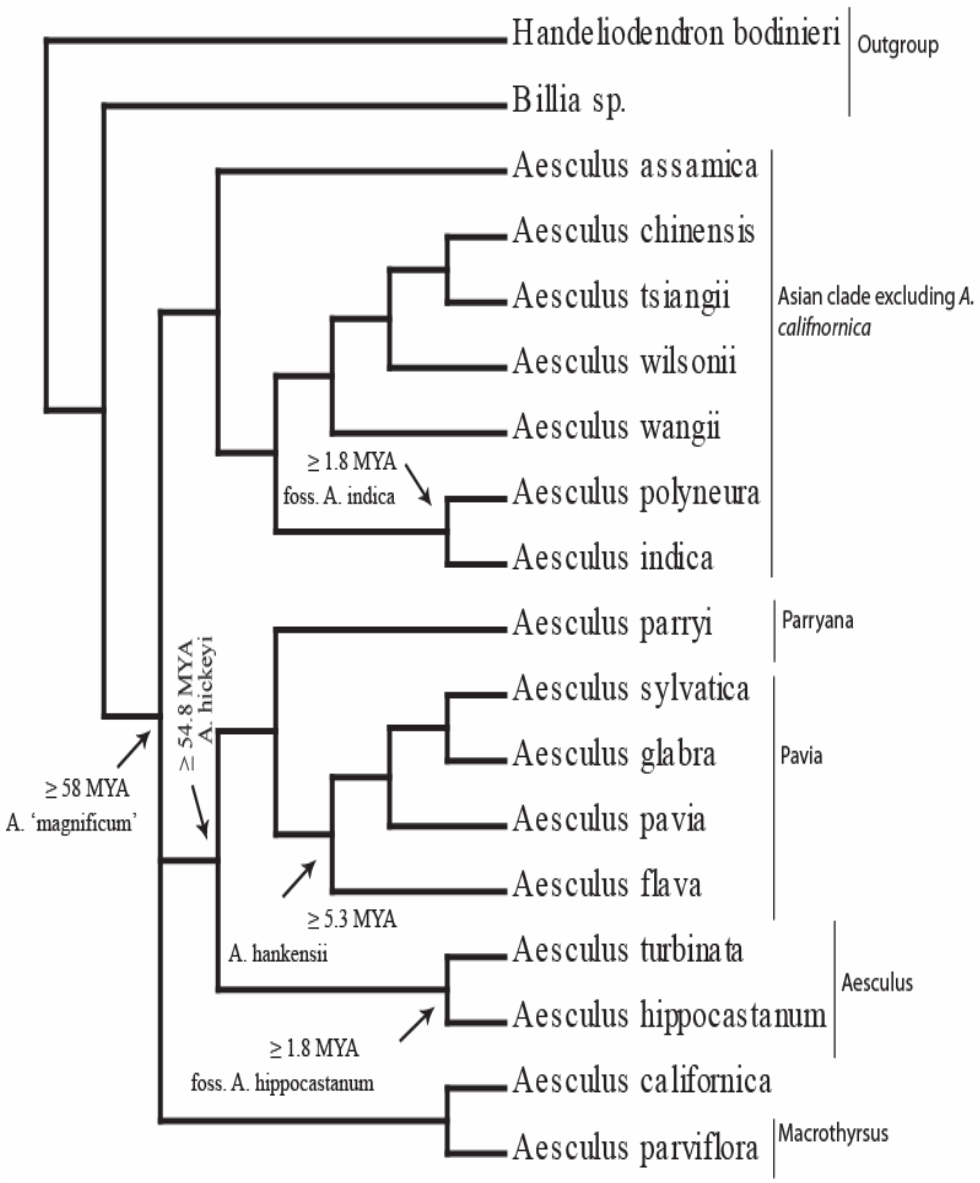

Fig. 2: Age constraints used in Multidivtime dating. Names of sections or clades are given to the right. Dates are given in millions of years. The fossil taxa used to constrain nodes are provided in the figure. The phylogeny will be discussed later in the text.

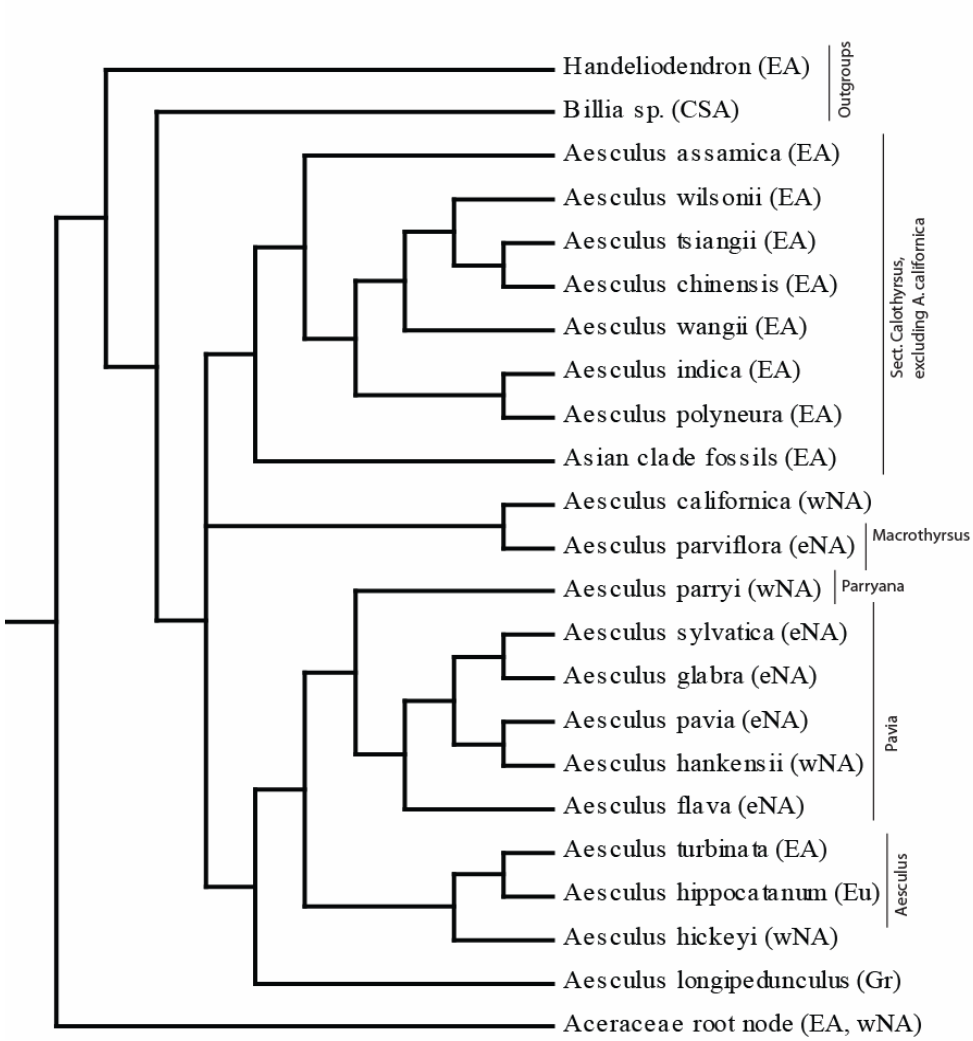

Fig 3: Base tree used for DIVA analysis – See text for information of resolution of the polytomy and probabilistic calculation. Distributions of modern species and fossil localities are shown to the right of terminals. Section affiliations are given for each clade. Where fossils are nested within traditionally recognized groups, they are shown as belonging to those groups. wNA = western North America, eNA = eastern North America, EA = East Asia, Eu = Europe, CSA = Central and South America, Gr = Greenland region

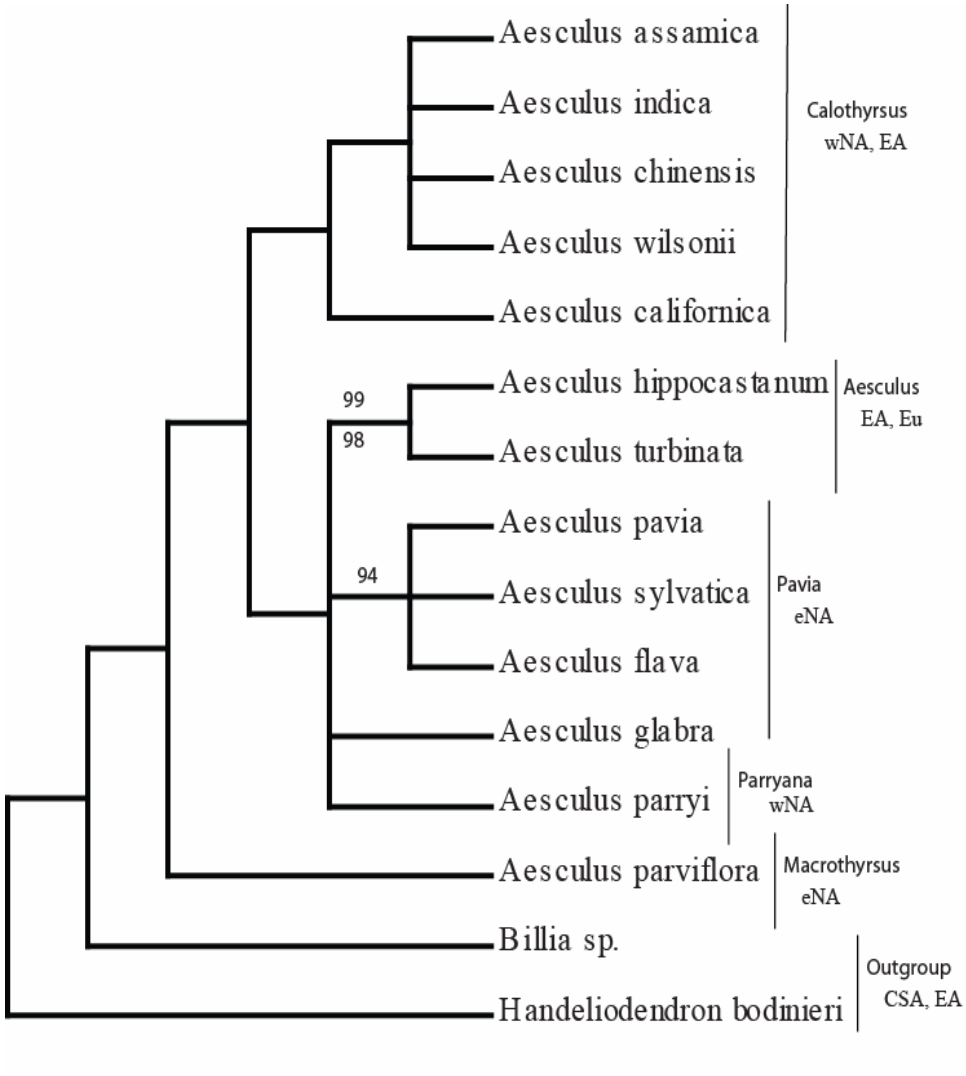

Fig 4: Strict consensus of 30 MP trees resulting from analysis of modified morphological matrix - Bootstrap values ≥ 70% are shown above branches, PP ≥ 90% is shown below. Section names and modern distributions are given to the right. Morphology matrix is modified from Forest et. al (2001), see text. wNA = western North America, eNA = eastern North America, EA = East Asia, Eu = Europe, CSA = Central and South America

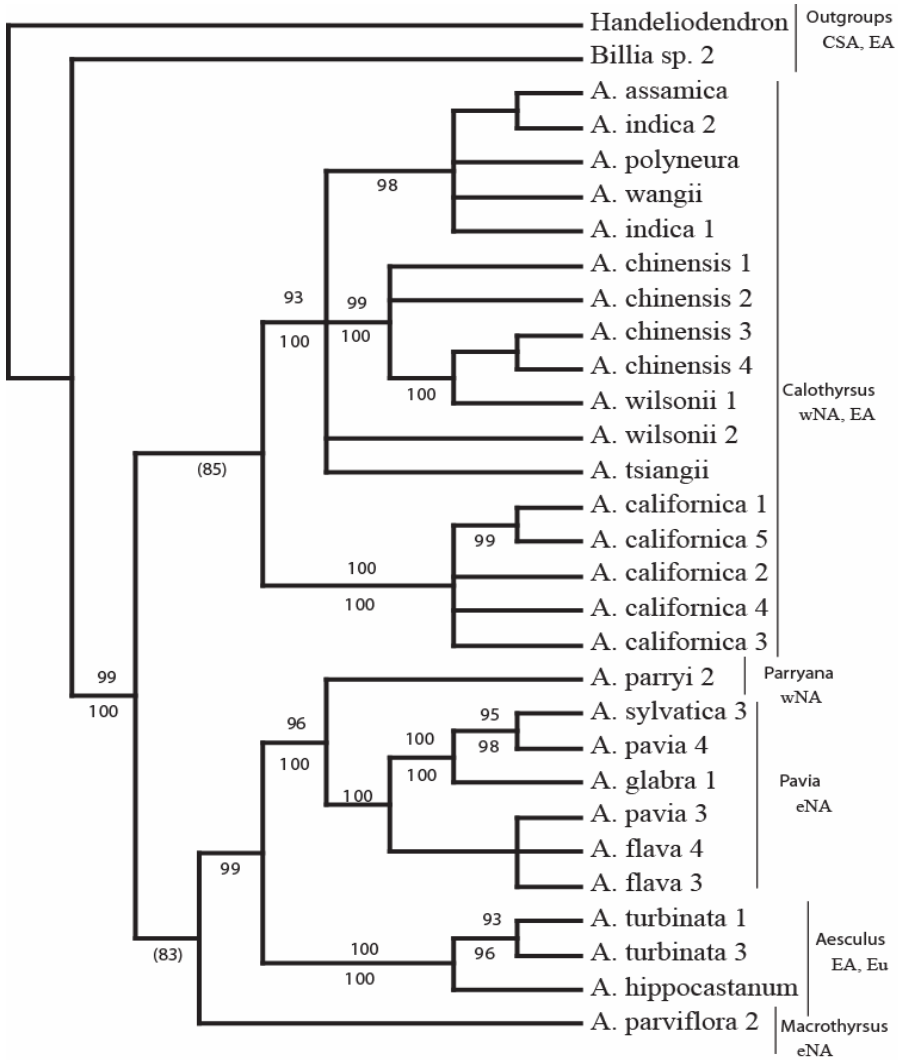

Fig. 5: Majority rule consensus of Bayesian cpDNA trees - Bayesian consensus topology of 19,800 MCMC sampled trees generated from analysis of the combined cpDNA matrix. PP $\geq$ 90% is shown below branches, bootstrap support from MP analysis $\geq$ 70% is shown below. Support for nodes less than these values but discussed in the text are shown in parentheses. wNA = western North America, eNA = eastern North America, EA = East Asia, Eu = Europe, CSA = Central and South America

