Molecular Phylogenetics and Evolution 145 (2020) 106726

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Phylogenomics, co-evolution of ecological niche and morphology, and historical biogeography of buckeyes, horsechestnuts, and their relatives (Hippocastaneae, Sapindaceae) and the value of RAD-Seq for deep evolutionary inferences back to the Late Cretaceous

Zhi-Yuan Du[a,b], AJ Harris[c,d], Qiu-Yun (Jenny) Xiang[b,*]

[a] Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China
[b] Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC 27695-7612, USA
[c] Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China
[d] Department of Biology, Oberlin College, Oberlin, OH 44074, USA

ARTICLE INFO

Keywords:
Aesculus
Biogeography
ddRAD-seq
Hippocastaneae
Niche and morphological evolution
Phylogenomics

ABSTRACT

In this study, we used RAD-seq data to resolve the phylogeny of the tribe Hippocastaneae (Sapindaceae) and conducted comparative analyses to gain insights into the evolution and biogeography of the group that had fossils dating back to the late Cretaceous. Hippocastaneae, including the horsechestnuts and buckeyes, is a well-supported clade in Sapindaceae that comprises 12–14 species in *Aesculus*, two in *Billia*, and one in *Handeliodendron*. Most species in the tribe are distributed in Eurasia and North America and exhibit a classic pattern of intercontinental disjunction in the Northern Hemisphere, while *Billia* occurs from southern Mexico to northern South America. The earliest fossils of *Aesculus* date back to at least the earliest Paleocene of eastern Asia and western North America, where there are also putative occurrences from the latest Cretaceous. The group provides an excellent system for understanding floristic disjunction in the Northern Hemisphere extending to the Neotropics. However, a strongly supported and well resolved phylogeny is presently lacking for the tribe. Previous phylogenetic studies using several gene regions revealed five well-supported clades in *Aesculus*, largely corresponding to five recognized taxonomic sections, but relationships among these clades and among *Aesculus*, *Billia*, and *Handeliodendron* were not well supported. In this study, we used RAD-seq data from 68 samples representing all clades and species of Hippocastaneae except *Billia*, for which we used one of two species, to further resolve relationships within the tribe. Our phylogenomic analyses showed strong support for a sister relationship between *Aesculus* and *Handeliodendron*, in contrast to previous findings which supported *Billia* as sister to *Aesculus*. Within *Aesculus*, relationships among sections were strongly supported as (sect. *Calothyrsus*, (sect. *Aesculus*, (sect. *Macrothyrsus*, (sect. *Parryana*, sect. *Pavia*)))). We found that the traditionally recognized section *Calothyrsus* was monophyletic, with all eastern Asian species sister to the western North American species, *A. californica*. Analyses of divergence times combined with biogeographic analyses suggested a Late Cretaceous origin of Hippocastaneae, in eastern Asia, western North America, and Central America (including southern Mexico), followed by isolation of *Billia* in Central America, extinction of the tribe ancestor in western North America, and divergence of *Aesculus* from *Handeliodendron* in eastern Asia. A Late Cretaceous origin of the common ancestor of *Aesculus* in eastern Asia was followed by dispersals into western North America, Europe, and eastern North America during the Late Cretaceous and the Paleogene. Our results support *Aesculus* as a relic of the boreotropical flora and subsequent intercontinental spread of the genus through the Bering and North Atlantic land bridges. We performed character mapping analyses, which revealed that biogeographic isolation and niche divergence may have played important roles in driving morphological evolution and lineage divergence in *Aesculus*. Our study demonstrates the value of RAD-seq data for reconstructing phylogeny back to the Late Cretaceous.

* Corresponding author.
E-mail address: jenny_xiang@ncsu.edu (Q.-Y.J. Xiang).

https://doi.org/10.1016/j.ympev.2019.106726
Received 10 August 2019; Received in revised form 26 December 2019; Accepted 28 December 2019
Available online 29 December 2019

1. Introduction

The origins of the intercontinental floristic disjunctions of species, such as *Aesculus* L., in the Northern Hemisphere are ancient and complex (Boufford and Spongberg, 1983; Wu, 1983; Donoghue and Smith, 2004; Manchester and Tiffney, 2001; Tiffney, 1985a; Wen et al., 2010). According to macrofossil and palynological evidence, a "boreotropical flora" spread rapidly over the Northern Hemisphere in the early Paleogene (66.0–23.03 million years ago [Ma]; GSA Geologic Time Scale v.5.0 followed throughout; https://www.geosociety.org), a globally warm period (Wolfe, 1975; Budantsev, 1992). Thereafter, climatic cooling during the Eocene-Oligocene transition (Prothero, 1994) shaped the boreal flora into a "mesophytic forest" (see Tiffney, 1985a). The early development of the boreal temperate flora was considered to have mainly taken place in the high latitudinal Beringian subcontinent (Anagaro-Beringia in Budantsev, 1992), which is typically defined in the present-day as comprising the region including Alaska and other areas between the Lena or Anagara Rivers in Russia and the Mackenzie River in Canada. Additional cooling of the climate in the Miocene, Pliocene, and Quaternary and increasingly dry climates in continental interiors associated with mountain formation eliminated elements of this mixed mesophytic forest from central Eurasia and central North America resulting in the present-day disjunct distribution of closely related species. The survivors of this forest are now found in a few "refugia" including in eastern and western Asia, southeastern Europe, and eastern and western North America (Tiffney, 1985a; Wolfe, 1975). Intercontinental exchanges occurred between eastern North America and Europe via the North Atlantic land bridge and between western North America and eastern Asia via the Bering land bridge (Milne, 2006; Graham, 2018). These exchanges occurred bidirectionally, but out of Asia dispersals across the Bering land bridge appear to be more common among investigated taxa (Xiang et al., 1998, 2000, 2005; Wen, 1999; Manos and Donoghue, 2001; Xiang and Soltis, 2001; Donoghue and Moore, 2003; Donoghue and Smith, 2004; Harris et al., 2009a; Harris and Xiang, 2009b; Harris et al., 2013; Wen et al., 2010; Zhao et al., 2013). Based on evidence from previous phylogenetic and historical biogeographic analyses, disjunctions in most investigated genera originated in the Neogene (23.03–2.58 Ma), and correlated with periods of climatic cooling (Wen, 1999; Xiang et al., 2000; Wen et al., 2010). This understanding of the floristic disjunction in the Northern Hemisphere can be refined with new data from studies of disjunct taxa with abundant fossils and robust dated phylogenies.

Historically, the patterns of disjunctions in the Northern Hemisphere have been most robustly investigated within temperate, subtropical, and boreal regions, but the pattern also extends into the Neotropics of Central and northern South America. The occurrence of boreotropical species in Central America may be attributed to gradual migration through western North America into northern Central America at different times since the late Cretaceous when the Chortis block, a continental fragment, emerged and was aligned to southern Mexico, and further into southern Central America after the Eocene when the Chorotega block, another continental fragment, emerged and was aligned to the Chortis block (Mann, 2007). The southward migration of the boreotropical flora may have mostly occurred during the climatic cooling period in the Oligocene and Neogene, as a response of thermophilic forest trees to lower temperatures (Graham, 1999). The occurrence of the boreotropical element in South America may have been recent through southward migration via the Panama Isthmus that was formed by at least the Pliocene (Graham, 1999; Stange et al., 2018). Alternatively, some South American constitutes may have arrived prior to the formation of the isthmus via long distance dispersal (Cody et al., 2010) rather than by gradual range expansion. Therefore, knowledge on the ages of divergence of the Neotropical outliers within the predominately temperate, subtropical, and boreal Northern Hemisphere disjunct taxa will be helpful for understanding the association between Central and South America and the relics of the boreotropical forest.

Following fragmentation of the boreotropical forest in the Northern Hemisphere including in Central America and northern South America (occasionally crossing the equator), geographic isolation of species is expected to have led to allopatric speciation and, thus, divergence of lineages in the different isolated continental areas. However, how morphology and genomes diverged in allopatry remains an open question for the disjunct taxa. A recent study revealed a common "mostly conserved-rarely adaptive" genomic architecture for putative single copy genes in a wide array of disjunct pairs in eastern Asia and eastern North America (Dong et al., 2019). The study identified a small set of genes under strong positive selection, and a number of these are associated with functions pertaining to responses to abiotic or biotic environmental stresses. The evidence suggested that divergence in environmental niches between geographic areas could have driven morphological and/or ecological divergence associated with speciation. This is consistent with the long-held view that lineages isolated by vicariance may initially undergo limited evolutionary change, followed by later adaptations in response to subsequent environmental change (Harris et al., 2018). At present, few studies have examined the evolution of both ecological niche and morphological character to gain insights into the evolution of the relic boreotropical forest after its fragmentation, and, to this end, the tribe Hippocastaneae, including *Aesculus*, provides an excellent system for study.

The tribe Hippocastaneae consists of three woody plant genera, *Aesculus* (buckeyes and horsechestnuts), *Billia* Peyr., and *Handeliodendron* Rehder, which are sister to the maples of tribe Acereae in the broadly circumscribed Sapindaceae (Acevedo-Rodriguez et al., 2011; Buerki et al., 2009, 2010). The three genera of Hippocastaneae constitute the traditionally recognized family, Hippocastanaceae (Hardin, 1957a, 1957b, 1960; Turland and Xia, 2005; Xia et al., 2007; Hammel et al., 2015). Within Hippocastaneae, *Aesculus* consists of 12–14 species that are disjunctly distributed in the Northern Hemisphere, primarily in eastern Asia and eastern North America, with one species native to Europe and two to western North America (Hardin, 1957a, 1957b, 1960; Xia et al., 2007). *Billia* has two described species in Central and northern South America, *B. rosea* C. Ulloa & P. Jorge, ranging from Costa Rica to Colombia, Venezuela, and Ecuador, and *B. hippocastanum* Peyr., from Mexico to Panama (Hardin 1957a, b; Woodson et al., 1975; Hammel et al, 2015). *Handeliodendron* has a single species, *H. bodinieri* Rehder, endemic to southern China (Xia et al., 2007). Although different biogeographic hypotheses have been inferred for *Aesculus* based on different lines of evidence (Raven and Axelrod, 1974; Hardin 1957a; Xiang et al., 1998; Forest et al., 2001; Harris et al., 2009a; Harris et al., 2016), the biogeographic history of the entire tribe has not been elucidated using a phyologenetic framework due to uncertainty of relationships among the three genera. Hypotheses regarding the origin of *Aesculus* included a Northern Hemisphere origin (Raven and Axelrod, 1974), a Central or South American origin from a *Billia*-like ancestor (Hardin, 1957a), and a northern high latitude of eastern Asia (Xiang et al., 1998) or Beringian origin (Harris et al., 2009a) in the Paleocene or earlier as an element of the Northern Hemisphere boreotropical flora (Wolfe, 1975). These hypotheses were based primarily on observations of geographic distributions of extant or fossil species, phylogenies of morphological data or a few gene regions, or a combination of these approaches, but in all cases relationships among extant species remained incompletely resolved limiting the power of inferences (Raven and Axelrod, 1974; Forest et al., 2001; Xiang et al., 1998; Harris et al., 2009a).

Within Hippocastaneae, *Aesculus* and *Billia* were assumed to comprise a clade that is sister to *Handeliodendron*, probably because *Aesculus* and *Billia* had long been the only two genera in the former Hippocatanaceae family until 1994 when *Handeliodendron* was added to the family, and because *Handeliodendron* is distinct from *Aesculus* and *Billia* in a number of morphological features (Rehder, 1935; Judd et al., 1994; Xiang et al., 1998; Xia et al., 2007; see more details below and in

discussion). *Handeliodendron* was added to Hippocastanaceae based on a phylogenetic analysis of morphological data (Judd et al., 1994), which resolved it as sister to the other two genera. A molecular study using two plastid gene *mat*K and *rbc*L for Sapindaceae by Harrington et al. (2005) also indicated that *Aesculus*, *Billia*, and *Handeliodendron* form a strongly supported clade, and suggested that *Billia* was more closely related to *Aesculus* than *Handeliodendron*. However, recent phylogenetic studies based on several loci from chloroplast and nuclear DNA show that the relationships among the three genera of the tribe are questionable and weakly supported (Buerki et al., 2010). In the taxonomic revision of *Aesculus* by Hardin (1957a, 1957b, 1960), the genus was divided into five sections on the basis of morphological characters, such as bud viscidity, fruit exocarp ornamentation, flower color, and petal morphology. According to Hardin, section *Aesculus* comprises two disjunct species, *A. hippocastanum* L. and *A. turbinata* Blume, in southeastern Europe and eastern Asia (Japan), respectively. Section *Macrothyrsus* has one species, *A. parviflora* Walter, which is endemic to the southeastern United States. Section *Parryana* has one species, *A. parryi* A. Gray, which is endemic to northern Baja California. Section *Pavia* has four species in eastern North America, and section *Calothyrsus* has four to ten species in eastern Asia and one, *A. californica* Nutt., in western North America. The status of some eastern Asian species in section *Calothyrsus* have been called into question (Xiang et al., 1998; Turland and Xia, 2005). For instance, *A.* "*wangii*" Hu, which was not validly published, was merged with *A. assamica* Griff., and *A. wilsonii* Rehder was reduced to a variety of *A. chinensis* Bunge (Turland and Xia, 2005; Xia et al., 2007). Nevertheless, the status of these species remains an open question (e.g., Zheng et al., 2018).

Previous studies showed differences in the placement of *A. californica* and the relationships among sections. The treatment of *A. californica* in section *Calothyrsus* is supported by morphological data (e.g., sticky buds; Hardin, 1957a, 1957b; Forest et al., 2001) and by chloroplast DNA sequences (Xiang et al., 1998; Harris et al., 2009a). However, phylogenies based on nuclear ribosomal gene ITS region and *LFY* gene have shown that *A. californica* is more closely related to the monotypic section *Macrothyrsus* of the southeastern United States (Xiang et al., 1998; Harris et al., 2009a). Among all prior molecular studies, the relationship of *A. californica* has not been strongly supported. Thus, the monophyly of section *Calothyrsus* has remained uncertain. In Harris et al. (2009a), analyses of both cpDNA and nrDNA strongly supported a sister relationship of section *Aesculus* to sections *Parryana* and *Pavia*, while a sister relationship of section *Macrothyrsus* to these three sections was poorly supported. As each section has a distinct geographic distribution, a robust phylogeny of the genus is needed to understand the biogeographic history, which can shed additional light on the classic pattern of floristic disjunction within the Northern Hemisphere.

Taxa within Hippocastaneae exhibit a wide range of morphological variation (Hardin, 1957a, 1957b, 1960; Judd et al., 1994; Judd et al., 2016). The three genera of Hippocastaneae share a number of features, including palmately compound leaves oppositely arranged on the stem, paniculate inflorescences, zygomorphic flowers, petals with basal appendages, unilateral floral discs, and large capsular fruits, and tricolporate pollen that is prolate in shape (Acevedo-Rodriguez et al., 2011; Harris et al., 2015). *Billia* is distinct in having free sepals, evergreen leaves, and 3 leaflets, whereas *Aesculus* and *Handeliodendron* have connate sepals, deciduous leaves, and 5 (*Handeliodendron*) or 5–11 (*Aesculus*) leaflets (Hardin, 1957a, 1957b, 1960; Xia et al., 2007; Acevedo-Rodriguez et al., 2011). Sepals of *Handeliodendron* are connate only at base, but sepals of *Aesculus* are connate half or more of their length. The winter buds of *Billia* do not have scales, whereas two sections of *Aesculus* (section *Calothyrsus* and section *Aesculus*) have sticky winter buds with imbricate scales (Acevedo-Rodriguez et al., 2011; Xia et al., 2007). Sticky winter buds are also present in *Handeliodendron* (Yuan Xu, personal communication, South China Botanical Garden). Within *Aesculus*, species vary in winter bud viscidity, leaf, flower, inflorescence, and fruit morphology, pubescence on different parts of the plants (Hardin, 1957a, 1957b, 1960; Forest et al., 2001), providing materials for studying morphological evolution and its relationships to biogeographic dispersal and ecological niche shifts in a framework of a robust phylogeny of the tribe.

RAD-seq is a recently emerged and cost-effective method to generate genome-wide molecular data for evolutionary studies (Baird et al., 2008; Peterson et al., 2012; see reviews in McCormack et al., 2013; Leaché and Oaks, 2017). The method has been mainly used for studies of population genetics and shallow phylogeny (i.e., closely related species; Boucher et al., 2016; Cavender-Bares et al., 2015; Eaton and Ree, 2013; Hipp et al., 2018; Hohenlohe et al., 2010; Pais et al., 2016; Qi et al., 2015; Rubin et al., 2012; Taranto et al., 2016; Zhou et al., 2018). Allelic dropout and loci dropout due to restriction site mutations in distant taxa has been a key concern on using RAD-seq data for deep phylogenetic studies. Too much allele drop off would result in too much missing data and fewer loci overlapping across distant species to be recovered (Leaché and Oaks, 2017). However, hierarchical missing data from RAD-seq has been shown to contain phylogenetic signals (DaCosta and Sorenson, 2016). In our lab, an experimental approach with careful pooling of high quality DNAs digested by *Pst*I-HF and *Msp*I enzymes and selection of a narrow range of fragments for sequencing has been demonstrated to be successful for phylogenetic studies of a number of taxa investigated at below and above the species-levels (Qi et al., 2015; Pais et al., 2016; Zhou et al., 2018). A few recent studies have shown the promise of RAD-seq data for inferring phylogenies of lineages that are as old as 60 million years in the Paleocene (e.g., Cariou et al., 2013; Leaché and Oaks, 2017; Rubin et al., 2012). Therefore, we investigate the value of RAD-seq data for a phylogenomic study of Hippocastaneae, which has rich fossil record dating back to at least the Paleocene (Hollick, 1936; Hu and Chaney, 1940; Schloemer-Jäger, 1958; Prakash and Barghoorn, 1961; Axelrod, 1966; Wolfe and Wehr, 1987; Manchester, 2001; Dillhoff et al., 2005; reviewed in Hardin, 1957; Mijarra et al., 2008; Harris et al., 2014). The objectives of this study are to (1) reconstruct a robust phylogeny of Hippocastaneae using ddRAD-seq generated sequence data to further resolve the relationships among *Aesculus*, *Billia*, and *Handeliodendron* and within *Aesculus*; (2) infer the evolutionary and biogeographic history of the tribe using the molecular phylogeny as a framework for integrating data from fossils and morphology; and (3) identify co-evolution of morphology and ecology within the tribe using character mapping and correlation analyses.

2. Materials and methods

2.1. Taxonomic sampling and RAD sequencing

We adopted a double digest Restriction site Associated DNA Sequencing (ddRAD-seq) method (Peterson et al., 2012) to generate genome wide molecular markers for this study. We collected fresh leaves from a total of 96 individuals in arboreta or from the field. Our samples represented all 13 accepted species of *Aesculus* according to Hardin (1957a, 1957b, 1960) plus *A.* "*wangii*", the singular species of *Handeliodendron*, and one of the two species of *Billia*, as well as two species of *Acer*, which served as the outgroup. We dried the collected material using silica and later extracted total DNAs with a modification of the CTAB method of Doyle (1991). We checked DNA quality and concentration and prepared the RAD-seq library following Peterson et al. (2012) with minor modifications as described in Pais et al. (2016); Zhou et al. (2018). After sequencing, we removed twenty-eight samples as a result of poor sequence quality or taxonomic uncertainty. The remaining 68 samples represented all the species in the tribe Hippocastaneae except *Billia hippocastanum* Peyr. (Supplementary data 1). For these samples, we used two restriction enzymes, *Pst*I-HF and *Msp*I, to digest 350 ng of total DNA. The *Pst*I adaptors and forward primer are the same as in previous studies (Pais et al., 2016; Zhou et al., 2018), but we used a different *Msp*I adaptor (top strand: GTG ACT GGA GTT CAG

ACG TGT GCT CTT CCG ATC T; bottom strand: CGA GAT CGG AAG AGC ACA CGT CTG AAC TCC AGT CAC) and NEBNext Index 4 Primer (NEB Inc, USA) in the present study. We performed 150 bp single end sequencing with fragments of 400–600 bp on Illumina HiSeq 2500 at Genomic Sciences Laboratory of North Carolina State University.