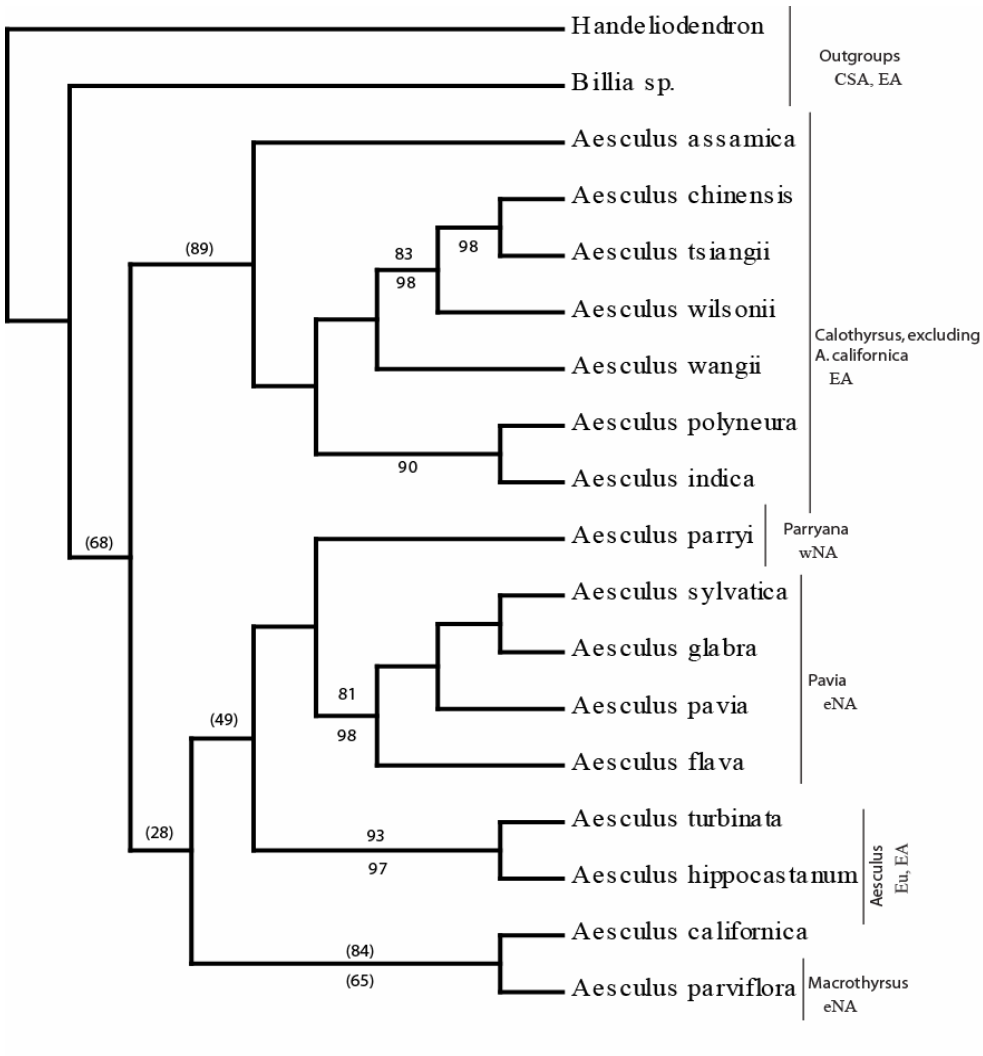

Fig. 6: Majority rule consensus of Bayesian trees from combined analysis of ITS and *LFY*: Bayesian consensus topology of 19,800 MCMC sampled trees generated from analysis of the combined analysis of ITS + *LFY*. PP $\geq$ 90% is below branches. bootstrap support $\geq$ 70% from MP analysis is shown above. Support for nodes less than these values but discussed in the text are shown in parentheses. wNA = western North America, eNA = eastern North America, EA = East Asia, Eu = Europe, CSA = Central and South America

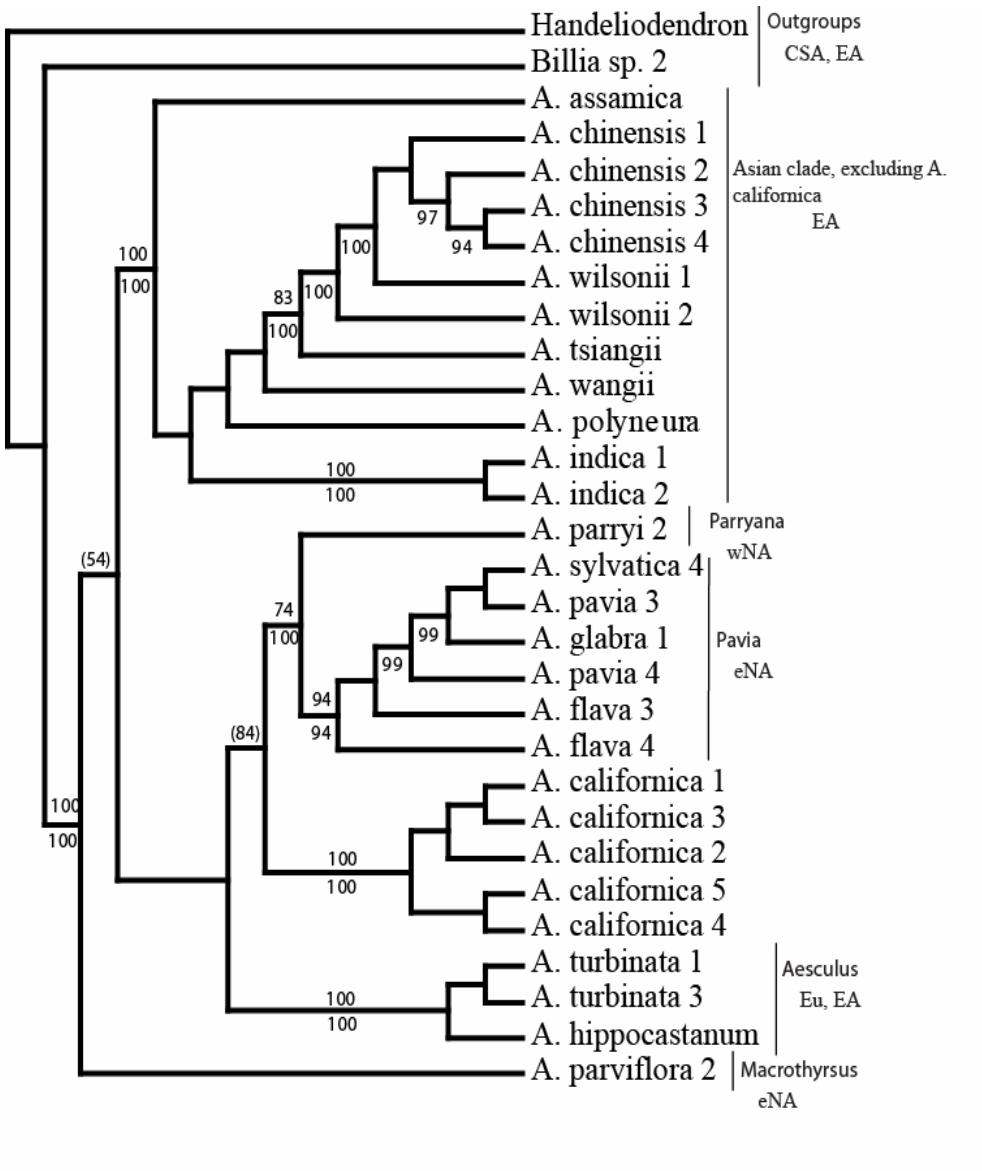

Fig. 7: Majority rule consensus of Bayesian trees from analysis of chloroplast and nuclear DNA: PP $\geq$ 90% is shown below branches, bootstrap support from MP analysis $\geq$ 70% is shown above. Support for nodes less than these values but discussed in the text are shown in parentheses. wNA = western North America, eNA = eastern North America, EA = East Asia, Eu = Europe, CSA = Central and South America

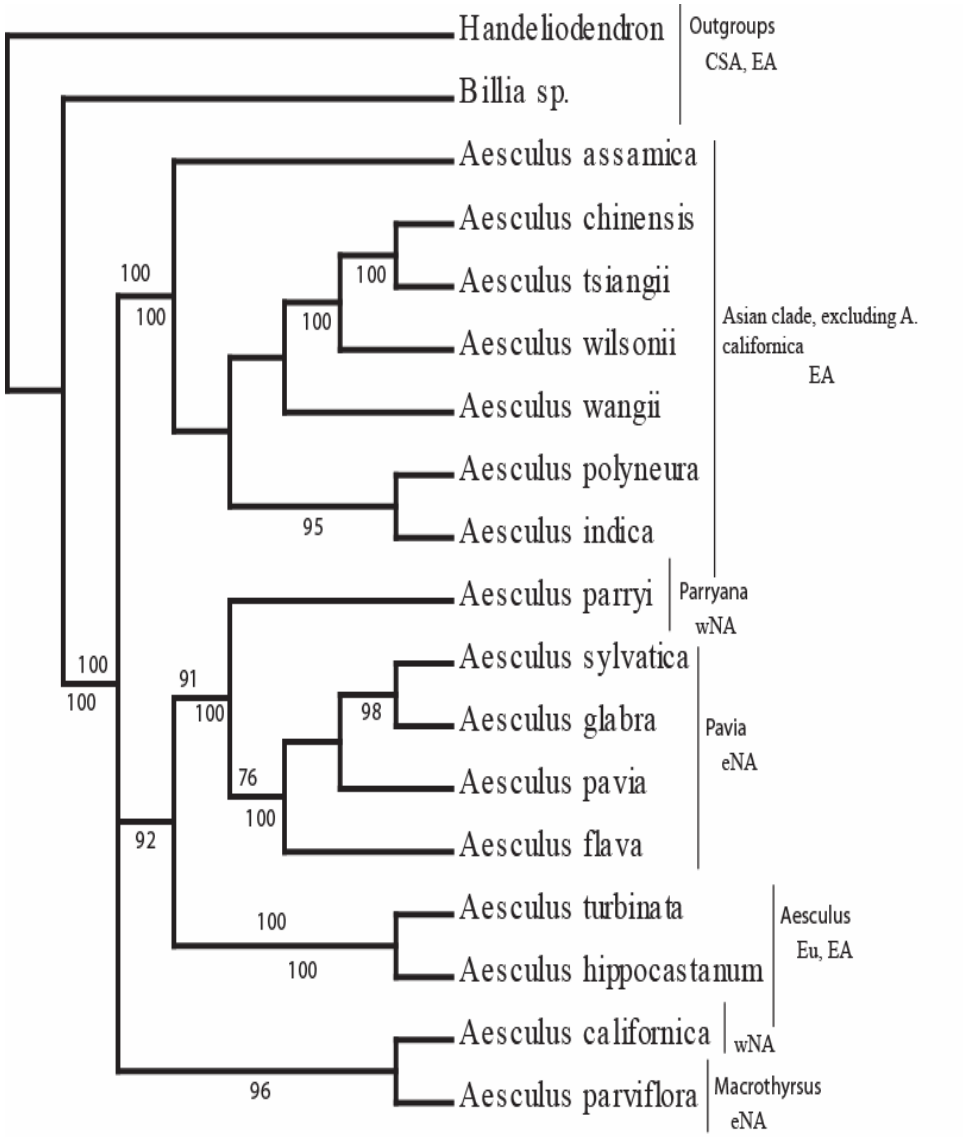
Handeliodendron
Billia sp.
Outgroups
CSA, EA
Aesculus assamica
Aesculus chinensis
Aesculus tsiangii
Aesculus wilsonii
Aesculus wangii
Aesculus polyneura
Aesculus indica
Asian clade, excluding A. californica
EA
Aesculus parryi
Parryana
wNA
Aesculus sylvatica
Aesculus glabra
Aesculus pavia
Aesculus flava
Pavia
eNA
Aesculus turbinata
Aesculus hippocastanum
Aesculus
Eu, EA
Aesculus californica
wNA
Aesculus parviflora
Macrothyrsus
eNA
100
100
100
100
100
95
91
100
98
76
100
92
100
100
100
96

Fig. 8a

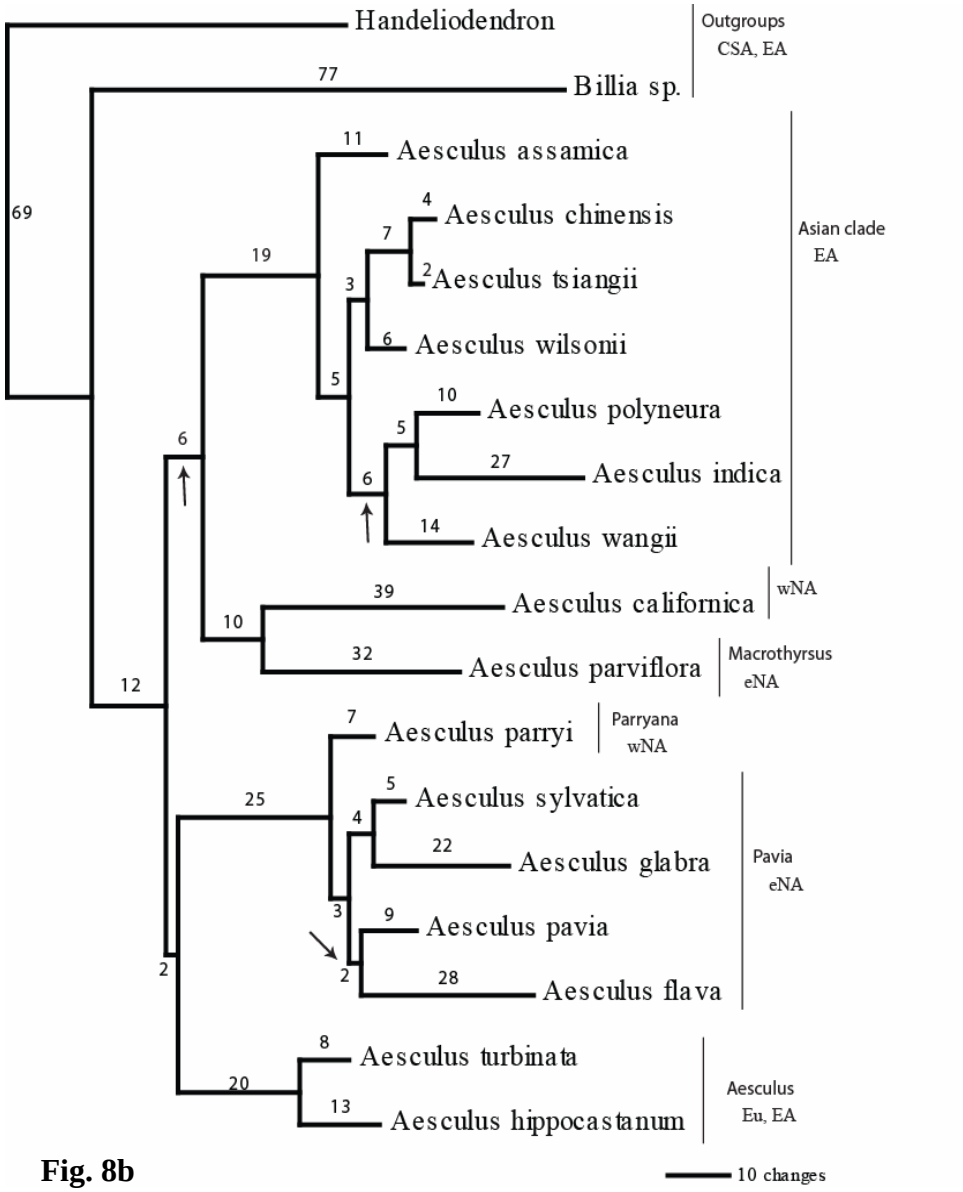

Fig. 8b

Fig. 8a-b: Majority rule consensus of Bayesian tree resulting from analysis of DNA + morphology, excluding fossils:. Section or clade affiliation is reported to the right of terminals as well as distributions of extant species. wNA = western North America, eNA = eastern North America, EA = East Asia, Eu = Europe, CSA = Central and South America. 8a - PP ≥ 90% is shown below branches, bootstrap support ≥ 70% from MP analysis is shown above. 8b – showing branch lengths using one of the MCMC sampled Bayesian trees from the analysis of DNA + morphology. Branch lengths, which include nucleotide and morphological characters, are shown adjacent to branches. Arrows indicate branches not present in the consensus.