2.2. RAD-seq data analyses

The sequence data were filtered and assembled into loci using iPyRAD v.0.6.15 (Eaton and Overcast, 2018). Filtering parameters were set to replace low quality base calls of Q < 20 and discard reads containing more than five ambiguous bases. Consensus base calls were made for clusters with a minimum depth of coverage greater than six. We also excluded loci containing more than two alleles as most of the taxa with reported chromosome counts are currently diploid (Krahulcová et al., 2017).

It is known that the number of loci in the assembled data matrix is significantly affected by the parameters of clustering threshold for de novo assembly and minimum number of samples per locus. We used the raw data of 17 samples representing16 species of Hippocastaneae and one outgroup species of *Acer* to explore the phylogenetic effect of data assembled by changing these assembly parameters. We first tested clustering thresholds at 0.80, 0.85, and 0.90 to cover the program recommended 0.85 threshold and up and down deviations. The resulting matrices contained different numbers of pre-filtered loci (18363, 18685, and 19466, respectively). We then filtered these matrices by retaining only loci present in at least 50% of samples and performed phylogenetic analyses for each of the filtered matrices using the maximum likelihood (see below for details). The filtered matrices each contained 17 taxa and 2537, 2620, or 2671 loci, respectively. As the results showed no difference in the resulting tree topologies (Supplementary data 2), we chose to use the recommended 0.85 threshold for assembly of the data for all samples, followed by filtering of loci at different levels of number of samples per locus, from 30% through 80% at 10% intervals. This process resulted in six matrices containing loci present in at least 30%, 40%, 50%, 60%, 70%, and 80% of the 68 samples by filtering the matrix containing 49,412 pre-filtered loci. The final number of loci and the percentage of missing data in each of the six matrices are listed in Table 1. The data matrices were submitted to Mendeley Data (https://data.mendeley.com). Using these data matrices, referring to as M30, M40, … M80, respectively, we tested to what level phylogenetic relationship are affected by variation of this parameter (for details in phylogenetic analyses, see below).

To detect chloroplast loci in the reads, we used the chloroplast genome of *A.* "*wangii*" (GenBank accession no. MF583747) as a reference to map our reads in ipyRAD. This analysis yielded a matrix of five loci that were short in length (525 bp in total) and contained considerable missing data. We reconstructed a phylogeny using the recovered cpDNA sequences, and the results showed little resolution of species relationships (tree not shown). Therefore, we used the assembly method of "de novo-reference" in ipyRAD to exclude the cpDNA sequences from the data matrix.

Table 1
Statistics for six nucleotide matrices containing loci present in different minimum percentages of taxa, where M30, for example, indicates that included loci are present in at least 30% of taxa.

Matrix	M30	M40	M50	M60	M70	M80
No. of loci	3858	2668	1853	1305	912	594
No. of SNPs	50,721	37,761	27,041	19,423	13,478	8826
No. of unlinked SNPs	3803	2625	1816	1278	889	578
Missing data in unlinked SNPs matrix	47.37%	38.50%	30.40%	23.45%	17.43%	12.13%

2.3. Phylogenetic analyses

We conducted phylogenetic analyses for all data matrices using maximum likelihood (ML) methods implemented in RAxML v.8.2.10 (Stamatakis, 2014), with GTRGAMMA model for bootstrapping phase and 1000 rapid bootstrap replicates in CIPRES v.3.3 Science Gateway (http://www.phylo.org/). We visualized the resulting trees in FigTree v.1.4.3 (Rambaut, 2014). Further, we also reconstructed phylogenetic relationships using MrBayes v.3.2.2 (Ronquist et al., 2012) via the CIPRES. The Bayesian analyses comprised 1.0×10^8 generations with sampling every 1000 generations under the HKY + Gamma model as inferred from jModelTest v.2.1.10 (Darriba et al., 2012). We examined the results of the Bayesian analyses in Tracer v.1.7.0 (Rambaut et al., 2018) to verify stationarity (ESS > 200) for each run as well as convergence between independent runs and the suitability of a 25% burnin. After discarding the burnin, the remaining trees were used to estimate posterior probability through the construction of a 50% majority rule tree in MrBayes v.3.2.2 (Ronquist et al., 2012). We displayed the resulting trees in FigTree v.1.4.3 (Rambaut, 2014) and rooted on the branch between Hippocastaneae and *Acer*.

To reconstruct a basic phylogenetic framework for divergence time and biogeographic analyses, we also conducted ML and Bayesian analyses for a reduced M40 matrix containing 17 species, each represented by a single accession that, in general, exhibited the least amount of missing data. The original M40 matrix (68 samples, 2668 loci, 280311 bp) was chosen because it resulted in the phylogenetic tree with the highest nodal support in both ML and Bayesian analyses of matrices with full taxon sampling (68 samples). The number of loci in the reduced M40 matrix was the same as in the original M40 matrix, and the resulting phylogenetic tree was consistent with the other trees including full taxon sampling, justifying for the utility of divergence time dating analysis. Due to phylogenetic divergence, only a small proportion of loci were recovered in the outgroup *Acer*. To test if the extensive amount of missing data from the *Acer* species affected inferences of relationships within Hippocastaneae, we also conducted ML and Bayesian analyses of the reduced M40 matrix (17 taxa) containing only the 340 loci detected in the species of *Acer*.

2.4. Divergence time estimation

Divergence times were estimated in BEAST v.1.8.0 (Drummond et al., 2012) using the aforementioned reduced M40 dataset (17 taxa, 2668 loci) with *Acer* as the outgroup. To calibrate the tree, we applied four fossil constraints as uniform priors to four nodes: (1) 66.25–89.8 Ma on the root node (see Stange et al., 2018 for importance of root calibration) that connects Hippocastaneae and *Acer* according to the oldest known fossils of *Acer* from western North America (WNA) (*Acer* sp., 66.25–72.1 Ma, the minimum boundary of the root; Harris, unpublished data; Denver Museum of Natural Science paleobotanical collections including 39000, 39001, 39002, 39004, 39006 of "Licking Leaves", Harding County, South Dakota, Hell Creek Formation) and the oldest fossil of Sapindaceae that does not show apparent affinity with any extant genera (Bell, 1957) from WNA of Coniacian to Campanian age (72.1–89.8 Ma), maximum boundary of the root; (2) 66–72.1 Ma for the stem node of section *Aesculus* based on phylogenetic evidence for placement of the Paleocene fossil species *A. hickeyi* Manch (Manchester, 2001) sister to section *Aesculus* (Harris et al., 2009a; also below for determining phylogenetic placement of fossils) and that section *Aesculus* has to be younger than the minimum age of Hippocastaneae; (3) 56–66 Ma on the stem node of the clade including sections *Pavia* and *Parryana* based on phylogenetic evidence for placement of the Paleocene fossil species *A.* "*magnificum*" det. Budantsev (Budantsev, 1983; Manchester, 2001) at the stem of this clade (see below) and that the maximum age of the stem of sections *Pavia* and *Parryana* cannot be older than the minimum age of the node below, i.e., stem of section *Aesculus*; and (4) 5.33–23.03 Ma for the crown node of

A. hippocastanum-A. turbinatum based on the Pliocene fossils of the modern species *A. hippocastanum* and close, extinct allies of *A. hippocastanum*, likely representing progenitors or stem lineages, from the Miocene of Europe and northern Africa (Velitzelos and Gregor, 1990; Mijarra et al., 2008). For each node, the ranges of dates presented above represent hard lower and upper bounds for the uniform prior. For uniform priors, lower bounds are usually clear and easy to set, because a group cannot be younger than its first fossil. Therefore, provided the fossils are interpreted correctly, a hard lower bound based on the fossil is a reasonable prior. Upper bounds are more difficult to set, because absence from the fossil record does not imply that a taxon did not exist. We used the minimum ages of the node below or fossils representing the stem of the node as the upper bounds. We deem it likely that the root of our tree is younger than the oldest known occurrence of the large Sapindaceae family that does not show apparent affinity to any living lineages of the family (Bell, 1957).

We performed four additional divergence time analyses to test how results are affected by prior calibration models, root constraint, and missing data in the outgroup. First, we explored using lognormal priors with soft upper bounds for the three inner nodes, while keeping the root age calibration as 66.25–89.8 Ma using uniform model. To set the lognormal priors, we applied offset values using the aforementioned minimum dates and adjusted the mean and standard deviations so that the aforementioned date ranges comprised the 95% HPD bounds. We then repeated the analysis by removing the root calibration. In addition, we conducted analysis using an older maximum root age with uniform prior distribution of the root constrained between 66.25 Ma and 125 Ma (the age of the oldest angiosperm fossil *Archifrutus*; Sun et al., 2002; Friis et al., 2003). Finally, we tested whether missing data in the outgroup *Acer* could affect the results, especially the root age, by using the matrix consisting of 340 loci present in *Acer*, with the uniform model, and then repeated with the lognormal model, for all the four calibrated nodes.

For dating, we enforced a lognormal relaxed molecular clock with an HKY substitution model with gamma site heterogeneity, four rate categories, and a Birth-Death tree branching process. We confined the calibrated groups to be monophyletic. Before dating analyses, we checked our priors with empty runs without data. The mean values and HPD of the four calibrated nodes were well within the fossil constraints we set, indicting the effectiveness of the calibrations. The priors of other parameters used in empty runs were then implemented in the analyses with data and the fossil calibrations. For the analyses of our data and empty runs, the MCMC length was 4×10^8 generations sampled every 1000 generations. The BEAST log file was analyzed with Tracer v.1.7.0 (Rambaut et al., 2018) to verify the stationarity of the distribution based on effective sample sizes of 200 or greater for all parameters. After discarding the first 10% of trees as burn-in, we used the tree file to generate a maximum clade credibility tree in TreeAnnotator v.1.8.0 (Drummond et al., 2012).

2.5. Placing fossil species to molecular phylogeny

To confirm the placement of *A. hickeyi* and other fossil species key to biogeographic analyses (see below), we performed phylogenetic analyses of morphology in MrBayes v.3.2.1 using the ML tree of extant species resolved from the M40 matrix as a partial constraint (i.e., allowing fossils to be resolved within constrained clades). The morphological data matrix consisted of 43 characters (Supplementary data 3) and was edited from Harris et al. (2009a) based on new personal observations, especially of *Billia* in the field. Explanations of these characters can be found in Forest et al. (2001; same order as in Harris et al., 2009a). The fossil species are *A. hickeyi* from WNA (Manchester, 2001), *A. longipedunculus* Schloemer-Jäger (1958) from the island of Spitsbergen, and three fossil species *A.* "*magnificum*" (Budantsev, 1983; Manchester, 2001), *A. majus* (Nathhorst) Tanai (1952), and *A. miochinensis* Hu & Chaney (1940) from eastern Asia (see Harris et al., 2009a; Harris and Xiang, 2009b). Bayesian analyses were performed using the GTR + G model for the molecular data (Yang et al., 1996; Huelsenbeck and Rannala, 2004) and the MK1 model for morphological characters. The analysis comprised two runs of 10 chains for 1.0×10^7 generations with sampling of trees every 1000 generations. We summarized the results using standard procedures in Tracer to verify stationarity, convergence, and suitability of a 10% burnin, and we obtained a maximum clade credibility tree from the combined runs using Log Combiner and Tree Annotator of the BEAST 1.8 package (Drummond et al., 2012).

2.6. Biogeographic analyses

Biogeographic analyses were conducted on the dated BEAST tree with placement of fossil species according to the phylogenetic analysis using morphology with the molecular tree as a constraint (see above). In this analysis, we used the BEAST tree inferred from the dating analysis using a root calibration of 66.25–89.8 Ma and the uniform prior distributions for all of the four calibrated nodes. The taxa were scored for presence in one or more of the following six areas that cover the entire range of the genus: Europe (EU), eastern Asia (EA), eastern North America (ENA), western North America (WNA), Central and South America (CSA, including southern Mexico), and Spitsbergen. We performed model test using the R-package BioGeoBEARS (Matzke, 2013) installed in RASP (Yu et al., 2015). The test selected the DEC model (Ree and Smith, 2008; Ree and Sanmartín, 2009) as the best for our data. We then performed biogeographic analyses using the DEC method implemented in RASP to reconstruct ancestral distributions and dispersal events. In the analysis, we implemented the time slice-dependent dispersal constraints (Supplementary data 4) based on evidence from literature (Sanmartín et al., 2001; Manchester and Tiffney, 2001; Wen et al., 2016; Graham, 2018). We performed DEC analyses with and without the outgroup *Acer*. The ancestral range of Acereae is uncertain as there is no up-to-date biogeographical study inferring the ancestral distribution. An alternative would be to code Acereae as having a wide distribution in the biogeographic analysis, because the genus occurs widely in Asia, Europe, Africa, and North America. However, this may bias the range reconstructions of Hippocastaneae to wide distribution. Therefore, in the first analysis, we removed *Acer* from the phylogeny and conducted analysis without defining the outgroup. In a second analysis, we included *Acer* using the inferred root area of Beringia from a paleobotanical study by Boulter et al. (1996). We followed Harris et al. (2009a) to set the ancestral distribution of Acereae as EA and WNA.

2.7. Examination of character and niche evolution

To gain insights into morphological evolution of Hippocastaneae in relation to niche evolution, we tracked the evolutionary history of 22 morphological characters and compared these with evolutionary changes in niche preference. The morphological characters varied among species of *Aesculus* were selected from the dataset of 43 characters utilized for the phylogenetic analysis including fossils (Supplementary data 3). We conducted character reconstructions using ancThresh implemented in Phytools (Revell, 2012) on the BEAST tree of the reduced matrix M40. The ancThresh function uses Bayesian MCMC to estimate ancestral states and thresholds for discrete characters on the phylogenetic tree.

Climate variables for each species were estimated from georeferenced occurrence records of herbarium specimens in the public database Global Biodiversity Information Facility (https://www.gbif.org/). The data were curated by removal of records labeled with cultivated origin or with questionable georeferences outside of known ranges of species. We also performed data thinning in R (R Core Team, 2019) to remove records in geographical proximity of 1.5 km of each other. Most species had at least 50 records, except *A. assamica*, *A.* "*wangii*", *A. chinensis* and *Handeliodendron bodinieri*, for which there are only 10–20

records available. For each occurrence point, we sampled elevation and the 19 precipitation and temperature variables comprising the WorldClim v.2.0 dataset (Fick and Hijmans, 2017) at five arc minute resolution (ca. 10 km^2 at the equator). We extracted the values of the grid cells containing occurrence records using Phytools (Revell, 2012). The mean values and variances of these variables were calculated within species and used for tracing the evolution of ecological niches using fastAnc in Phytools on the BEAST tree of the reduced matrix M40. The fastAnc in Phytools infers ancestral values of continuous variables based on a continuous stochastic process that behaves like a Brownian motion, but attracted toward some central value, and the strength of the attraction increases with the distance from that value (Joy et al., 2016). We also performed a tests of phylogenetic signal (Pagel's λ and Blomberg's K) and pairwise correlation between bioclimatic variables (adjusted R^2), and between each pair of bioclimatic variable and morphological character in Phytools using *Aesculus* species. Values of Pagel's λ (Pagel, 1999) and Blomberg's K (Blomberg et al., 2003) less than 1.0 correspond to characters being less similar among species than expected from their phylogenetic relationships, whereas values greater than or equal to 1.0 represent strong phylogenetic signal. Adjusted R^2 close to 0 indicates no correlation between variables, whereas 1.0 means strong correlation.

3. Results

3.1. RAD-seq data and phylogenetic analyses

The resulting sequences from RAD-seq were 105 bp long after removal of the adapters and recognition sequences. The samples had an average of 652,172 reads each (range from 110,663 to 2,533,769) that passed quality filtering. The filtered reads were clustered into an average of 43,636 clusters per sample (range from 12,485 to 255,184) and had a mean depth of 22.3 (range from 3.1 to 77.7). Among these, an average of 7332 clusters per sample had a depth greater than six.

The M30–M80 matrices vary significantly in the number of loci and amount of missing data (Table 1). However, phylogenetic analyses of them using the ML and Bayesian Inference methods resulted in highly similar topologies that varied mainly in nodal support (trees not shown). The tree from the M40 matrix with 68 samples showed the highest support overall (Fig. 1) and resolved all accessions from the same species together. The tree also revealed *Handeliodendron* as the sister to *Aesculus* with 99% ML bootstrap (BS) support and 0.97 Bayesian posterior probability (PP). Within *Aesculus*, each of the non-monotypic sections was monophyletic (Fig. 1). In section *Calothyrsus*, the western North American *A. californica* is sister to the eastern Asian species, and within the Asian clade, the divergence order is (*A. indica*, (*A. assamica*, (*A.* "*wangii*", (*A. chinensis*, *A. wilsonii*)))). In section *Pavia* from eastern North America, the divergence order is (*A. glabra*, (*A. pavia*, (*A. sylvatica*, *A. flava*))). Relationships among sections were completely resolved as the following (sect. *Calothyrsus*, (sect. *Aesculus*, (sect. *Macrothyrsus*, (sect. *Parryana*, sect. *Pavia*)))). All of the relationships among genera of Hippocastaneae and sections of *Aesculus* were strongly supported (Fig. 1).

The ML and Bayesian Inference analyses using the reduced M40 matrix with 17 samples showed identical relationships to those described above (Supplementary data 5a). The same topology was also recovered with the matrix of 340 loci present in *Acer* (Supplementary data 5b). In all analyses of the M40 matrix with full taxon and loci sampling, reduced taxon sampling (17) or reduced loci sampling (340), the monophyly of *Aesculus* and a sister relationship between *Aesculus* and *Handeliodendron* were strongly supported (BS 100%, PP 1).

3.2. Divergence time estimation

The phylogeny of extant and fossil species shows support for our placement of *A. hickeyii* and *A.* "*magnificum*" as constraints for divergence time dating (Supplementary data 5c). The dating analyses in BEAST using the uniform root calibration of 66.25–89.8 Ma and uniform or lognormal prior distributions for the other three inner nodes showed similar results (Fig. 2 and Supplementary data 6a). The analyses using lognormal or uniform model without root calibration also showed little difference except older nodes below *Aesculus* (e.g., the root fell into the Late Triassic with both models; 231.1 Ma and 236.14 Ma, respectively; Supplementary data 6b and 6c). In the analysis with the root constraint of 66.25–125 Ma, the estimated nodal ages were similar to those from analyses with root constraint of 66.25–89.8 Ma, but the nodes below *Aesculus* were substantially increased (e,g., the root age increased to the Early Cretaceous, 119.06 Ma; Supplementary data 6d). Results from analyses testing whether missing data in the outgroup *Acer* could affect the root age estimation showed that the nodal ages were also similar to those estimated from the analysis with the full dataset (Supplementary data 6f, 6e). There were little difference between the uniform and lognormal models except the nodes below *Aesculus* whose ages were different between the two models; the difference was especially pronounced for the root (lognormal: 213.04 Ma vs uniform: 118.99 Ma).