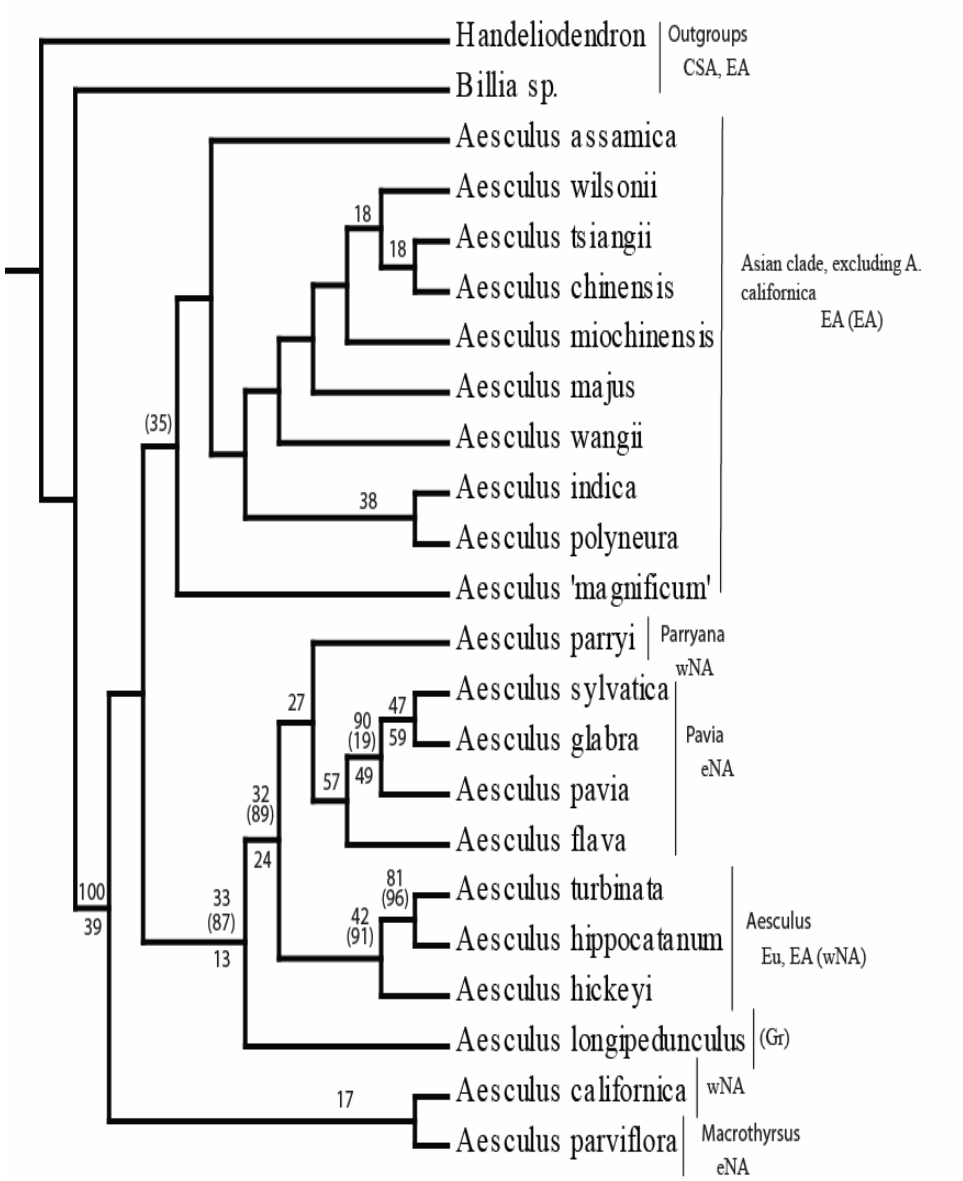

Fig. 9: Majority rule tree from MP analysis including fossils using the total evidence tree as a backbone constraint – Boot strap support is shown above branches. Support from majority rule consensus discussed in the text is shown parenthetically. PP support values for non-conflicting relationships is shown below branches. Section or clade affiliations and distributions of living taxa are shown to the right of terminals. Known localities of fossil taxa are given parenthetically.

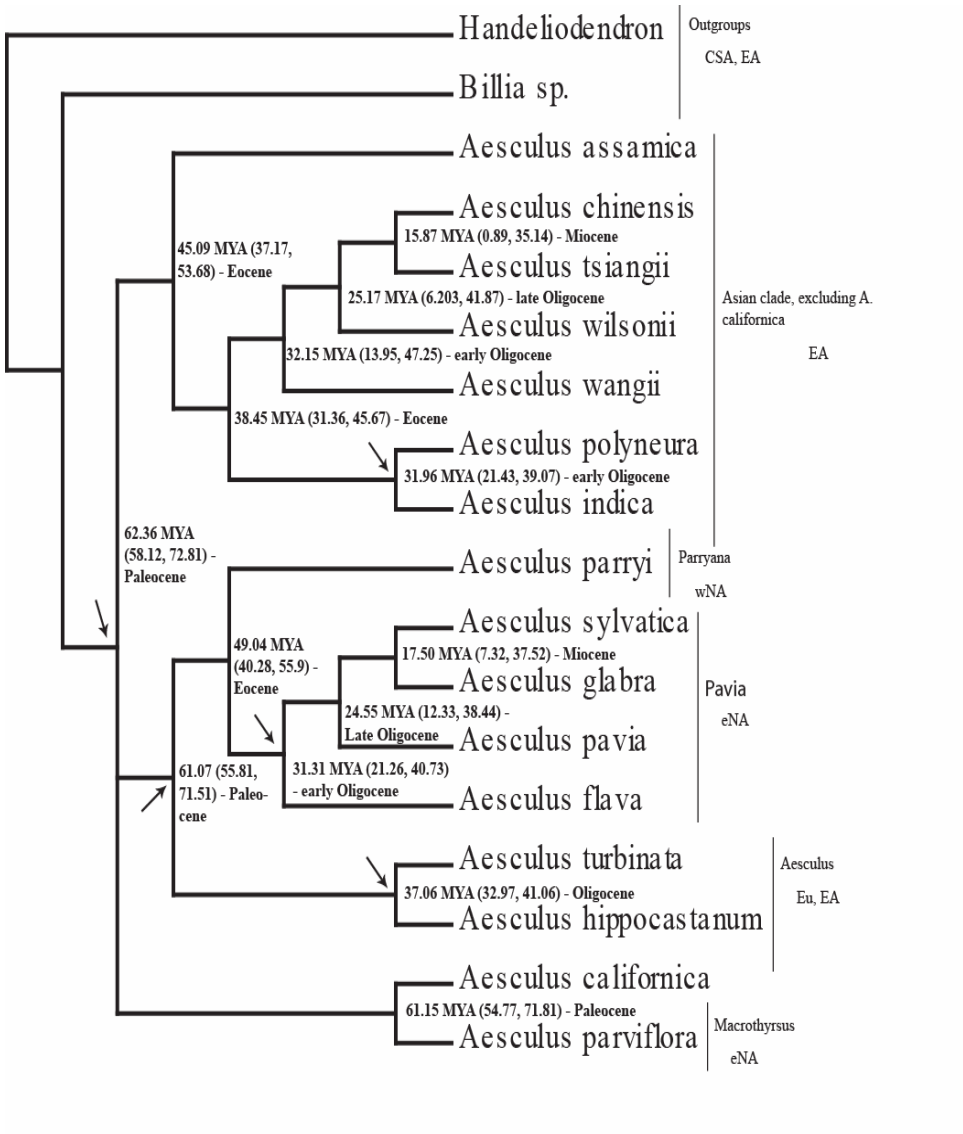

Fig. 10: Phylogeny of *Aesculus* showing divergence times of lineages – Divergence times are given in millions of years with 95% CIs in parentheses followed by the corresponding epoch. Arrows indicate nodes constrainted using fossil data (see text and Table 2). Section afffiliations are shown to the right of terminals with distributions of modern species.

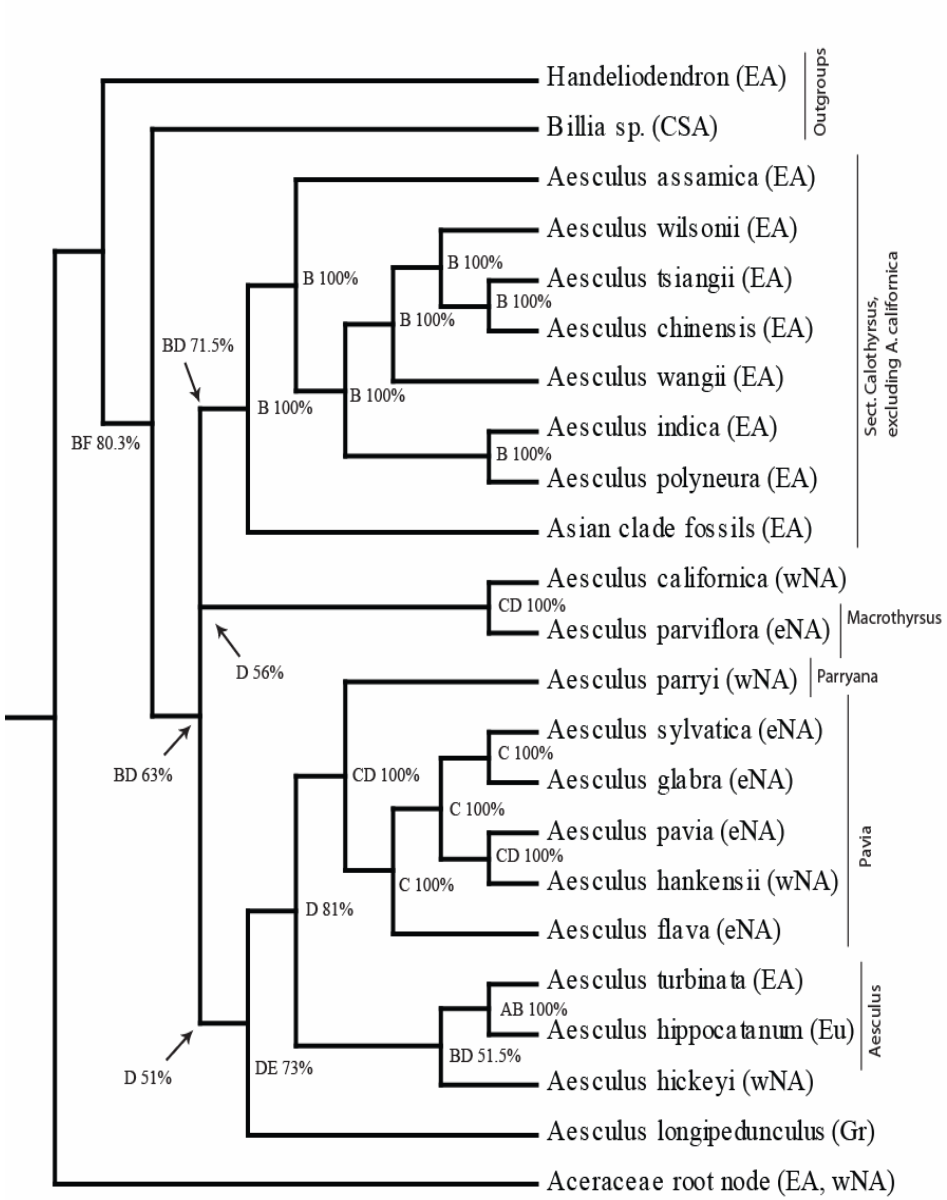

Fig. 11: Biogeographic reconstruction results from DIVA - Distributions of modern species and fossil localities are shown to the right of terminals. Section affiliations are given for each clade. Where fossils are nested within traditionally recognized groups, they are shown as belonging to those groups. The most probable ancestral ranges and associated probabilities are given at each node. See text for information on calculation at the tritomy. A = Eu (Europe), B = East Asia (EA), C = eNA (eastern North America), D = wNA (western North America), E = Gr (Greeland region), F = CSA (Central and South America).

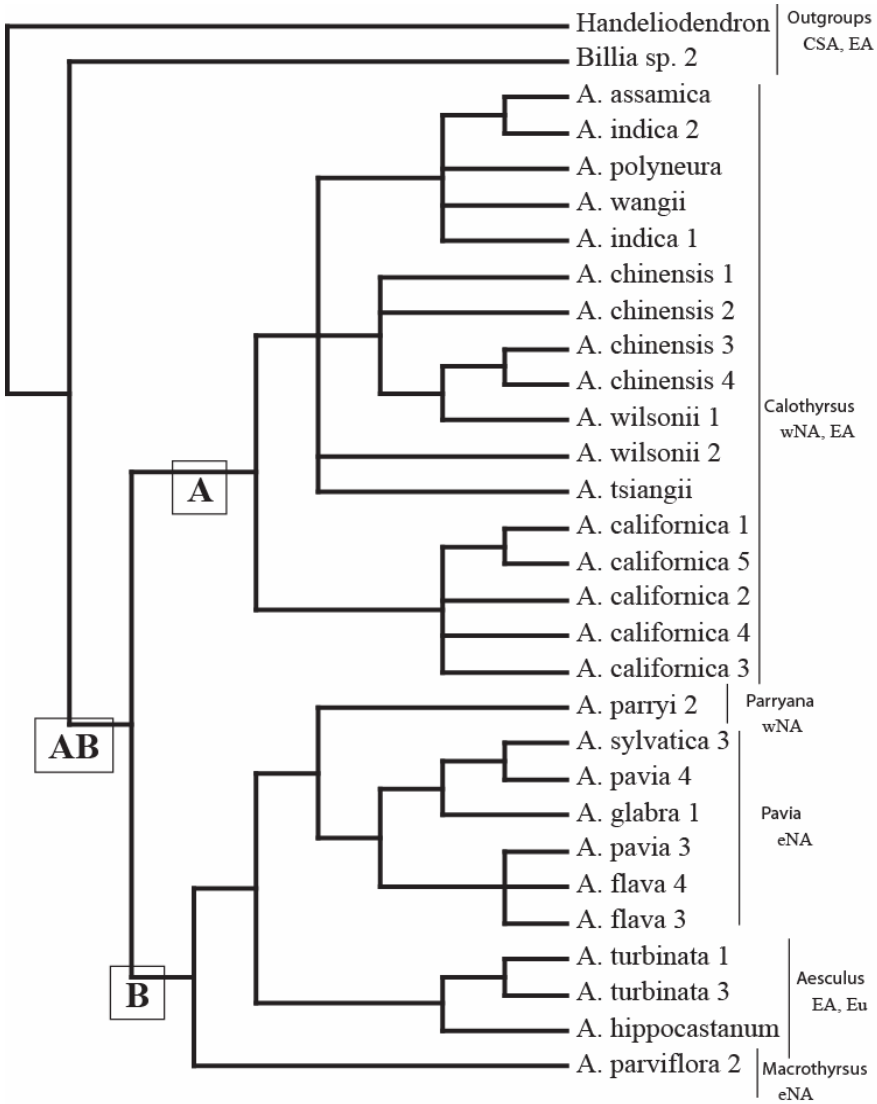
Handeliodendron
Billia sp. 2
A. assamica
A. indica 2
A. polyneura
A. wangii
A. indica 1
A. chinensis 1
A. chinensis 2
A. chinensis 3
A. chinensis 4
A. wilsonii 1
A. wilsonii 2
A. tsiangii
A. californica 1
A. californica 5
A. californica 2
A. californica 4
A. californica 3
A. parryi 2
A. sylvatica 3
A. pavia 4
A. glabra 1
A. pavia 3
A. flava 4
A. flava 3
A. turbinata 1
A. turbinata 3
A. hippocastanum
A. parviflora 2
A
AB
B
Outgroups
CSA, EA
Calothyrsus
wNA, EA
Parryana
wNA
Pavia
eNA
Aesculus
EA, Eu
Macrothyrsus
eNA

Fig. 12a

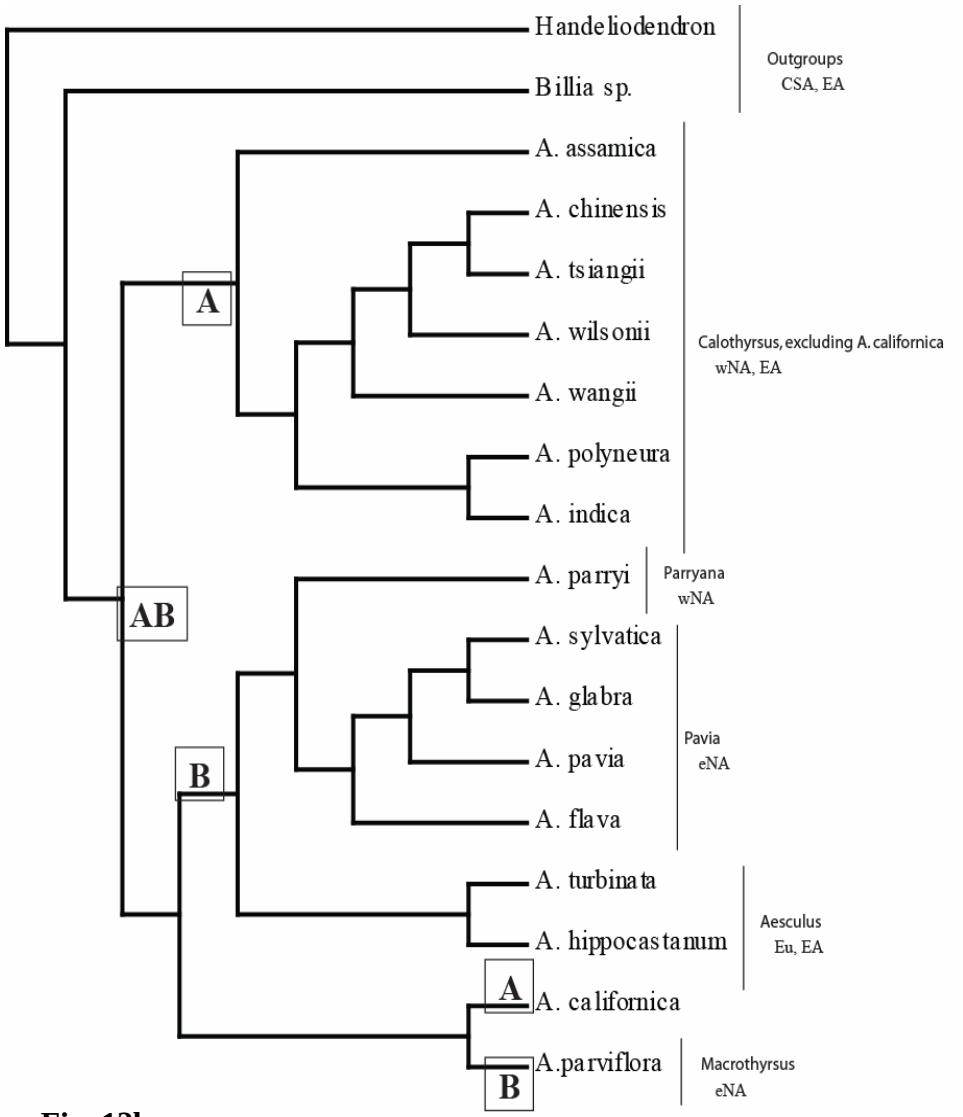

Fig. 12b

Fig. 12a-b: Plastid lineage sorting of haplotypes A and B: Section affiliation is reported to the right of terminals as well as distributions of extant species. wNA = western North America, eNA = eastern North America, EA = East Asia, Eu = Europe, CSA = Central and South America. Boxes overlap the branches with which they are associated. Letters within boxes refer to haplotypes. 8a – Plastid haplotypes shown on the cpDNA tree. Like haplotypes are grouped together. 8b – Plastid haplotypes are shown on the nuclear DNA tree.