Results in Fig. 2 indicated a Late Cretaceous crown age of the tribe Hippocastaneae (88.21 Ma, 95% HPD: 83.76–89.8 Ma) and relatively rapid divergence of the three genera (Fig. 2). *Billia* diverged 85.86 Ma (95% HPD: 80.42–89.59 Ma), while the time of split between *Handeliodendron* and *Aesculus* was 78.8 Ma (95% HPD: 73.29–84.17 Ma), both in the Late Cretaceous (66–100.5 Ma). Divergence of section *Calothyrsus* from the rest of *Aesculus* was shown to be 71.39 Ma (95% HPD: 67.92–75.65 Ma) in the late Cretaceous (Fig. 2). Within section *Carothyrsus*, the divergence of the western North American *A. californica* was shown to be 59.79 Ma (95% HPD: 42.65–70.25 Ma), while diversification within the Asian clade was shown to be from 22.13 Ma to 4.66 Ma (Fig. 2). The result also showed a Late Cretaceous divergence (67.48 Ma, 95% HPD: 66.1–70.63 Ma) of the Eurasian section *Aesculus* from the North American clade of sections *Macrothyrsus*, *Parryana* and *Pavia* (Fig. 2). Divergence time between the Japanese *A. turbinata* and the southeastern European *A. hippocastanum* in section *Aesculus* was shown to be 6.27 Ma (95% HPD: 5.33–10.2 Ma) in the Miocene (Fig. 2). Divergence times within the North American clade are estimated at 59.68 Ma (95% HPD: 56–64.14 Ma) in the Paleocene between the eastern North American species *A. parviflora* (section *Macrothyrsus*) and the other two sections, and at 44.42 Ma (95% HPD: 30.37–56.61 Ma) between the western North American species *A. parryi*, comprising section *Parryana*, and the eastern North American section *Pavia* (Fig. 2).

3.3. Biogeographic analyses

The DEC analyses with and without the outgroup of Acereae showed the same results, suggesting that the crown ancestor of Hippocastaneae was widely distributed in eastern Asia, western North America, and Central America (Fig. 3 and Supplementary data 7). The results indicated that the common ancestor of *Handeliodendron* and *Aesculus* was distributed in eastern Asia. Six dispersal events in the biogeographic history of *Aesculus* were inferred (Fig. 3). Combined with information from divergence time dating, the results indicated that *Aesculus* evolved in eastern Asia during the Late Cretaceous and diverged into two lineages, forming a Eurasian-North American clade comprising four sections and an eastern Asian-western North American clade consisting species of section *Calothyrsus*. Within the former clade, there are two subclades. One subclade consists of fossil species *A. longipedunculus* and *A. hickeyi*, and the Eurasian section *Aesculus*. This subclade underwent separate dispersals from EA to Spitsbergen, WNA, and EU. The European species, *A. hippocastanum*, diverged from a Eurasian ancestor of section *Aesculus* during the Miocene. Another subclade included eastern Asian fossil species *A.* "*magnificum*" and three North American sections *Macrothyrsus*, *Parryana*, and *Pavia*. Biogeographic reconstruction indicated that the ancestor of this subclade dispersed from EA to EU and

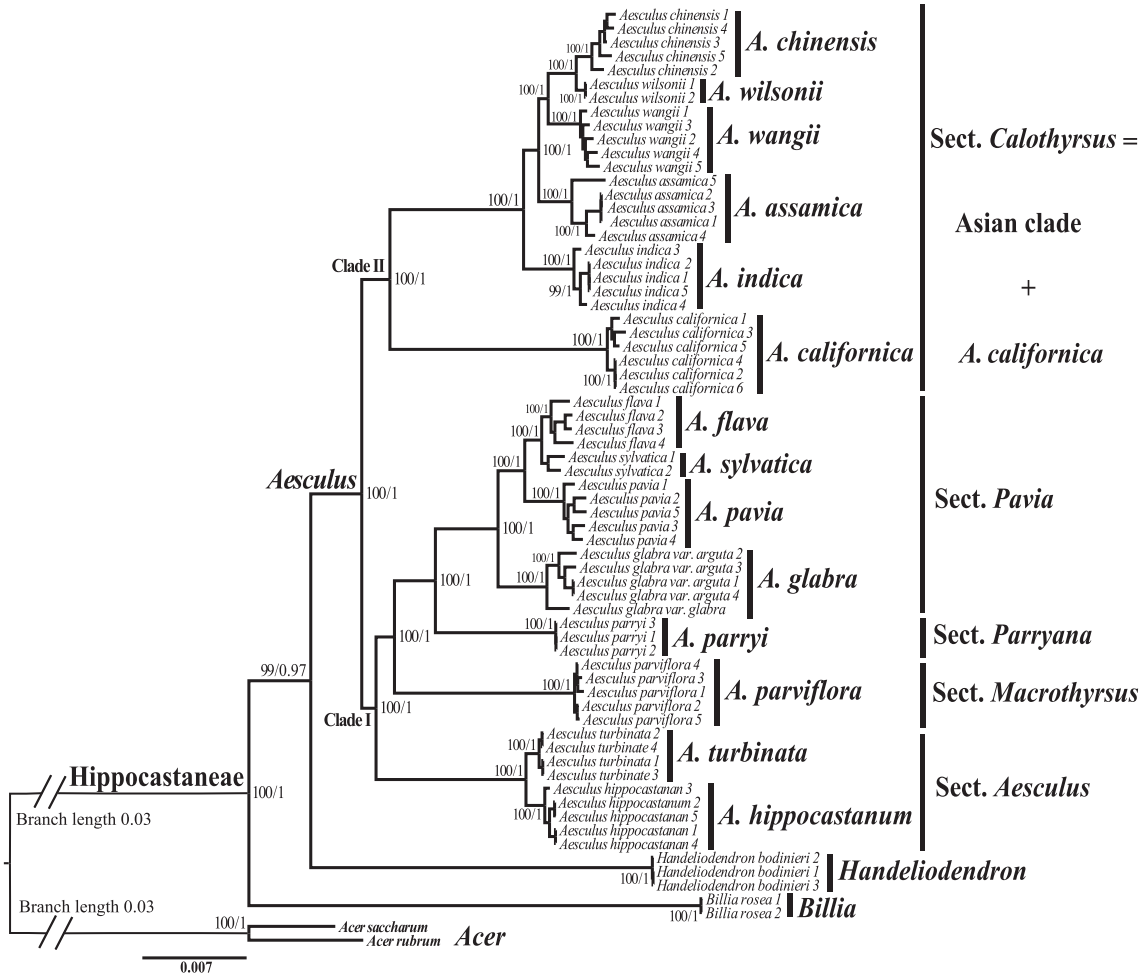

Fig. 1. Phylogenetic tree inferred from RAD-seq Matrix40 for 68 samples of the tribe Hippocastaneae and outgroup species of *Acer* using RAxML. Maximum likelihood bootstrap support values from RAxML and Bayesian posterior probabilities from analysis using MrBayes are indicated at the nodes. The five sections of *Aesculus* of Hardin (1957, 1960) are indicated.

ENA in the Paleocene through the North Atlantic land bridge (NALB). Within this subclade, section *Macrothyrsus* of ENA split in the Paleocene. Section *Parryana* of WNA and section *Pavia* of ENA shared a common ancestor with an EA-EU-ENA range prior to their divergence in the Eocene. Section *Pavia* later diversified within the temperate region of ENA during the Miocene. Another clade from the first split of the genus, section *Calothyrsus*, was shown to have an eastern Asian ancestor. Within this lineage, the western North American *A. californica* diverged from the eastern Asian lineage in the Paleocene following dispersal from EA to WNA through the Bering land bridge (BLB).

3.4. Reconstruction of morphological characters and ecological niches

Results of morphological character analyses are presented in Supplementary data 8, which showed that the ancestor of *Aesculus* was a tree or shrub (character 1) with petiolulate leaves (character 6), completely fused sepals (character 16), and reflexed petals (character 20). Some characters useful for distinguishing sections showed the following evolutionary trends: loss of bud viscidity in three North American sections *Macrothyrsus*, *Parryana* and *Pavia* (character 2), loss of petiolule in sections *Aesculus* and *Parryana* (character 6), growth of red pubescence on young shoots in the Eurasian section *Aesculus* compared to glabrous young shoots in the ancestor (character 8), gain of patches of white trichomes in vein axils in section *Pavia* (character 9), gain of short trichomes on inflorescence peduncle, flower pedicels and outer calyx in eastern Asian species of section *Calothyrsus* (character 10), loss of petal dimorphism in section *Macrothyrsus* (character 19), growth of straight petals in sections *Pavia* (except upper petals) and *Macrothyrsus* compared to reflexed petals in the ancestor (character 20), gain of petal ligule at the base of the petal blade in section *Parryana* (character 23), and gains of curved filaments (character 30) and pistils (character 34) in section *Macrothyrsus* from straight filaments and pistils of the ancestor. Petal grooves were absent in sections *Calothyrsus* and *Parryana* and present in the other three sections, but the ancestral state was uncertain (character 22). The rest of the morphological characters did not show a pattern consistent with resolved clades.

Results of ecological niche reconstructions of the 19 bioclimatic variables from WorldClim and altitude are presented in Supplementary data 9. The evolutionary trends of ecological niches estimated that the ancestor of *Aesculus* lived in an environment with annual mean temperature of 16.08 °C (BIO1), temperature annual range of 23.81 °C (BIO7), mean annual precipitation of 1456.8 mm (BIO12), moderate precipitation seasonality (BIO15), and elevation about 954.6 m (BIO20). Results of phylogenetic signal (Pagel's λ and Blomberg's K) and correlation of the bioclimatic variables (adjusted R^2) are listed in Supplementary data 10. Only two bioclimatic variables (BIO15 precipitation seasonality and BIO19 precipitation of coldest quarter) showed strong phylogenetic signals, while three bioclimatic variables (BIO2 mean diurnal range, BIO14 precipitation of driest month and BIO17 precipitation of driest quarter) exhibited medium levels of phylogenetic signals. Some variables had strong correlations (adjusted $R^2 > 0.8$), such as BIO12 (annual precipitation), BIO13 (precipitation

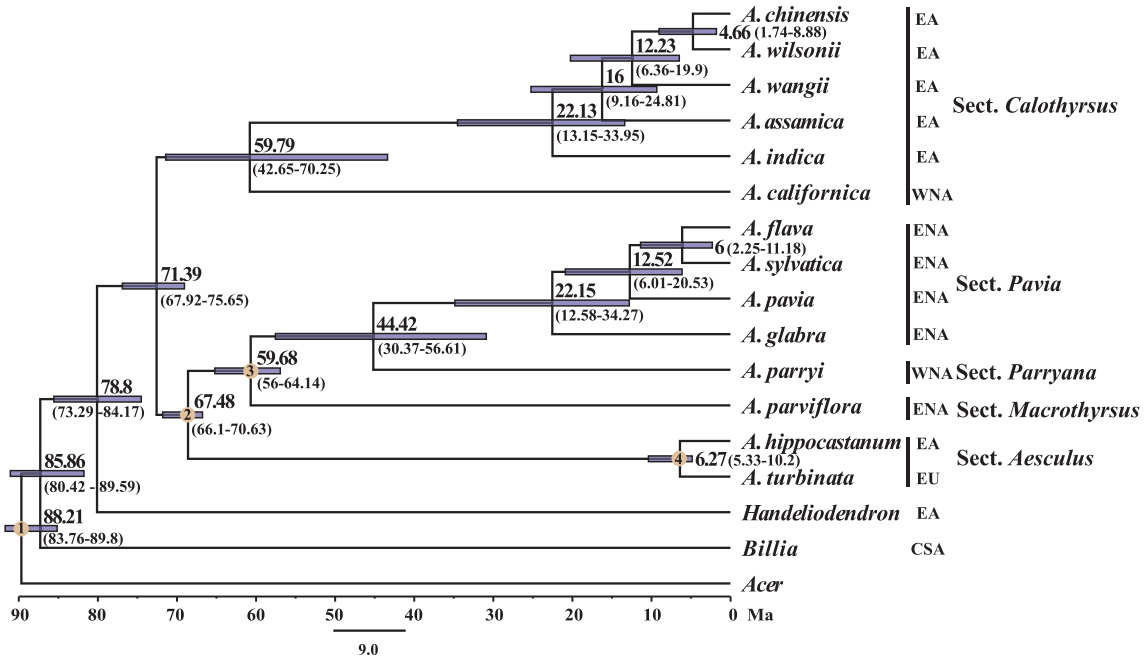

Fig. 2. Time-calibrated phylogeny inferred from RAD-seq Matrix40 for 16 species of the tribe Hippocastaneae and one outgroup species of *Acer* using BEAST. Numbers 1–4 highlighted in brown solid circles indicate fossil calibrated nodes set with uniform prior distributions. Median values of divergence time in millions of years are shown in bold faces at each node. Blue bars and values in parentheses indicate the 95% highest posterior density values (HPD). Distributions of the taxa are indicated as follow: EU = Europe, EA = eastern Asia, ENA = eastern North America, WNA = western North America, and CSA = Central and South America. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of wettest month), BIO16 (precipitation of wettest quarter), and BIO18 (precipitation of warmest quarter), or BIO14 and BIO17 (precipitation of driest month and driest quarter).

Results of phylogenetic signal of each trait (Pagel's λ and Blomberg's K), and correlation between bioclimatic variables and morphological characters (adjusted R^2) are listed in Table 2. Most traits showed strong phylogenetic signals (Pagel's $\lambda = 1.00$, $P < 0.001$), and in general their Blomberg's K values were above one. However, trait1 (habit), trait 26 (glandular trichomes on petal margin), trait 32 (trichomes on anther), and trait 33 (pistil length) had low phylogenetic signals. When comparing the results from morphological features and environmental factors, we observed co-occurrences of several transitions of morphological and niche characters on different phylogenetic branches: (1) Increase of elevation in the eastern Asian species of section *Calothylsus* (Fig. 4) and gains of short trichomes on inflorescence peduncle, flower pedicels and outer calyx (character 10, state 1) (adjusted $R^2 = 0.41$, *p*-value < 0.01); decrease of elevation in the North American clade of sect. *Macrothyrsus*-sect. *Parryana* -sect. *Pavia* (Fig. 4) and loss of bud viscidity (character 2, state 0) (adjusted $R^2 = 0.51$, *p*-value < 0.01). (2) Decrease of annual mean temperature (BIO1, Fig. 5) and gains of red pubescence on young shoot (character 8, state 1) (adjusted $R^2 = 0.36$, *p*-value < 0.05), and petal protuberances holding stamens (character 24, state 2) (adjusted $R^2 = 0.36$, *p*-value < 0.05), on the branch of section *Aesculus*, the two species in Europe and Japan. (3) Decrease of annual precipitation on the branch of the Baja Californian species, *A. parryi* (section *Parryana*; BIO12, Fig. 6) and growth of petal ligule (character 23, state 1) (adjusted $R^2 = 0.75$, *p*-value < 0.001), as well as a change to make sepals fused to only half of its length in the species (character 16, state 1) (adjusted $R^2 = 0.74$, *p*-value < 0.001). (4) Increase of precipitation seasonality in section *Calothyrsus* and section *Parryana* (BIO15, Fig. 7) and absence of petal groove in each of these lineages, respectively (character 22, state 0) (adjusted $R^2 = 0.51$, *p*-value < 0.01). The above transitions of morphological characters and ecological niches co-occurred in some sections, with moderate to strong correlations between these bioclimatic variables and morphological traits in the phylogeny. Besides the aforementioned traits, several other morphological characters (e.g. traits 1, 14, 20, and 33) also showed some correlation with niche variables in the phylogeny, but not within a specific section.

4. Discussion

4.1. Relationships among genera of the tribe Hippocastaneae and utility of RAD-seq data for resolving Late Cretaceous divergences

Relationships among Hippocastaneae, comprising *Aesculus*, *Handeliodendron*, and *Billia*, have remained somewhat uncertain. A closer relationship between *Aesculus* and *Billia*, compared to *Handeliodendron*, had been supported by observations of morphology (Rehder, 1935) and phylogenetic analyses of morphological characters (Judd et al., 1994; Forest et al., 2001), whereas molecular phylogeny showed low supports (Harrington et al., 2005; Harris et al., 2016). Alternatively, a molecular phylogenetic study of Sapindaceae resolved *Aesculus* as paraphyletic with *Billia* and *Handeliodendron* nested within (Buerki et al., 2010) but was limited by considerable missing data for representative species of Hippocastaneae. In contrast to the previous hypotheses, our study with genome-wide molecular data from RAD-seq supported a sister relationship of *Aesculus* and *Handeliodendron* (Fig. 1). This relationship was strongly supported by our phylogenetic analyses of different datasets using both ML and Bayesian Inference methods (e.g., M40 with 2668 loci and 68 samples: BS 99%, PP 0.97, Fig. 1; M40 with 2668 loci and 17 samples: BS 97%, PP 1, Fig. S1; and M40 with 340 loci only present in the outgroup *Acer*: BS 100%, PP 1, Fig. S2).

The relationship revealed in our study was somewhat surprising given that *Handeliodendron* was only recently allied with *Billia* and *Aesculus* and is distinct from *Billia* and *Aesculus* by a number of morphological characters, such as leaflets with glandular spots, petals with appendages, absence of a petal claw, sessile glandular hairs on the ovary and outer calyx, and having a petiolate ovary, or gynophore (Acevedo-Rodriguez et al., 2011). We questioned if the sister relationship between *Aesculus* and *Handeliodendron* might be an artifact of long-branch attraction biased by potential saturation of nucleotide