APPENDICES

All literature referenced in the Chapter 1 appendices can be found in the Chapter 1 list of references.

APPENDIX 1

MORPHOLOGICAL CHARACTERS INCLUDED IN PHYLOGENETIC ANALYSIS IN ADDITION TO THOSE DESCRIBED BY FOREST ET. AL. (2001)

40. Secondary vein curvature: 0 – camptodromic, 1 – craspedromic

Craspedodromic secondaries run directly into the leaf margin. This condition common among taxa that have serrate leaves, as in *Aesculus* (Manchester pers. comm. 2007). Camptodromic secondaries diverge before the margin with a smaller vein running into the margin and a (typically) larger vein looping upward and merging with the next secondary.

While all *Aesculus* species have a tendency towards the camptodromic condition, craspedodromic veins are observed in some species such as *A. hippocastanum* and the constituent species of Sect. *Pavia*. Ideally, multiple leaflets are available for determination of this character state.

41. Intermediate secondary veins: 0 – absent or obscure, 1 – prominent

This character of *Aesculus* and *Billia* has been described and pictured by Hardin (1957a, see: page 152, and Fig. 2a-2f).

42. Pollen mesocolp structure: 0 – striate, 1 – spinulose

This character refers to the ornamentation of the pollen mesocolp surface. Spinulose refers t tapering pointed elements. Striate refers to elongate, usually parallel or nearly parallel ornamentation separated by grooves.

This character was scored using a morphology matrix included in Manchester (2001).

APPENDIX 2

ADDITIONAL DESCRIPTION OF AESCULUS FOSSILS INCLUDED IN THIS STUDY

Fossil remains of *A. hickeyi* found in Wyoming and North Dakota include palmately compound leaves, associated pollen, and associated fruits. Mature fruits are spiny, a character shared with two extant species; *A. glabra* of Sect. *Pavia* and *A. hippocastanum* of Sect. *Aesculus*. These prickles are the result of enlargement of the stalk of stiptate glands on the ovary wall. These enlarged gland stalks do not always develop into prickles at maturity as is the case with *A. parryi* (Hardin 1957b, Forest et al. 2001). The open locucidal capsule of one fossil fruit (Accession # UF 31059, Manchester 2001) shows a portion of a hilum that appears to cover more than 1/3 of the seed surface, a character common to Sect. *Aesculus* and the Asian clade. Impressions of palmately compound leaves show sessile leaflet attachment, an autoapomorphic feature of Sect. *Aesculus*. Secondary veins of *A. hickeyi* leaflets generally run directly into the serrulate leaflet margins (craspedromous venation), as in Sect. *Aesculus*, Sect. *Pavia*, and Sect. *Parryana* though all species of *Aesculus*, including the fossil, exhibit a tendency towards the camptodromous condition. These characters unite *A. hickeyi* most closely with Sect. *Aesculus*, supporting the result of phylogenetic analysis.

The fossil described from wood, *A. hankensii*, is known from the Columbia Basalts in an area near Vantage, Washington, USA. The authors report that the fossil flora has been considered to be Upper Miocene in age, but the taxa present suggest an earlier Miocene assemblage (Prakash and Barghoorn 1961). The fossil wood is diffuse and porous, a character shared by all *Aesculus* species. Prakash and Barghoorn (1961)

compared sections of this fossil by eye to sections from living *Aesculus* species. The authors report that *A. hankensii* differs from *A. pavia* only in the less frequent occurrence of vessel multiples and the more compact crowding of vessels in the early wood (Prakash and Barghoorn 1961). This fossil was included in the biogeographic analysis and as an age constraint on the divergence time of Sect. *Pavia* in divergence time dating.

Leaflet (Budantsev 1983, Schloemer-Jager 1958) and partial leaf material with attached leaflets (Golovneva 2000) of *A. longipedunculus* has been described from the Lower Eocene flora of Spitsbergen, an island in the Artic Ocean or far North Atlantic considered part of Europe (map in Golovneva 2000). Leaflets are petiolate, a condition shared by all members of the Asian clade, Sect. *Macrothyrsus* + *A. californica*, and Sect. *Pavia*. Craspedromous secondary veins characteristic of Sect. *Aesculus*, Sect. *Parryana*, and Sect. *Pavia*, are reported in *A. longipedunculus*. Bayesian and MP analysis resolve this fossil as sister to Sect. *Aesculus* + Sect. *Parryana* + Sect. *Pavia* indicating that the sessile condition is derived within Sect. *Aesculus*.

Budantsev (1983, Manchester 2001) described *Aesculus* fossils from the Late Paleocene/Lower Eocene of Kamchatka, in the far northeast of Russia. These fossils included a palmately compound leaf (Budantsev 1983, Manchester 2001) easily identifiable as *Aesculus*. Manchester reports that while Budantsev attributed this material to '*A.* magnificum' (Newberry) Iljanskaya, that the leaflets are petiolate and appear to have a closer affinity to extant East Asian species. Manchester reports (2001) that the '*A.* magnificum' type specimen (baisonym: *Hicoria magnifica*, Knowlton 1904) is probably Juglandaceous.

Aesculus majus was first described from leaflet material in an Upper Oligocene-Lower Miocene (~23.8 MYA) flora of the Nishitagawa coal basin in Yamagata Prefecture in Japan (Tanai 1952). The author synthesized his own findings with a previously reported fossil form; *Aesculiphyllum majus* Nahorst described in a Miocene flora of Akita Prefecture. Tanai believed that his fossil and the material described by Nahorst was most closely affine to *A. turbinata* and *A. hippocastanum*, the constituent species of Sect. *Aesculus*. His assessment may have been based, in part, on geography since *A. turbinata* is the only species of *Aesculus* endemic in Japan. However, Tanai (1952) describes lateral or secondary veins that "arquées en avant" or arc before [the margins]. This camptodromic condition is common to the constituents of the East Asian clade, to *A. californica*, and to Sect. *Macrothyrsus* but not to Sect. *Aesculus*. Tanai (1952) also reports that his leaflets are finely serrate, while mature leaflets of *A. turbinata* and *A. hippocastanum* are doubly serrate or nearly so (Hardin 1960). This character, however, as noted by Forest et al. (2001) can vary depending on leaflet age and position in the compound leaf. Although we have not examined Tanai's leaflet materials, the description suggests that the *A. majus* may have more of an affinity with the Asian clade constituents of Sect. *Calothyrsus* than with Sect. *Aesculus*. That a fossil with closer affinity to the Asian clade constituents of Sect. *Calothyrsus* than to the Japanese endemic, *A. turbinata* (Sect. *Aesculus*), is present in the Upper Oligocene or Lower Miocene of Japan is explained by timing of the formation of the Sea of Japan. The formation of the Sea of Japan, which separates Japan from the Asian mainland, began with crustal subsidence and basaltic volcanic activity around 21 MYA (Lower Miocene) and reached a peak of activity around 15-16 MYA (middle Miocene), when Japan began to turn in

counterclockwise rotation out into the Pacific (Horikoshi 1990). Until some time after 21 MYA, there does not appear to be a geographic barrier to floristic exchange between Japan and mainland Asia. *Aesculus majus* probably represents an Asian clade constituent that survived the opening of the Sea of Japan. *Aesculus majus* is also known from the Pliocene Kabutoiwa flora of central Honshu (Ozaki 1991), though the determination of the Kabutoiwa leaflets and their affinity remains unclear. Bayesian analysis supported an Asian clade including this fossil at the base of the clade (10% PP). Parsimony analysis failed to place this fossil due to insufficient characters.

Aesculus miochinensis was described from leaflet material as a constituent of the Miocene flora of Shantung Province in eastern-central China (Hu and Chaney 1940). The authors suggest that their fossil leaflets are most closely related to *A. chinensis*. Although we cannot confirm that association, parallel tertiaries and camptodrome secondaries seem to support an affinity of this fossil taxon with the Asian clade.

Fossil leaflets of *A. indica* were discovered in a Pliocene formation in the Kashmir region of northern India (Puri 1945); Manchester (pers. comm. 2007 at UF) concurred with this determination, suggesting that pictures in Puri (1945) appear to be *Aesculus* and that the affinities of Pliocene fossils are often most easily determined by geographic proximity to living species. Two living species of *Aesculus* are found in modern-day India; *A. indica* and *A. assamica*, though only *A. indica* currently has a range that includes Kashmir (Hardin 1960). Since this fossil represents the modern species, *A. indica*, it was used only in establishing a node age constraint for divergence time dating.

Several authors (Szafer 1947, Szafer 1954, Mai and Walther 1988, and deLumely 1988) have described fossil fruits of *A. hippocastanum* from the Pliocene and Pliestocene

of Europe. Prickles observed on the fruits, age of the fossils, and geographic proximity to the living species, *A. hippocastanum*, have been important factors in determination of these fossils. These fossil taxa appear to represent the modern species and share the same range. Thus, these fossils were used only for establishing node age constraints for divergence time dating.

APPENDIX 3

FOSSILS OF AESCULUS EXCLUDED FROM THIS STUDY OR FROM PHYLOGENETIC ANALYSIS

Two fossil species, *A. ashleyi* Axlerod (wNA, Miocene) and *A. montanus* (wNA, Eocene)Axlerod, described only from leaflets were excluded from phylogenetic analysis due to insufficient characters. Affinities of these fossil determined by the author were questionable (Erwin pers comm.), so these taxa were not included in Table 4 and could not reliably be placed on the tree topology using posterior mapping. Initially, the leaflet material from *A. paleocastanum* Ett. was included for phylogenetic analysis, but later excluded due to insufficient informative characters. Examination of MP trees showed that this fossil behaved as a wildcard (Nixon and Wheeler 1992, Kearney and Clark 2003). The author suggested no affinity for this fossil for use in posterior mapping. The fossil species *A. preglabra* Condit, known from leaflets, was excluded because published photographs of holotypes do not appear to represent the same species and may not all be *Aesculus*. The prologue of *A. preglabra* is a collective description of all holotypes.

The wood fossils *A. hankensii* Prakash & Barghoorn (1961) (western North America, Miocene), *A. mioxyla* Suzuki & Terada (1996) (East Asia, Miocene), and *Aesculus* sp. determined by Jeong et al. (2004) (East Asia, Miocene) were also excluded from phylogenetic analysis. To our knowledge, wood characters that distinguish living *Aesculus* taxa at the section or species level have not been thoroughly investigated and published. *Aesculus mioxyla* may belong to Sect. *Aesculus* (Suzuki and Terada 1996), based primarily on its occurrence in Japan (Suzuki and Terada 1996), where the modern species *A. turbinata* (Sect. *Aesculus*), a Japanese endemic, occurs. Since this fossil occurs

within the range of the modern species, excluding it from the biogeographic analysis is unlikely to affect the results. *Aesculus hankensii* is thought to be most closely related to Sect. *Pavia* (see Materials and Methods, Chapter 1). Since this fossil falls outside the modern range of Sect. *Pavia*, it is considered in biogeographic reconstruction and is placed by posterior mapping as sister to *A. pavia* as suggested by the authors (Prakash and Barghoorn 1961).

The fossil fruit species *A. spinossissima* Van Der Burgh from the Pliocene of Europe is distinguished from the extant species, *A. hippocastanum*, by smaller prickles on the fruit. This fossil taxon appears to represent *A. hippocastanum*, also known from the European Pliocene fossil record (See Table 4). Since its closest affinity is with *A. hippocastanum*, if it is a distinct species, and has the same range (Europe) as the modern species, it was unnecessary to include this taxon in biogeographic reconstruction.

APPENDIX 4

NOTES ON 'AESCULUS' FOSSIL NOMENCLATURE AND TAXONOMY

***Aesculus antiquorum* (Newberry) Iljinsk.**

1. nom. ambig.; applied to *Carya antiquorum* and *A. hickeyi*
2. Concept currently split into *A. hickeyi* Manchester and *Carya antiquorum* Newberry
3. synonym: *Hicoria antiquora* (Newberry) Knowlton

***Aesculus antiquus* Dawson**

1. Proposed new combination: *Spinifructus antiquus* (Dawson) McIver (McIver 2002)

***Aesculus ashleyi* Axlerod**

1. Affinity may not be as author described (Diane Erwin pers. comm. UC Berkeley, see also Axlerod 1956)
2. May be more closely related to *A. hickeyi* or *A. glabra* (The Paleobiology Database)

***Aesculus magnificum* (Knowlt.) Iljinsk.**

1. basionym: *Hicoria magnifica* Knowlton
2. May be *Carya antiquorum* Newberry (Hollick 1936, Manchester 2001)

CHAPTER 2

A STATISTICAL APPROACH TO INFERRING BIOGEOGRAPHY IN THE FACE OF PHYLOGENETIC UNCERTAINTY USING DIVA

ABSTRACT

Reconstructing the biogeographic history of a taxon and developing hypotheses regarding the events responsible for its current distribution typically requires a well-supported, bifurcating phylogenetic tree. Biogeographic reconstructions normally involve some method of mapping ancestral distribution areas as characters across all nodes of a single phylogenetic tree. In the real world of systematics, phylogenies are not always well-supported nor does the data always produce perfectly bifurcating trees. We propose that biogeographic information about taxa or lineages of interest need not be restricted to reconstruction of a single scenario across a tree topology. Here, biogeographic analysis using DIVA (Dispersal Vicariance Analysis) software is repeated for a set of randomly sampled Bayesian trees generated using MrBayes. We define a node as the hypothesized ancestor of a specific lineage and its unspecified sister (x). For each node, the probability of each possible ancestral range is calculated as follows: $[P_{T1} + \ldots P_{T100}] / N$, where P is the frequency of a range set estimated for a node and T*n* is one of N randomly sampled trees.Thus, ranges for each node are reported as a set of probabilities from which biogeographic hypotheses may be developed and tested. This approach obviates the need for use of a single tree topology with well-supported relationships between groups for biogeographic reconstruction. The utility of this approach was tested using *Aesculus* L.

INTRODUCTION

DIVA (Ronquist 1999) is a program designed to estimate the ancestral distribution of the taxon or lineage at each node of a given phylogenetic tree. The program requires that two parameters are defined: the phylogeny and the present distribution of terminal taxa. The latter is entered into a matrix (e.g. in PAUP or MacClade) and distribution areas are treated as character states. Unlike Bremer's (1992) Ancestral Area (AA) approach to ancestral biogeography, DIVA does not consider the size, shape or arrangement of landmasses. Instead, the program optimizes distributions at each ancestral node by minimizing extinctions and dispersals. Extinction or dispersal events (weight 1) are more costly than vicariance or sympatric or allopatric speciation events (weight 0). The parsimony criterion is used for optimization. Importantly, DIVA calculates the most parsimonious biogeographic scenarios across the entire tree topology. This often results in multiple most parsimonious scenarios. Among these multiple parsimonious scenarios, a single node may have different optimal ancestral distributions. A summary of optimizations produced by DIVA at each node shows all ancestral distributions suggested for that node that occur at least once in the set of most parsimonious scenarios.