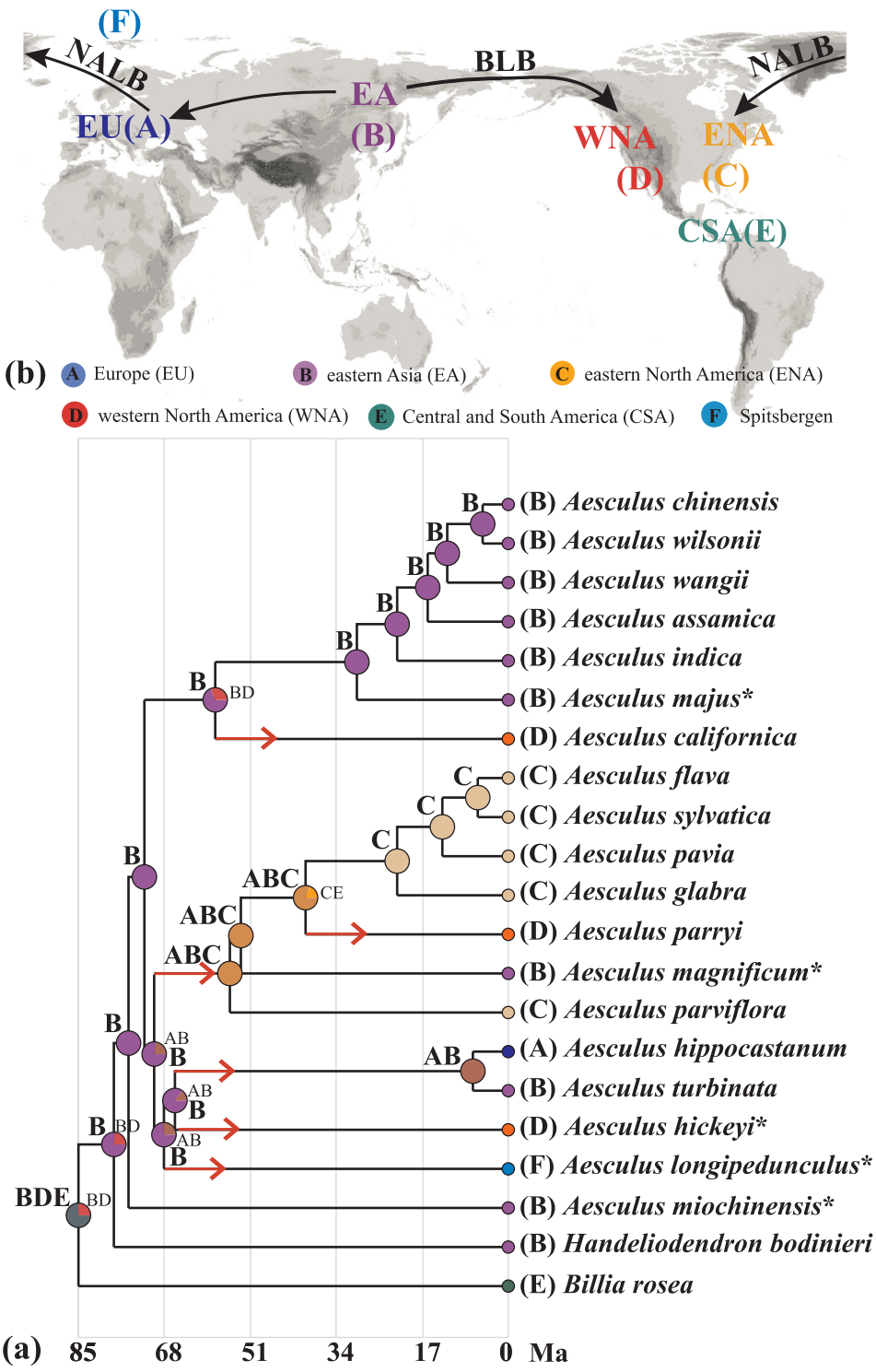

Fig. 3. (a) Reconstructed ancestral distributions from DEC analysis in RASP using dated phylogeny including fossils. The outgroup, *Acer*, is removed from the phylogeny. Distribution ranges are indicated as follow: A = Europe (EU), B = eastern Asia (EA), C = eastern North America (ENA), D = western North America (WNA), E = Central and South America (CSA), F = Spitsbergen. Asterisks indicate fossil species. Pie charts represent the marginal probabilities for each alternative ancestral area. The ancestral areas at each node are indicated by letters. Arrows indicate the direction of dispersal events. The time scale is in millions of years. **(b)** Map showing distribution areas and dispersal routes. BLB = the Bering land bridge; NALB = the North Atlantic land bridge.

substitutions in the RAD-seq data. Evidence suggests that this is unlikely. The branches uniting the two genera are actually short (Fig. 1). The RAD-seq data contained DNA sequences of several thousand loci (fragments) that represent various parts of the genome, including many coding genes, which are expected to evolve relatively slowly. Compared to previous studies, the present study included all species and multiple accessions for each species of *Aesculus*, and much more nuclear data for resolving the phylogeny. Thus, the phylogenetic relationships inferred from the RAD-seq data have better resolution and support. Furthermore, *Aesculus* and *Handelidendron* also share some morphological and ecological characteristics; both are temperate species and have deciduous palmate leaves, typically with five or more leaflets, a denser and leafless thyrsoid inflorescence, connate sepals, and pilose stamens, whereas *Billia* is a tropical genus bearing evergreen leaves, usually with only three leaflets, and having a less dense and leafy cymose inflorescence, free sepals, and glabrous stamens (Hardin, 1957a, 1957b; Judd et al., 1994; Forest et al., 2001; Manchester, 2001; Xia et al., 2007). Additionally, sticky buds are present in section *Calothyrsus* and section *Aesculus* (Hardin, 1957a), representing ancestral state in *Aesculus* (Supplementary data 8b), and reportedly also present in *Handeliodendron* (Yuan Xu, personal communication, South China Botanical Garden). These shared characteristics corroborate the RAD-seq data

Table 2
Results of phylogenetic signal tests (Pagel's λ and Blomberg's K) for each character and adjusted R^2 for correlation between bioclimatic variables and morphological characters. Bioclimatic variables that do not have correlation with morphological characters are not shown. *P*-values: * < 0.05, ** < 0.01, *** < 0.001.

No.	λ	K	BIO1	BIO2	BIO3	BIO4	BIO5	BIO7	BIO9	BIO12	BIO13	BIO14	BIO15	BIO16	BIO17	BIO18	Altitude
trait1	0.53	0.36	0.15	0.35*	0.21	−0.08	0.25*	−0.07	0.30*	0.02	0.17	−0.05	−0.04	0.21	−0.06	0.28*	0.47**
trait2	1.00***	2.26***	−0.02	−0.03	−0.03	−0.01	0.36*	0.02	−0.02	−0.05	0.32*	−0.07	0.05	0.32*	−0.07	−0.08	0.51**
trait6	1.00***	1.15**	0.12	−0.05	−0.08	−0.01	0.02	−0.08	−0.06	0.23*	0.15	0.19	0.11	0.11	0.19	0.07	−0.03
trait8	1.00***	1.80**	0.36*	0.00	0.08	−0.07	0.23*	−0.06	0.02	−0.05	−0.06	−0.04	−0.08	−0.05	−0.04	−0.06	−0.03
trait9	1.00***	2.04**	−0.08	−0.07	−0.06	0.14	0.16	0.11	−0.08	0.06	−0.04	0.21	0.20*	−0.03	0.21	0.04	0.18
trait10	1.00***	3.42***	−0.06	0.35*	−0.07	0.00	0.07	0.16	−0.06	−0.06	0.25*	−0.08	−0.03	0.27*	−0.08	0.08	0.41**
trait11	1.00***	1.04**	0.05	−0.06	−0.01	0.01	−0.05	−0.01	−0.07	0.05	−0.03	0.09	0.26*	0.01	0.11	−0.02	−0.08
trait14	1.00***	2.56***	0.02	0.57***	0.06	−0.03	0.52**	0.05	0.05	−0.01	0.31*	−0.06	−0.06	0.35*	−0.05	0.11	0.58***
trait15	1.00***	1.18**	0.16	−0.07	−0.06	−0.08	−0.04	−0.02	0.01	0.11	0.08	0.02	0.18	0.09	0.00	−0.03	−0.07
trait16	1.00***	2.56**	−0.05	−0.08	0.30*	0.16	−0.06	−0.01	−0.01	0.74***	0.57**	0.61*	0.23*	0.51*	0.60***	0.31*	−0.08
trait19	1.00***	2.31**	−0.05	−0.07	−0.08	−0.05	−0.05	−0.07	−0.01	0.01	−0.07	0.01	0.06	−0.08	0.02	−0.05	0.12
trait20	1.00***	1.43***	−0.13	−0.16	−0.16	0.12	0.18	0.06	−0.11	0.17	−0.10	0.42*	0.45*	−0.11	0.43*	0.04	0.48*
trait22	1.00***	1.57**	0.02	−0.07	0.06	0.22*	−0.08	0.23*	−0.06	0.35*	0.04	0.68***	0.51**	0.03	0.69***	0.19	0.25*
trait23	1.00***	1.23	−0.05	−0.08	0.30*	0.16	−0.06	−0.01	−0.01	0.75***	0.57**	0.61*	0.23*	0.51**	0.60***	0.31*	−0.08
trait24	1.00***	1.52**	0.36*	−0.09	0.11	−0.01	0.18	0.10	0.09	−0.15	−0.16	−0.13	−0.12	−0.14	−0.13	−0.14	−0.11
trait25	1.00***	0.98*	−0.06	−0.08	−0.08	0.00	0.08	−0.04	−0.05	0.01	−0.06	0.03	0.31*	−0.02	0.05	−0.04	−0.05
trait26	0.24	0.27	−0.06	−0.08	0.09	−0.06	0.08	−0.08	−0.01	0.10	0.27*	−0.05	−0.05	0.23*	−0.05	−0.03	−0.06
trait29	1.00***	1.02*	−0.02	−0.07	0.06	−0.08	0.15	−0.08	0.01	0.02	0.29*	−0.04	−0.04	0.21	−0.05	−0.03	−0.06
trait30	1.00***	0.85	−0.05	−0.07	−0.08	−0.05	−0.04	−0.07	−0.01	0.01	−0.07	0.01	0.06	−0.07	0.02	−0.05	0.12
trait32	0.00	0.49	0.26*	−0.03	−0.07	−0.06	0.02	0.02	−0.06	0.08	0.03	−0.07	0.01	0.11	−0.07	−0.03	−0.07
trait33	0.00	0.44	0.15	−0.01	−0.03	0.00	−0.05	0.66***	−0.01	0.11	0.16	−0.08	−0.06	0.28*	−0.05	−0.05	−0.08
trait34	1.00***	2.31**	0.06	−0.07	−0.08	−0.05	−0.05	−0.07	−0.01	0.01	−0.07	0.01	0.06	0.04	0.02	0.05	0.12