Uncertainty in biogeographic reconstruction using DIVA arises from two sources. These are uncertainty inherent in a phylogenetic reconstruction and multiple optimal states for a given node. With a phylogeny weakly supported at some nodes or containing polytomies at the tree base, inference of biogeographic history using DIVA can be non-meaningful and infeasible. The potential mobility of a weakly supported branch across the tree topology causes obvious uncertainty in the results of almost any biogeographic

analysis that employs an analytical method reliant on a phylogeny. In DIVA, moving a branch can change the optimal ancestral distributions at one or many nodes. Stated another way, a different tree topology can result in different optimizations. Further, DIVA requires a fully resolved tree topology to run the analysis. Some studies take a best guess at unresolved relationships to produce a fully bifurcating tree for the analysis. This approach works fine if the multiple lineages in a polytomy have the same terminal distributions. However, if the terminal taxa in a polytomy do not all have the same distribution a different guess can generate quite different results (Fig. 1).

Summary DIVA optimizations may include one or more sets of character states (range sets) for a given node in the phylogeny. A set may consist of one or more distribution areas (character states). Nodes for which more than one range set is optimal introduce uncertainty. Although some range sets can be eliminated based on prior knowledge, such as physical locations of areas, availability of land routes for connecting areas, and fossil evidence, etc., others remain difficult to evaluate and may suggest widely disparate biogeographic histories. A tree with three ancestral nodes each with three optimal range sets equates to 27 different possible biogeographic histories for the taxa represented by that tree. Each scenario or hypothesis that cannot be eliminated or explained as unlikely is equally likely to be the true biogeographic history of the group.

Optimally, a single tree with well supported branches is submitted to DIVA with the goal of producing a single most parsimonious biogeographic scenario. As discussed above, this optimal situation is frequently confounded with phylogenetic uncertainty and multiple summary optimizations at ancestral nodes. Lutzoni et al. (2001) demonstrated the utility of considering ancestral character states as probabilities calculated by repeating

an ancestral character state reconstruction procedure over a set of phylogenetic trees generated in a Bayesian framework and sampled using a Markov chain Monte Carlo (MCMC). Here we propose applying this principle to biogeographic analysis using DIVA under the condition of phylogenetic uncertainty described above. The probability of an ancestral distribution of a lineage with uncertain placement on the phylogenetic tree can be estimated using DIVA. Probabilities are calculated from DIVA optimizations of a sub-set of MCMC sampled Bayesian tree topologies.

Optimized ancestral ranges at each node can be considered as biogeographic hypotheses with calculated probabilities. The probability of each hypothesis at each node can be calculated as follows assuming a sample of 100 trees: $[P_{T1} + \ldots P_{T100}] / 100$, where P is the frequency of a range set estimated for the node and T*n* is one of 100 randomly sampled trees. The probability (P) can be obtained using the printrecs command in DIVA or from the summary DIVA output. In the latter case, probability is calculated as follows: $[(1/S)_{T1} + \ldots (1/S)_{T100}] / 100$; where 1 indicates that a range set was proposed by DIVA and S is the number of optimal range sets for a node on a tree (T). Multistate range sets should be considered distinct from single state ranges: Range set [AB] is a hypothesis distinct from range set [B]. An example of this approach, using the $[(1/S)_{T1} + \ldots (1/S)_{T100}] / 100$ calculation of P, is given using *Aesculus* L. Table 1 provides an example of probability calculation using the $[(1/S)_{T1} + \ldots (1/S)_{Tn}] / N$ approach, where N is the total number of trees (T).

This approach seeks to minimize the two sources of uncertainty in DIVA. First, the utility of a Bayesian approach to handling uncertainty in phylogenetic and character state reconstruction is well reported (Huelsenbeck 2000, Lutzoni et al. 2001, Ronquist

2004). This method allows the researcher to sample from the set of possible topologies supported by the data. This is an alternative to selecting a single topology when the data supports several or many topologies that are in conflict. Second, multiple repetitions of DIVA over a representative range of Bayesian tree topologies can provided statistical support for a range set when multiple optimal range sets are estimated for a node. An optimized range set may emerge as more probable than others when the optimal ancestral range is reconstructed on a set of topologies rather than on a single topology. For example, A, B and C represent ancestral areas for a group of taxa and tree 1, 2 and 3 are randomly sampled Bayesian trees. For node Y, three range sets are estimated for tree 1 [(A), (AB), (C)], two range sets are estimated for tree 2 [(A), (AC)], and two range sets are estimated for tree 3 [(A) (BC)]. Clearly the range set (A) is more likely than (AC), (BC), or (C) when all three trees are considered.

Preliminary analyses of *Aesculus* sequence data for the molecular markers matK, ITS, and rps16 showed strong support for the monophyly of the polytypic sections (excluding *A. californica* from Sect. *Calothyrsus,* see Chapter 1), lack of resolution of deeper nodes, namely the relationships among the six sections. Thus, the utility of DIVA, applied in the traditional way, for biogeographic reconstruction of *Aesculus* was limited. Here, the ancestral range of each section of *Aesculus* was explored using the proposed probabilistic method of applying DIVA. The goal was to determine the utility of this method of biogeographic reconstruction in the circumstance of highly supported crown groups and deep node polytomies.

Nodes of interest in the *Aesculus* phylogeny to which the method described will be applied are (1) the last shared ancestor of each section with its sister section or

sections and, (2) for polytypic sections, the last shared ancestor of the taxa within the section (Fig 2). In case 1, nodes are defined as a hypothetical common ancestor shared by the section of interest and *x* where *x* may be any group of taxa represented on the tree. The node need not define an ancestor shared by the section of interest and some specific group. Thus in a random sample of Bayesian trees, *x* may vary from tree to tree but the section of interest will always have a node shared with some *x*.

In a separate analysis (R3, below), we also test the utility of this approach in determining the impact of the addition of wildcard taxa on biogeographic analysis. Wildcard taxa (Nixon and Wheeler 1992, Kearney and Clark 2003) are defined as those taxa that, due to significant missing characters, may be placed algorithmically at many or all nodes on the tree topology. In the case of *Aesculus*, the wildcard taxa are fossils. Thus, the results may be applicable to other biogeographic analyses which include fossils in phylogenetic analysis.

MATERIALS AND METHODS

Generating data matrices

All 13 species recognized by Hardin (1957a, 1957b, 1960) in his monograph of the Hippocastanaceae were included as well as material from *A. polyneura*, *A. tsiangii*, and *A. wangii*. The morphological data set used in run 1 and run 2 (see below) was a modification of the 39-character data set published by Forest et al. (2001). Modifications: (1) Combined the scoring of *Billia hippocastanum* Peyr. and *B. columbiana* Planchon & Linden into a single taxon, *Billia* sp., (2) Removed *A. glabra* var. *arguta* (Buckl.)

Robinson, and all outgroup taxa except *H. bondinieri* from the data set, (3) Scored two of the six newly described Chinese species – *A. tsiangii* Hu & Fang and *A. wangii* Hu – from the Latin prologues (Hu & Fang 1960, Hu 1960; respectively). These changes differ slightly from changes to the same data set applied in Chapter 1. The alterations to the morphological data set described in Chapter 1 represent additional data not available at the time of this study. In run 3 (see below), the morphological matrix from Chapter 1 was used. DNA sequence data obtained for reconstructing the phylogeny of *Aesculus* (Chapter 1) was used for this study. See Table 1, Chapter 1. All references to ambiguously aligned regions below refer to regions fitting the operational definition discussed in Miadlikowska et al. (2003).

DNA matrix for Run 1 (R1): DNA sequence data included the markers ITS, *mat*K, and the *rps*16 intron Simple gap coding (presence/absence) and ARC software (Kauff et al. 2003, Miadlikowska et al. 2003) were used to code gaps and ambiguously aligned regions in *mat*K. Ambiguously aligned regions and gaps in ITS and *rps*16 were neither coded nor excluded.

DNA matrix for Run 2 (R2): DNA sequence data included the markers ITS, *mat*K, and the *rps*16 intron Simple gap coding (presence/absence) and ARC software (Kauff et al. 2003, Miadlikowska et al. 2003) were used to code gaps and ambiguously aligned regions all three markers.

DNA matrix for Run 3 (R3): DNA sequence data included the markers ITS, *mat*K, the *rps*16 intron, *trn*HK, and *LFY*. The R3 matrix is identical to the DNA + morphology, including fossils matrix described in Chapter 1, Materials and Methods. In

this matrix, gaps and ambiguously aligned regions were uncoded. Gaps were treated as missing data.

Phylogenetic analysis

For each molecular marker, ModelTest 3.0 (Posada & Crandall 1998) was used to determine the best model of evolution. Since recoding gaps and ambiguous regions using standard states (output of ARC) caused nucleotide sequences to vary between R1 and R2, ModelTest, ModelTest was re-run for all markers in R2. Parameters proposed by ModelTest were set as priors in MrBayes. Gap and ambiguous region coding was included in R1 for *mat*K and for all molecular markers in R2. For each data set the Markov Chain ran for 22 million generations in two simultaneous runs. The stop rule was set to 0.001 and was in effect for all data sets but was never implemented. Split frequencies between chains were ~.002 for each data set after 22 million generations. Three hot chains and a single cold chain were employed for each run to ensure mixing. Trees were sampled every 2000 generations. Burnin was set to 2.2 million generations or 1100 trees. The 19,800 remaining trees were summarized using a 50% majority rule consensus in PAUP*4.0 (Swofford 2002). The data for R3 is the same as that collected from the Bayesian analysis of the DNA + morphology matrix, including fossils, described in Chapter 1.

To test for disagreement between individual makers and the morphological data, independent Bayesian MCMC analyses for each matrix were run using the conditions in MrBayes described above. Bootstrap values for these trees were generated in PAUP*4.0 using parsimony bootstrapping. Bootstrap analyses were done using the following

settings in PAUP: 1000 addition sequence replicates, heuristic search, gaps treated as a 5th base, gap and ambiguous region coding included, characters unweighted and unordered. Branches on the Bayesian consensus trees with less than 70% bootstrap support were considered mobile across the tree topology though not mobile into groups with a branch supported by >70%. Disagreement was defined as two trees having greater >70% support for a conflicting topology. Since the goal is to explore the biogeographic history of the six sections, disagreement within sections was not considered.

Data sets, including gap and ambiguous region coding were combined. Bayesian analyses of the R1 and R2 combined data were done using settings in MrBayes as described above. The final split frequency of the two simultaneous runs was 0.00274 in R1, 0.083559 in R2, and 0.00715 in R3. To test for accuracy of the Bayesian phylogeny, the consensus trees from R1, R2, and R3 were compared independently to parsimony trees for the same data sets generated in PAUP*4.0 (1000 addition sequences, heuristic search, gaps treated as a 5th base, gap and ambiguous region coding included, characters unweighted and unordered). Parsimony bootstrapping did not support disagreement between the MP and Bayesian trees for R1, R2, or R3 (considered independently of one another) based on the same criterion used to determine conflict between individual markers: >70% bootstrap support for a group that could not exist on the other tree regardless of mobility of branches into or out of groups with <70% support.

Random sampling of Bayesian trees

A sub-set of one hundred trees was randomly sampled independently from each set of tres from the analyses of the R1, R2, and R3 matrices. These 100 trees were

sampled from the 19,800 MCMC Bayesian trees sampled in each analysis. Random sampling was done using RandomTree (Kauff).

DIVA *analysis*

Extant taxa were scored as present (1) or absent (0) in each of five large and geographically isolated ranges: Europe (A), Asia (B), eastern North America (C), western North America (D), and South and Central America (E). Range E was included to accommodate *Billia* sp., which is found only in Central and South America. Each of the 100 randomly selected Bayesian trees were optimized in DIVA using the default settings of the program. Maxareas was not enforced.

Range sets may appear once or numerous times in the DIVA optimized scenarios. This information is available via the "printrec" command in DIVA. The advantage of this approach is that it can be used to calculate the exact probability (P) of each range set for nodes on each tree (T). These probabilities are then used to calculate the total probability over the entire sub-set of randomly sampled trees using $[P_{T1} + \ldots P_{T100}] / N$, where N is the number of trees in the sub-set. However, calculation of probabilities using the printrec data may result in calculations that are undesirably dependent on estimated distributions at deep nodes. The goal here is to reconstruct biogeography in cases where deep nodes are unsupported or poorly supported and where these deep nodes are not considered nodes of interest. For this reason, we have opted to use the $[(1/S)_{T1} + \ldots (1/S)_{T100}] / 100$ approach to probability calculation. This method of calculation seeks to minimize the influence of deep node estimations on the results. Of course, no method of biogeographic

reconstruction using phylogeny, such as DIVA, can entirely dissociate a single speciation event or single hypothesized ancestor from the rest of the tree.

Polytypic groups of interest may not be supported by 100% PP, such as Sect. *Macrothyrsus* + *A. californica* in R3. Thus, a node defining the group's shared ancestor may not appear in every randomly sampled Bayesian tree. For sampled topologies in which the node of interest did not appear, all range sets should be scored as having a probability of P = 0. Done in this way, remaining uncertainty regarding the monophyly of the group is inherient in the resulting range probabilities

For each node of interest, the probability of each possible range set was summed for all 100 trees and divided by 100 (using the equation given above). In R3, Sect. *Macrothyrsus* was not treated independently. Instead, Sect. *Macrothyrsus* + *A. californica* were treated as a monophyletic, polytypic group. Nodes of interest for this group included the last shared ancestor before the divergence of the two lineages and the last shared ancestor of the group and its sister (*x*). Record keeping and calculation was done by hand an in Microsoft Excel (example provided in Table 1).

Testing the impact of a fossil wildcard on biogeographic analysis using this approach to DIVA

Statistical analysis to show the impact of inclusion of the EA wildcard fossil *A.* 'magnificum' (Budantsev 1983, Manchester 2001) and the Eu fossil *A. longipedunculus* was done using a two tailed z-test comparing population proportions at a 0.95 level of significance. This test was used to determine if there was significant change in the probability of optimal range sets including Europe or EA between R1 (excluding fossils)

and R3 (including fossils) when fossils were included in the Bayesian analysis. Probabilities for each possible range set were calculated as above. The probabilities for all range sets which included Europe were summed for each run, R1 and R3. This was repeated for range sets including EA. The resulting summed probabilities for R3 were compared to those from R1, independently for EA and Europe, using the statistical test.

RESULTS

Results presented here are not given to propose a particular biogeographic history of *Aesculus* because the phylogenetic analyses in R1 and R2 were conducted using partial data available for *Aesculus*. Results that follow are intended to illustrate our statistical approach to use of DIVA software.

R1 and R2 results

The Bayesian consensus tree topologies resulting from the combined data set runs in R1 and R2 were not in disagreement with one another in the arrangement of sections. There was some incongruence that received low support, less than the conflict criterion. (Figs. 3a,b). The following results, unless otherwise noted, are data collected from R1.

DIVA optimizations of the ancestral range of the last common ancestor of the species in polytypic sections– Sect. *Aesculus*, Sect. *Pavia*, and the Asian clade – were resilient to the mobility of sections over the tree topology. For each section, the optimal distribution of the last common ancestor shared by the members of the section was a single range set with a probability of 1. The last common ancestor of Sect. *Pavia* is 100% likely to have been limited to eastern North America according to this method. The last

shared ancestor of the Asian clade grew in Asia. The last common ancestor of Sect. *Aesculus* was widespread across Asia and Europe (Table 2).

The optimal distributions of the last shared ancestor of each of the six groups with its sister (*x*) were much more sensitive to branch hopping. However, in all groups except Sect. *Macrothyrsus* a range set with probability >0.5 emerged. Figures 4a-f graphically summarize the distribution probabilities of the last shared ancestor of each section with its unspecified sister (*x*). Table 2 lists the three most likely ancestral distributions with probability values.

Before divergence of Sect. *Parryana*, an ancestor of this section and some group (*x*) was distributed across eastern and western North America (P=0.90). The Bayesian consensus tree supports (70% PP, Fig. 3a) a sister group relationship of *A. parryi* and Sect. *Pavia* that had an eastern North American distribution (P=1.0). The last shared ancestor of the Asian clade was East Asian (P=1.0). The last shared ancestor between this group and its sister (*x*) was also suggested to have an East Asian distribution (Fig. 2, Table 2). The ancestral distributions of this clade inferred here are consistent with the hypothesis of Xiang et al. (1998) supports this hypothesis. Using fossil evidence and ancestral state reconstruction, Xiang and colleagues proposed that the *Aesculus* genus evolved in the high latitudes of East Asia in the Paleocene and that the Asian clade lineage (Sect. *Calothyrsus* in Xiang et al.) has been present in Asia since that time. The last shared ancestor of *A. californica* and some group (*x*) had a range spanning East Asia and western North America (P=1.0). Thus, the current limited range of *A. californica* in western North America is best explained by a vicariant event disrupting the ancestral range, followed by either speciation of *A. californica* in western North America and *x* in

Eeast Asia or by extinction of *A. californica* in East Asia. The last ancestor shared by Sect. *Aesculus* and some group (*x*) had an East Asian distribution (P=0.832), while the last shared ancestor of the taxa in Sect. *Aesculus* had a range across Europe and Asia (P=1.0). Thus, the distribution of the extant species in this section – *A. turbinata* (Japan) and *A. hippocastanum* (Europe) – is explained by a vicariant event that isolated the ancestor in Asia and Europe oand caused the divergence of *A. turbinata* and *A. hippocastanum*.

The probabilities of the optimal ancestral ranges of the last shared ancestor of Sect. *Macrothyrsus* and some sister (*x*) do not provide support above 50% for any ancestral range for this section. However, summing the probabilities of the three most likely ranges – eastern North America (P=0.245), eastern North America/western North America (0.395), and eastern North America/western North America/Asia (P=0.119) – results in a probability of 0.759. The ancestral range Asia/eastern North America (P=0.193) was excluded since this range would require a conneting area such as western North America or Europe. These three ranges are not in conflict with one another. The high probability of the three ranges when combined suggests that the ancestral range of Sect. *Macrothyrsus* is a question of the extent to which the ancestor was distributed; widespread or more limited.

Probabilities calculated from R2 (data not shown), were generally similar to those calculated using trees from R1. Results differed primarily in support for the placement of *A. californica* as sister to Sect. *Macrothyrsus*. This sister group relationship received moderate support (86% PP, Fig. 3b) in the R2 analysis. The single range set (CD: eastern North America – western North America) associated with the *A. californica* – Sect.

Macrothyrsus sister group relationship emerged with the highest (0.914) probability as the ancestral distributions for both of these groups.

R3 results

Results from R3 are generally similar to those found in R! and R2 except that almost all range sets were supported by lower proability values (Table 3). Section Pavia appeared resilient to the addition of fossils and retained a highly supported ancestral range of eastern North America (P = 1.00). Section *Aesculus* was shown to diverge from its unspecified sister from an ancestor with a Eu range (P = 0.150). This differs from the highly supported ancestral range of East Asia (P=0.832, Table 2) found in R1. show that the last shared ancestor of all Asian clade constituents has a high probability, The ancestral range of the Sect. *Macrothyrsus* + *A. californica* lineage, not considered as a group in R1, and its sister shows the highest probability of having a range covering all or part of Eu, wNA, and eNA (P = 0.20). The shared ancestor of Sect. *Macrothyrsus* and *A. californica*, which diverged to form these two lineages, had an eNA-wNA distribution (P = 0.840). Section *Parryana* was shown to have diverged from a common ancestor with a wNA-eNA distribution (P = 0.770), a 13% drop in support from R1 for the same range. An infrequent placement of *A. hickeyi* as sister to Sect. *Parryana* observed in the sub-set of trees may account almost entirely for the 13% support for a wNA range of the ancestor of Sect. *Parryana* and its sister. The highest three ancestral distribution probabilities for each node of interest are shown in Table 3.

Impacts of wildcard taxa

In R3, the fossil species *A. hickeyi* was included in the analysis in addition to *A. longipedunculus* and *A.* 'magnificum' (Budantsev 1983, Manchester 2001). However, the *A. hickeyi* fossil does not appear to fit the operational definition of a fossil wildcard. In 49% of the 19,800 MCMC sampled Bayesian trees, this fossil was resolved as sister to Sect. *Aesculus*. In the sub-set of 100 randomly sampled Bayesian trees, it was observed that alternative placements included placement as sister to all *Aesculus* (33%), sometimes with *A. longipedunculus*, and as sister to the monotypic Sect. *Parryana* (17%). Four percent of the sub-set of trees showed other alternative placements, each occurring only once (1%).

Although the most probable placement of the European fossil, *A. longipedunculus*, is in a position sister to Sect. *Pavia* + Sect. *Parryana* (discussed in Chapter 1), remaining uncertainty regarding the placement of this fossil appears to significantly increase the probability for some range including Europe for ancestors of all sections and their unspecified sister groups (*x*) except Sect. *Parryana* (Table 4). The increase in the probability for an ancestor with a European distribution may be due, in part, to the changes in the data set used (inclusion of additional molecular markers, and coding gaps as missing data) for the R3 analysis compared to that used for R1 and R2. Phylogenetic analysis of the R3 data set shows that *A. hippocastanum*, the only extant European species, forms a monophyletic group with *A. turbinata* (EA) supported by 100% PP (Chapter 1). This relationship similarly reconstructed and supported in R1 and R2. The influence of *A. hippocastanum* on the increase in probabilities of European ancestral ranges for each section may be assumed to be minimal, leaving *A.*

longipedunculus to account for most of the increase. The greatest increase in the probability of some range including Europe was for Sect. *Aesculus* (ΔP=0.512), while the smallest increase was observed for Sect. *Parryana* (ΔP=0.32).

The percent change related to the introduction of the Asian wildcard fossil *A.* 'magnificum' (Budantsev 1983, Manchester 2001) is best measured by considering the impact of the addition of this fossil on the ancestral range of Sect. *Macrothyrsus* + *A. californica* and some unspecified sister (*x*), Sect. *Parryana* and some unspecified sister, and Sect. *Pavia* and some unspecified sister. This is because we already expect that there is a high probability that the Asian clade and its unspecified shared sister have an East Asian distribution (P= 100%, R1) and we expect the fossil *A. hickeyi* to have a greater impact than the Asian wildcard on Sect. *Aesculus* due to support (49% PP, Chapter 1) for its placement in that section.

The Asian wildcard was scored for two characters: Petiolule - Absent/Present, (absent), and Leaf Margin - Entire/Serrate (serrate). The petiolule character links this taxon with Sect. *Pavia* and the Asian clade. The leaf margin character links this taxon with all species currently classified within *Aesculus* as well as with the fossils included in this analysis. A visual survey of the randomly sampled trees used for DIVA analysis shows that the Asian wildcard was sometimes included within Sect. *Pavia*; a result of the petiolule character. However, strong support for the sister-group relationship between Sect. *Pavia* and Sect. *Parryana*, results in strong support for an eNA – wNA distribution of the Sect. *Pavia* ancestor. If *A.* 'magnificum' is nested within Sect. *Pavia*, observed results in DIVA suggest that the presence of A. 'magnificum' in East Asia is explained by later dispersal.

The most dramatic increase in probability of some ancestral range including East Asia, compared to results from R1, was seen for the Sect. *Macrothyrsus* + *A. californica* group (ΔP=0.49). There remains some uncertainty as to the monophyly of this group. It receives only 96% support in the DNA + morphology analysis. As with the well-documented and strongly supported crown groups (100% PP for all other sections), there was no support above 50% PP for the placement of Sect. *Macrothyrsus* + *A. californica* with respect to other sections. The movement of fossil taxa, including the wildcard, within and around this group greatly impact its probable evolutionary origins (Table 4).

DISCUSSION

Can the probabilities be inferred directly from Bayesian posterior probabilities?

One might wonder what relationship the ancestral area probabilities have to PP values and whether the probabilities can be inferred from PP support for nodes using the Bayesian consensus topology. In general, resulting range probabilities for each group and its sister (*x*) are slightly higher than the posterior probability values supporting the position of each section (Fig. 3, Table 2), though not all resulting probabilities differed significantly from the posterior probability values. For example, in the phylogenetic analysis Sect. *Aesculus* was resolved as sister to the Asian clade with a PP of 76%. The most likely range of the ancestor of Sect. *Aesculus* is Asia (P = 0.832, significant at 0.99 level). The fact that the probability of an Asian ancestor is higher than the occurrence of a specific relationship between Sect. *Aesculus* and another group or section means that the ancestral range probabilities cannot be explained entirely by the occurrence of a node in

the set of Bayesian posterior probabilities. Other possible topologies add support for an Asian ancestor in DIVA analysis. Sect. *Parryana* shows the greatest difference between posterior probability (70%) and probability of the most likely ancestral range (P = 0.90, significant at 0.99 level). A comparison of posterior probability values and most likely ancestral range probabilities is made in Table 2. It is important to recall that the goal is to generate a set of hypotheses with associated probabilities that reflect the uncertainty in the data. Thus, while the ancestral distribution with the highest probability value is of interest, alternative ancestral range probabilities should not be ignored or go unreported.

Probable range sets for groups supported by 100% PP, such Sect. *Pavia* (Fig. 3 a,b), appear resilient to the mobility of these groups over the tree topology. Thus, the ancestral area of a group receiving 100% PP support should be estimated using DIVA only once. In our results for R1, each group supported by 100% PP had a single optimal range set estimated by DIVA (data not shown). Thus, ancestral distributions are these nodes received 100% probability support. Theoretically, DIVA may estimate more than one optimal range from a single tree for a node with 100% PP support. In this case, the probability of each distribution would be calculated using $[P_{T1}] / 1$.

Are the sampled trees representative of the entire pool

The estimated area probabilities based on the 100 randomly sampled trees may be biased due to sampling error. In other words, trees with certain topologies may be sampled more frequently than their frequency in the pool of MCMC sampled Bayesian trees. Conversely, some tree topologies may be sampled less frequently than their frequency in the MCMC sampled pool. This over or under sampling of topologies would

clearly influence the probability calculations to a greater or lesser extent. We examined two nodes to gain insight into this issue.

In R2, the ccurrence of the *A. californica* + Sect. *Macrothyrsus* sister group relationship was supported by 86% PP (Fig. 3b). In the set of 100 randomly sampled trees using Random Tree, we observed seven variant placements of Sect. *Macrothyrsus* (Table 5). A topology showing the *A. californica* + Sect. *Macrothyrsus* relationship was observed in 87% of the sub-set (n = 87). Similarly, in R1, the sister relationship between Sect. *Aesculus* and the Asian clade was recovered in 73% (73% PP) of the MCMC sampled Bayesian trees. This relationship was observed in 76% (n = 76) of the sub-set of trees.

CONCLUSIONS

DIVA may be applied to the set or a sub-set of sampled trees from a Bayesian MCMC run to determine the probabilities of variant biogeographic hypotheses for nodes of interest. This approach to exploring biogeographic history can generate a set of hypotheses, some of which may have been overlooked, if only a single tree was used as the framework for biogeographic analysis. This approach better explores all variant hypotheses given the data and assigns each hypothesis a probability so that it can be considered quantitatively.

Consideration of the biogeography of a group of taxa is possible in the absence of a well-supported bifurcating tree using this approach. This is particularly useful when the

monophyly of a crown group of interest is well supported but relationships among the crown groups remain uncertain.

This method of biogeographic analysis using DIVA generates quantitative values that can be used for testing the impact of wildcard taxa on various biogeographic hypotheses for crown lineages of interest.

REFERENCES

Bremer, K. Ancestral Areas: A cladistic reinterpretation of the center of origin concept. *Systematic Biology*. 41(4): 436-445 (1992)

Budantsev, L.Y. History of the Artic flora of the early Cenophytic epoch. Nauka, Leningrad. 156pp. (1983) (In Russian, cited and discussed in Manchester 2001).

Forest, et al. A morphological phylogenetic analysis of *Aesculus* L. and *Billia* Peyr. (Sapindaceae). *Canadian Journal of Botany*. 79: 154-169 (2001)

Hardin, J. A revision of the American Hippocastanaceae. *Brittonia*. 9(3): 145-171 (1957a)

Hardin, J. A revision of the American Hippocastanaceae II. *Brittonia*. 9(3): 173-195 (1957b)

Hardin, J. Studies in the Hippocastanaceae: Old World Species. *Brittonia*. 12(1): 26-38 (1960)

Hu. *Aesculus tsiangii*. *Journal of Sichuan University: Natural Science Edition*. 3: 93-95 and pl. 6-7 (1960)

Hu and W. Fang. *Aesculus wangii*. *Journal of Sichuan University: Natural Science Edition*. 3: 99-100 and pl. 9-10 (1960)

Huelsenbeck, J. Accomodating phylogenetic uncertainty in evolutionary studies. *Science*. 288:2349-2350 (2000)

Huelsenbeck, J. and F. Ronquist MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17:754-755 (2001)

Kauff, F., J. Miadlikowska, and F. Lutzoni. ARC: a program for Ambiguous Regions Coding. Available from the authors http://www.lutzonilab.net/pages/download/ shtml (2003).

Kauff, F. RandomTree. <http://www.lutzonilab.net/downloads/index.shtml>

Kauff, F., J. Miadlikowska, and F. Lutzoni. ARC: a program for Ambiguous Regions Coding. Available from the authors http://www.lutzonilab.net/pages/download/ shtml (2003).

Kearney, M. and J. Clarke. Problems due to missing data in phylogenetic analyses including fossils: A critical review. *Journal of Vertebrate Paleontology.* 23(2): 263-274 (2003)

Lutzoni, F.; M. Pagel and V. Reeb. Major fungal lineages are derived from lichen symbiotic ancestors. *Nature*. 411:937-940 (2001)

Manchester, S. Leaves and fruits of *Aesculus* (Sapindales) from the Paleocene of North America. *International Journal of Plant Science.* 162(4):985-988 (2001).

Miadlikowska, et al. New approach to an old problem: Incorporating signal from gap-rich regions of ITS and rDNA large subunit into phylogenetic analysis to resolve the Peltigera species complex. *Mycologia*. 95(6):1181-1205 (2003).

Nixon, K and Q. Wheeler. Extinction and the origin of species. Ed. M. Novacek and Q. Wheeler. *Extinction and Phylogeny*. Columbia University Press, New York. pp 119-143 (1992)

Posada D and Crandall K*A.* Modeltest: testing the model of DNA substitution. Bioinformatics 14 (9):817-818 (1998).

Ronquist, F. DIVA version 1.1. Computer program and manual available by anonymous FTP from Uppsala University (ftp.uu.se or ftp.systbot.uu.se) (1996)

Ronquist, F. Bayesian inference of character evolution. *Trends in Ecology and Evolution.* 19(9):475-481 (2004)

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

Simmons, M. et al. How meaningful are Bayesian support values? *Molecular Biology and Evolution*. 21(1): 188-199 (2004)

Swofford, D. L. 2000. PAUP* v.4.0b10.: Phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland, M*A.*

Xiang, Q.Y. et al. Origin and biogeography of *Aesculus* L. (Hippocastanaceae): A molecular phylogenetic perspective. *Evolution* 52(4):988-997 (1998)

Table 1: Example of probability calculation - Hypothetical Bayesian tree sample of three trees (T_1-T_3). Node Y is some node (or ancestor) of interest. A, B, and C are symbols used to represent distribution areas. For example A: Europe. B: Asia, C: eastern North America. A, B, C, AB, BC, and ABC are therefore all possible range sets or all possible distribution ranges of Y given the data. The shaded element shows 1/S for distribution A of T_1 for node Y. 1 indicates that A is a possibility given in the summary DIVA optimizations for T_1. Three (3) is the total number of possible ancestoral distributions for node Y reported in the DIVA summary of optimizations for T_1. P is the sum of each column divided by the total number of trees and is thus the probability(P) of a distibution at node Y. The probability of a European distribution at node Y (or of ancestor Y) is 0.611 or 61.1% for this sample of trees.

	All Possible Range Sets for Node Y					
Trees	**A**	**B**	**C**	**AB**	**BC**	**ABC**
T_1	1/3	0	1/3	0	0	1/3
T_2	1	0	0	0	0	0
T_3	1/2	0	1/2	0	0	0
P	0.611	0	0.278	0	0	0.111

Table 2: Most probable ancestral ranges from R1 - Shows the most probable ancestral range of the last shared ancestor of each section with an unspecified sister group or section (x) from analysis of R1 data. Probabilities for Sect. *Aesculus* and Sect. *Parryana* cannot be explained entirely by Bayesian credibility values, i.e., the reoccurance of a relationship on a majority of Bayesian tree topologies.

Section	Most probable range	Probability	Bayesian Credibility supporting connetion to sister group	Z-statistic for Probability vs. Bayesian Credibility	Significant difference (compared to Zα/2 at 99%)?
Asian clade	Asia	0.755	73%	0.58	no
Aesculus	Asia	0.832	73%	2.719	yes
A. californica	Asia / western North America	0.76	75%	0.234	no
Pavia	eastern North America / western North America	0.663	70%	-0.781	no
Parryana	eastern North America / western North America	0.9	70%	6.628	yes
Macrothyrsus	none >50%	-	-	-	-

Table 3: Most probable ancestral ranges from R3 at nodes of interest from probabilistic analysis using DIVA - Showing the three most probable ancestral ranges calculated using a probabilistic approach to DIVA at each node of interest. Nodes of interest described in text and highlighted in Fig. 7. A = Europe, B = Asia, C = eastern North America, D = western North America

	Total Number of Range Sets with P > 0	Most Probable Ranges		
		1	2	3
Ancestral Node of Clade + unspecified sister				
Asian Clade	13	B (0.261)	AB (0.15)	ABCD (0.13)
Section Aesculus	15	A (0.15)	ABD (0.12)	B (0.10)
Section Macrothyrsus + A. californica	19	ACD (0.20)	BCD (0.14)	C (0.12)
Section Pavia	8	CD (0.93)	AC (0.032)	BCD (0.012)
Section Parryana	13	CD (0.77)	D (0.13)	BCD (0.019)
Ancestral Node of Clade (Multispeciate Only)				
Asian Clade	2	B (0.995)	AB (0.005)	-
Section Aesculus	3	AB (0.75)	ABC (0.15)	BD (0.10)
Section Pavia	1	C (1.00)	-	-
Section Macrothyrsus + A. californica	6	CD (0.84)	BCD (0.05)	BD (0.03)

Table 4: Comparison of ancestral ranges of each section of Aesculus and an unspecified Aesculus sister group (x) when fossils are included and excluded in DIVA analysis - Showing the % increase or decrease in the probability for ancestral ranges that include Europe and Asia for each section of Aesculus and some shared sister (*x*) when fossils are included in the probabilistic analysis in DIVA. Negative z-statistic values indicate that there was an increase in the probabilty when the fossils were included. Caution is advised when interpreting the resulting z-statistics, since the analysis in DIVA including fossils also included molecular markers in addition to those included in the analysis excluding the fossil taxa. % change is caluclated as the difference between the probability for each range when fossils are included and excluded (respectively). P = probability; [1] = Data from R1, * = calculations for Sect. Macrothyrsus (no AEP), since later data from Chapter 1 better support the relationship between Sect. *Macrothyrsus* and *A. c*

	excluding all fossils[1]		**including fossils**		**% change**		**ΔP(Europe) significant at α = 0.05?, (z stat)**	**ΔP(Asia) significant at α = 0.05?, (z stat)**
Section	**P(Europe)**	**P(Asia)**	**P(Europe)**	**P(Asia)**	**P(Europe)**	**P(Asia)**		
Aesculus	0.088	0.990	0.600	0.580	51.2%	-41.0%	yes (-9.047)	yes (8.143)
Asian Clade	0.000	1.000	0.436	0.961	43.6%	-3.9%	yes (-8.792)	yes (2.014)
Macrothyrsus + A. californica	0.024*	0.025*	0.410	0.434	38.6%	40.9%	yes (-7.493)	yes (-7.870)
Parryana	0.007	0.036	0.039	0.051	3.2%	1.5%	no (-1.518)	no (-0.520)
Pavia	0.000	0.022	0.370	0.027	37.0%	0.5%	yes (-7.663)	no (-0.228)

Table 5: Possible placements of Sect. *Macrothyrsus* observed in R2 - All possible placements of Sect. *Macrothyrsus* observed in the 100 randomly sampled trees in R2, shown in set notation. Occurrences refers to the number of times the topology was observed in the sample. Distribution refers to ancestral distribution(s) generated by DIVA for the shared ancestor of Sect. *Macrothyrsus* and its sister taxon. These distributions were the same for each occurrence of topology 1-7. Sect. *Macrothyrsus* is shown in bold in each possible topology. Note in Topology 5, Sect. *Macrothyrsus* is sister to the rest of *Aesculus*. Ancestral range notation: A: Europe, B: Asia, C: eastern North America, D: western North America. Range sets including more than one range indicate a widespread ancestor.

Topology #	Topology	# of Occurrences	Distribution (range sets)
1	(*A. californica*, **Sect. *Macrothyrsus***)	87	CD
2	((*A. californica*, Sect. *Aesculus*), **Sect. *Macrothyrsus***)	1	AC BC ABC BCD ABCD
3	(((Sect. *Pavia*, Sect. *Parryana*), *A. californica*), **Sect. *Macrothyrsus***)	4	CD
4	((Sect. *Pavia*, Sect. *Parryana*), **Sect. *Macrothyrsus***)	4	C
5	(**Sect. *Macrothyrsus***, *Aesculus* L.)	2	BC CD ACD BCD ABCD
6	(**Sect. *Macrothyrsus***, Sect. *Aesculus*)	1	AD CD ACD BCD ABCD
7	(((Sect. *Pavia*, Sect. *Parryana*), Sect. *Aesculus*), **Sect. *Macrothyrsus***)	1	C

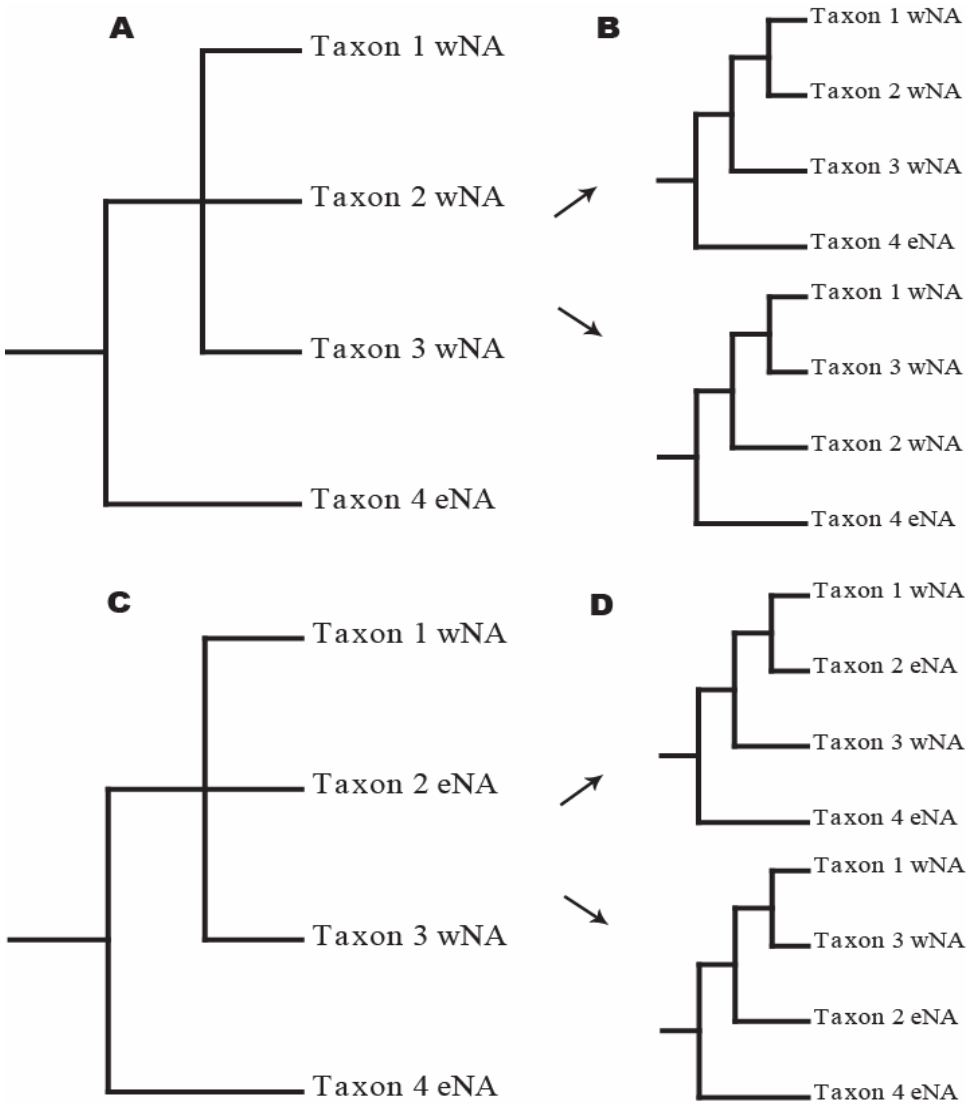

Fig. 1: **Resolution of polytomies and effect on DIVA output - A and B:** A polytomy in Tree A is solved in Tree B. Since Since Taxon 1, 2, and 3 have the same distributions, any solution to the polytomy will not affect the outcome of DIVA. **C and D:** A polytomy in Tree C has two possible solutions (D).. Since Taxon 1, 2, and 3, do not have the same distributions, different resolutions of the polytomy may produce different optimizations in DIVA. **wNA and eNA**: eastern and western North America

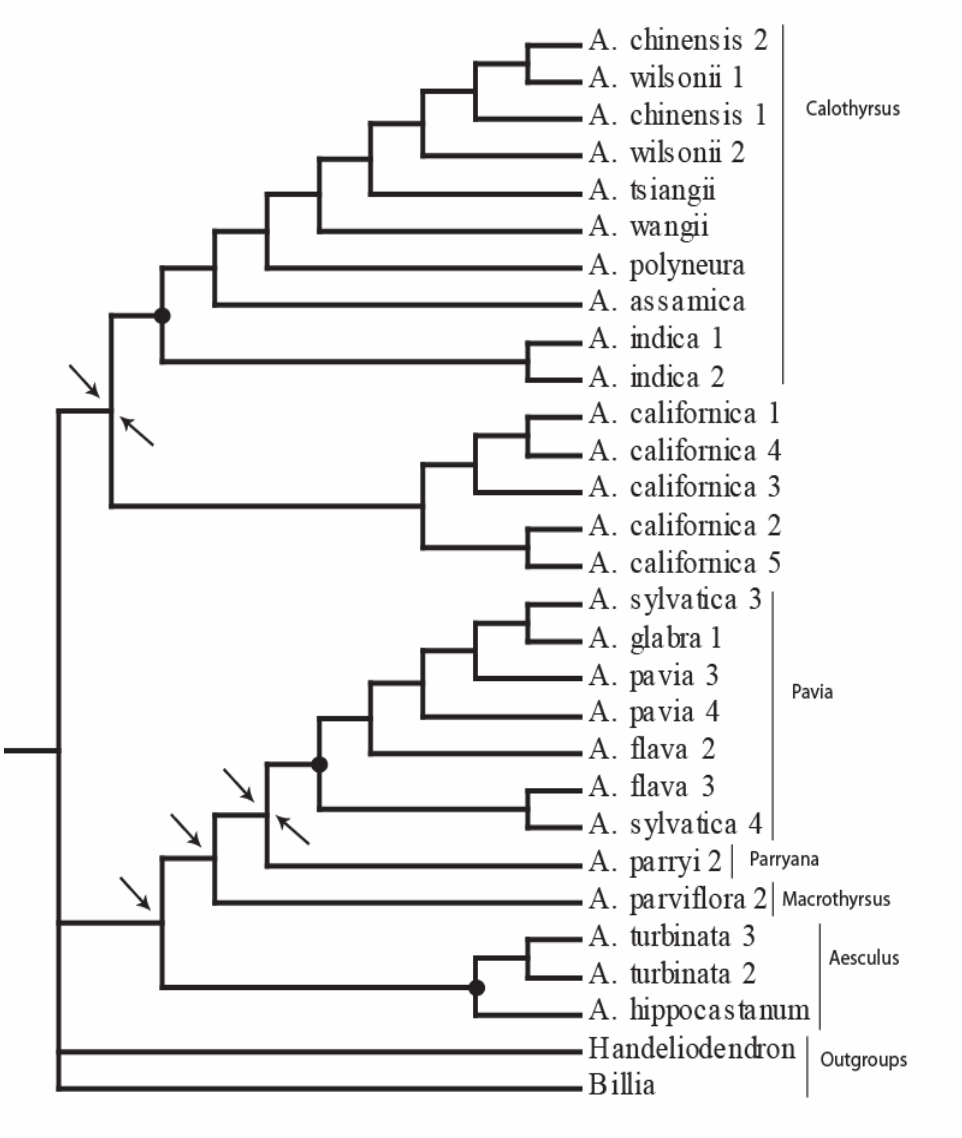

Fig. 2: **Nodes of interest considered in the present study -** One of 100 randomly sampled Bayesian trees from R1 shown here to highlight nodes (or ancestors) of interest. Circles indicate the last common ancestor shared by members of polytypic sections. Arrows indicate the last shared ancestor of each section and *x*, an unspecified sister. Two arrows pointing to the same node indicate that this node represents an ancestor shared by two sections. Section names are shown to the right of the tree. Section *Calothyrsus* excludes *A. californica* (se text) We do not propose that this tree is representative of the phylogeny of *Aesculus*. The purpose here is only to highlight nodes considered in this study.

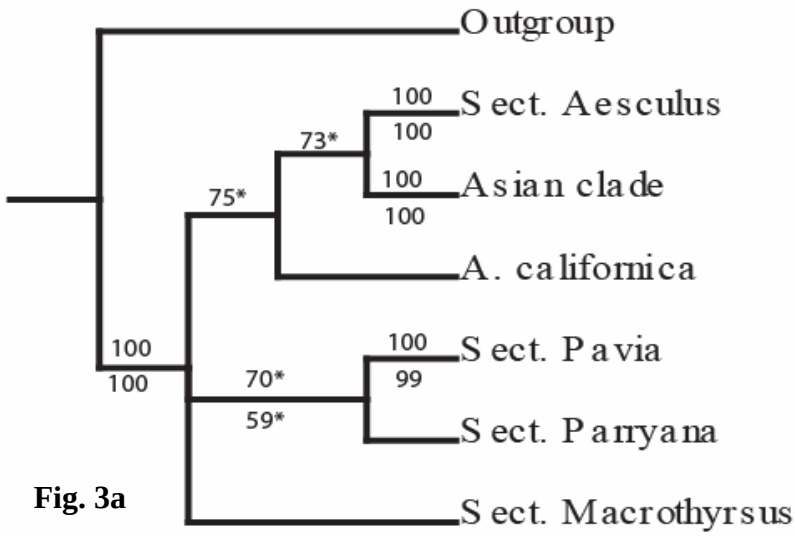

Fig. 3a

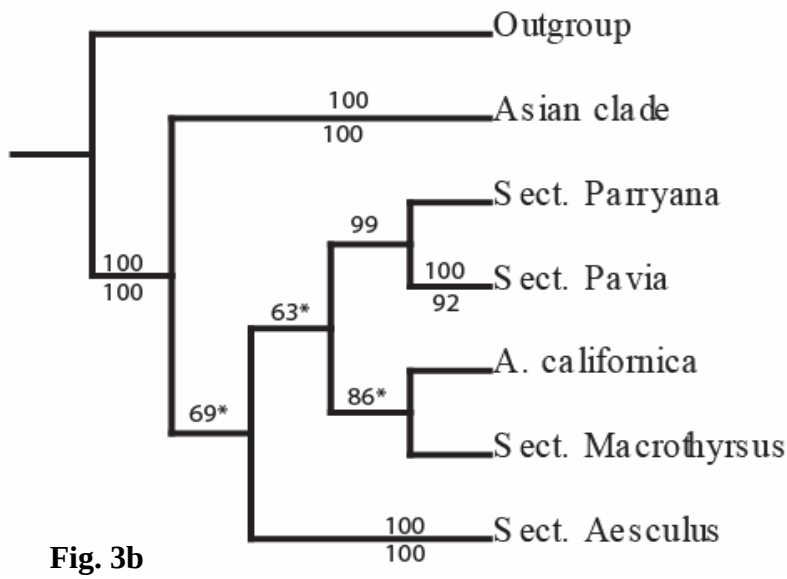

Fig. 3b

Fig. 3: **Summary Bayesian trees from R1and R2** – **3a** – Summary 50% Majority rule consensus tree of 19,800 post-burnin Bayesian trees from replicate 1 (rps16 and ITS gaps and ambiguous regions uncoded). **3b** – Summary 50% majority rule consensus tree of 19,800 post-burnin Bayesian trees from R2. PP support shown above branches. Bootstrap support shown below branches. Support for polytypic sections shown on terminal branches. All available accessions were included in the two replicates. * = By convention Bayesian and Bootstrap support >50% is shown. However, * indicates that the value did not meet the criterion for determining whether there was disagreement between the trees.

Distribution Probabilities of Last Shared Ancestor of the Asian Clade and Sister Group

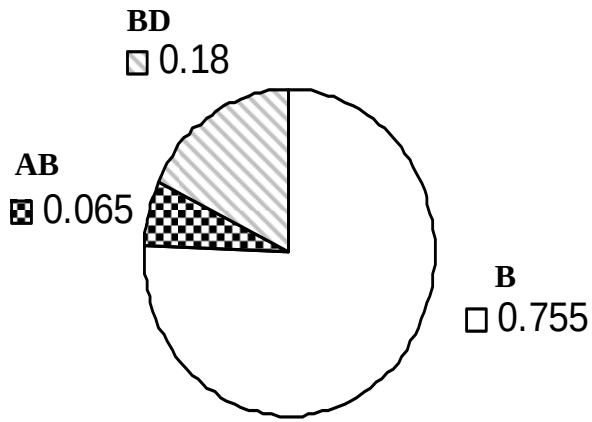

Fig. 4a

Distribution Probabilities of Last Shared Ancestor of Sect. Aesculus and Sister Group

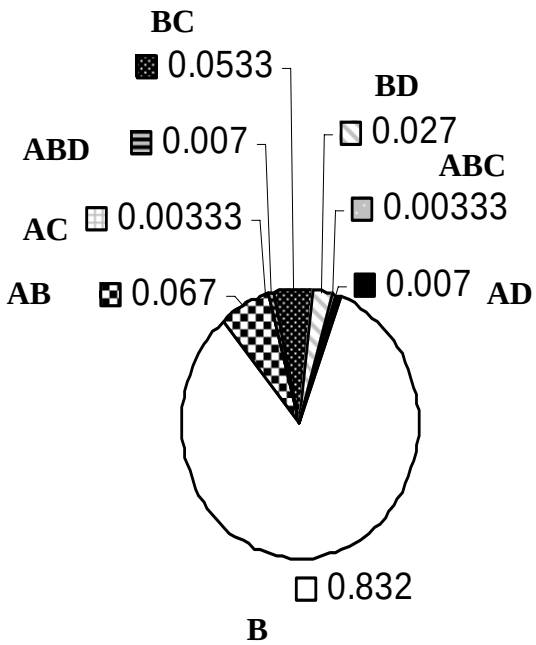

Fig. 4b

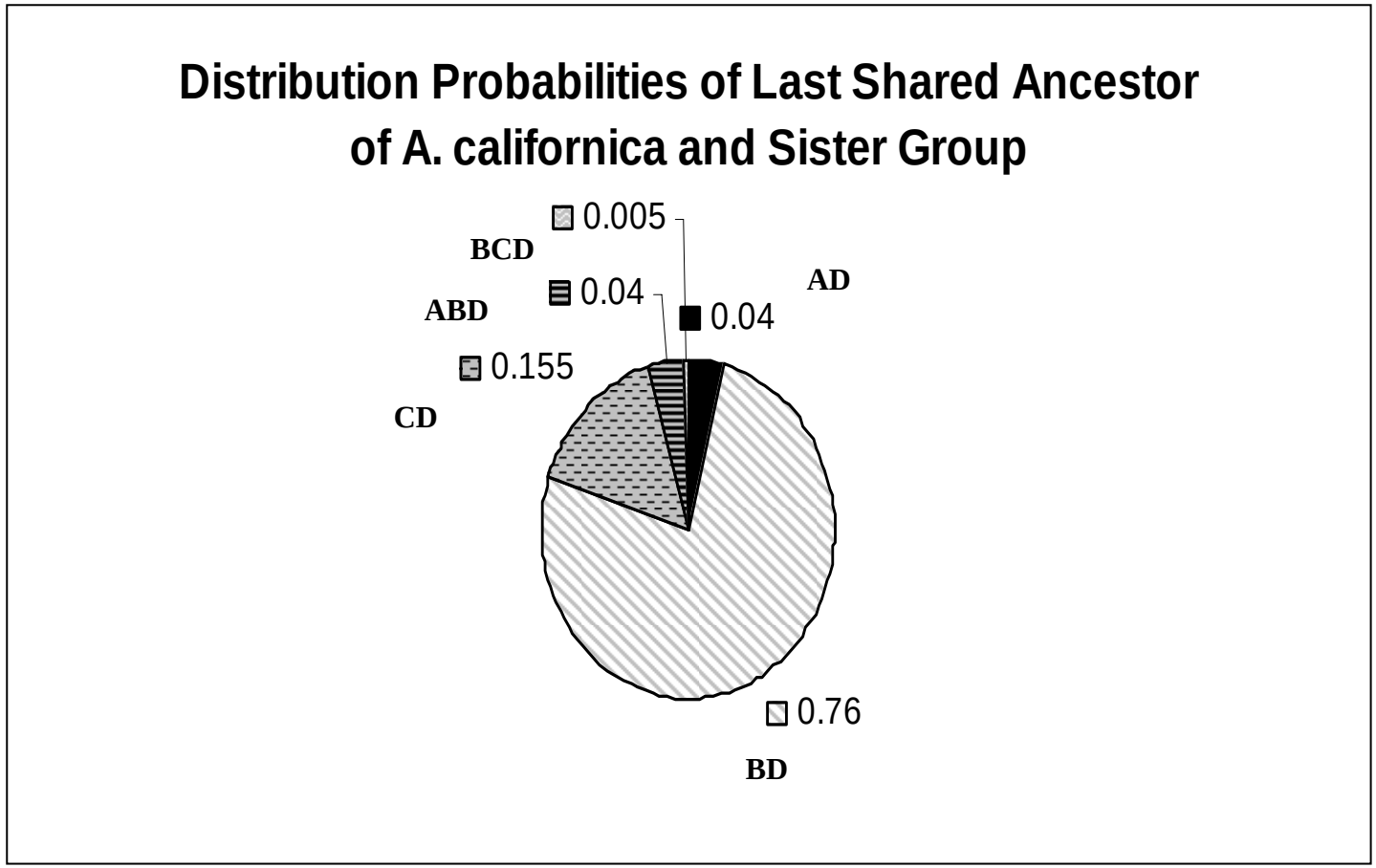
Distribution Probabilities of Last Shared Ancestor of A. californica and Sister Group
0.005
BCD
AD
0.04
ABD
0.04
0.155
CD
0.76
BD

Fig. 4c

Distribution Probabilities of Last Shared Ancestor of Sect. Pavia and Sister Group

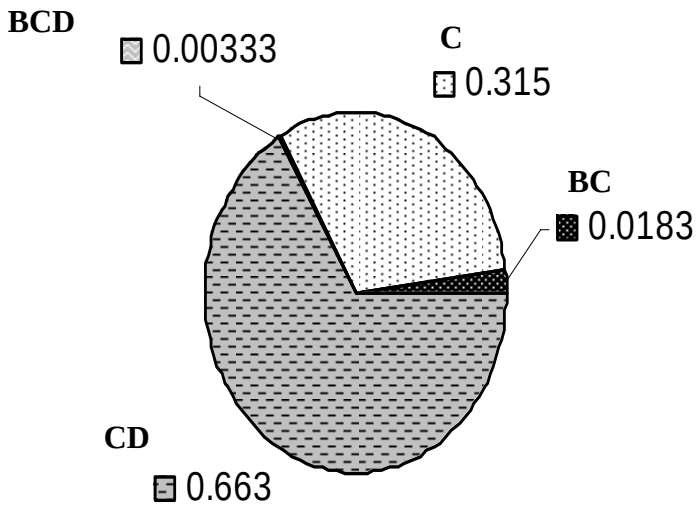

Fig. 4d

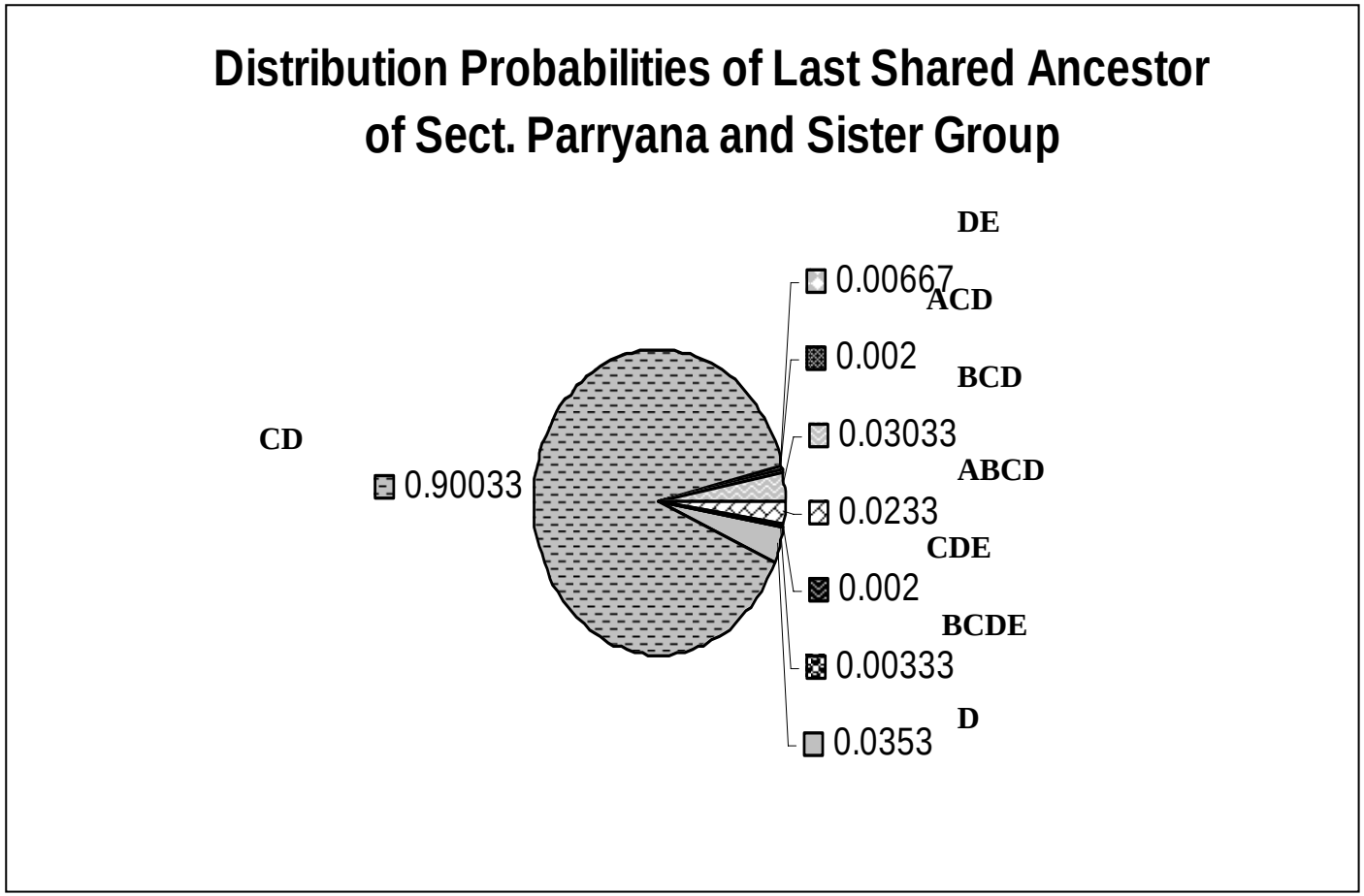
Distribution Probabilities of Last Shared Ancestor of Sect. Parryana and Sister Group
DE
0.00667
ACD
0.002
BCD
0.03033
ABCD
0.0233
CDE
0.002
BCDE
0.00333
D
0.0353
CD
0.90033

Fig. 4e

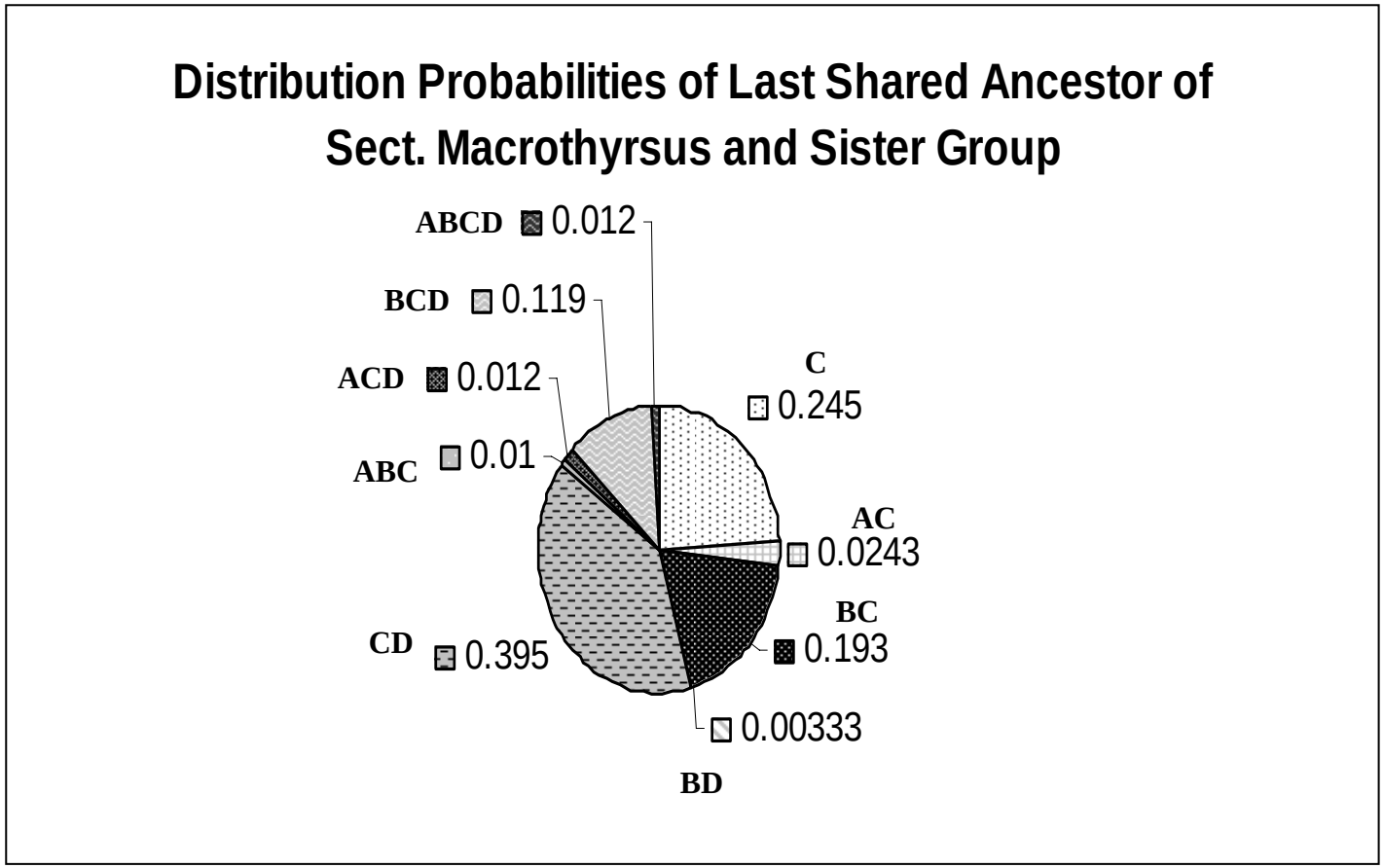
Distribution Probabilities of Last Shared Ancestor of Sect. Macrothyrsus and Sister Group
ABCD 0.012
BCD 0.119
ACD 0.012
ABC 0.01
C 0.245
AC 0.0243
BC 0.193
CD 0.395
BD 0.00333

Fig. 4f

Fig. 4a-f: **Distribution probabilities of the last shared ancestor of each of the 6 sections of *Aesculus* and some sister group (*x*)**. Data shown are the probabilities of ancestral ranges considered over 100 randomly sampled Bayesian trees from R1. Charts include all hypotheses where $P \neq 0$.

Fig. 4a-f Legend 1: Distribution Area Input for DIVA and Representative Symbols

A : Europe

B: Asia

C: eastern North America

D: western North America

E: Central/South America

Fig. 4a-f Legend 2: Possible Range Sets