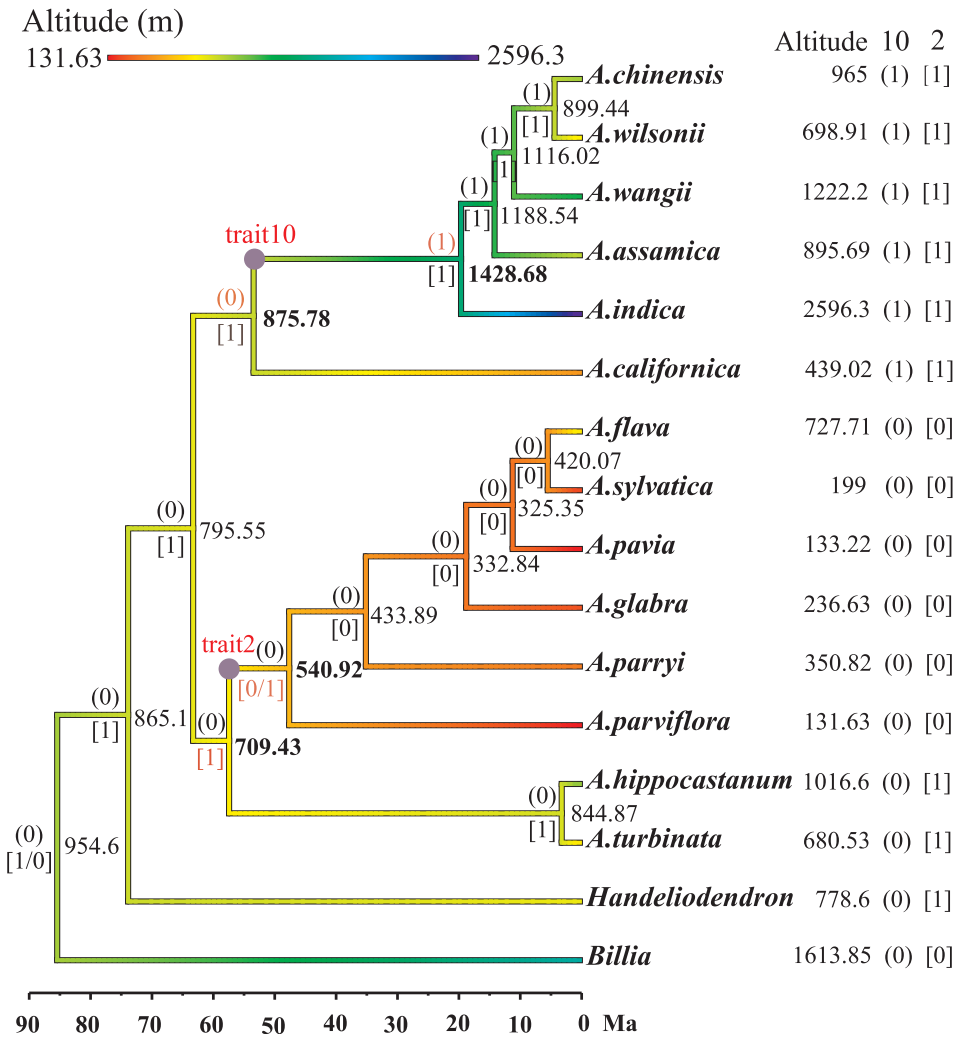

Fig. 4. Reconstructed evolutionary trends of altitude, character 10 (short trichomes on inflorescence peduncle, flower pedicels, and outer calyx), and character 2 (bud viscidity). Numbers at nodes and after terminal taxon names are mean altitudes in meters above sea level. Numbers 0, 1 in parentheses and square brackets are character states of character 10 and character 2, respectively (0, absent; 1, present). Numbers separated by a forward slash indicate uncertainty between the two character states. Purple markers on branches indicate shift of character states of the indicated character (2 or 10). Changes in branch color signify transitions in altitude. Shifts are interpolated along branches according to Felsenstein (1985) and annotated in color. The figure shows increase of elevation in the eastern Asian clade of section *Calothyrsus* is associated with gains of short trichomes on peduncle, pedicel and calyx. Decrease of elevation in the clade of three North American sections co-occurs with loss of bud viscidity. The time scale is in millions of years. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

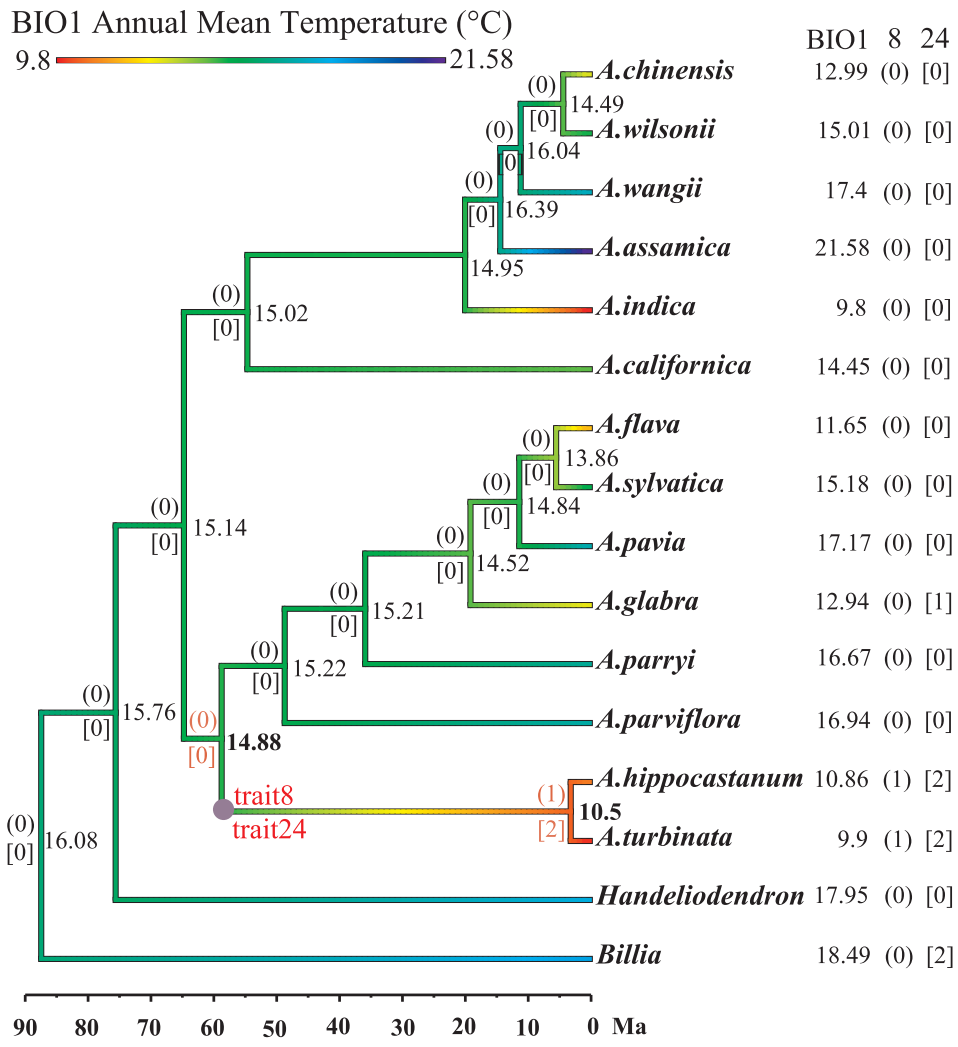

Fig. 5. Reconstructed evolutionary trends of annual mean temperature (BIO1), character 8 (red pubescence on young shoot), and character 24 (petal protuberances). Numbers at nodes and after terminal taxon names are mean annual mean temperature in Degrees Celsius. Numbers 0, 1, and 2 in parentheses and square brackets are character states of character 2 and character 24, respectively (0, absent; 1, present; 2, present, enveloping stamens). Purple marker on branch indicates shift of character states of the indicated character (8 or 24). Changes in branch color signify transitions in annual mean temperature. Shifts are interpolated along branches according to Felsenstein (1985) and annotated in color. The figure shows decrease of annual mean temperature in section *Aesculus* is associated with gains of red pubescence on young shoot and petal protuberances holding stamens. The time scale is in millions of years. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

supporting a close relationship between the two genera. The morphological features shared by *Aesculus* and *Billia* may represent pleisiomorphies of the tribe inherited from a sapindaceous ancestor. Nonetheless, we inferred short time intervals between the successive divergence events separating *Billia*, *Handeliodendron*, and *Aesculus*, suggesting a rapid divergence of the three genera (Fig. 2).

RAD-seq data have been considered suitable for evolutionary studies at low taxonomic levels (below species or among closely related species) (see reviews in McCormack et al., 2013; Yu et al., 2018 and case studies in Boucher et al., 2016; Cavender-Bares et al., 2015; Eaton and Ree, 2013; Hipp et al., 2018; Hohenlohe et al., 2010; Pais et al., 2016; Qi et al., 2015; Rubin et al., 2012; Taranto et al., 2016; Zhou et al., 2018). A major concern with application of RAD-seq data for deeper evolutionary history is too much missing data due to allele "drop off" (i.e., mutation at the restriction site due to deep divergence, leading to a small number of homologous fragments for comparison across all taxa). Indeed, our study showed that most of the loci shared among species of *Aesculus* were missing in *Acer*. However, our results demonstrated that RAD-seq data can be valuable for resolving phylogenetic relationships among plant lineages in Hippocastaneae that diverged in the Late Cretaceous, adding to previous support for the use of RAD-seq for resolving groups of at least Paleocene age (Cariou et al., 2013; Leaché and Oaks, 2017; Rubin et al., 2012). Even with a dataset that was reduced from several thousand loci to the set of 340 loci detected in the outgroup *Acer*, a robust phylogeny of Hippocastaneae was still supported. This result suggests that allele drop-off is likely hierarchical and do not significantly impact phylogenetic inference, as also shown in DaCosta and Sorenson (2016). Our result also demonstrates that large dataset comprising many nuclear markers such as RAD-seq data can overcome noises associated with missing data among distantly related organisms, as also pointed out in Herrera et al. (2015). The present study supports the use of RAD-seq for resolving phylogenies among taxa diverged in the Late Cretaceous and Paleogene.

Our study also suggests that divergence time estimation is not significantly affected by using uniform or lognormal model, if the constrained time range was applied consistently and properly (Supplementary data 6). However, our study showed that proper calibration of the root node in the phylogeny is critical to obtain reasonable estimation of deep nodes. Our analyses indicated that the deep nodes were overestimated for their ages if a root calibration was not applied or was constrained with a maximum bound that is too old (e.g., using *Archifrutus*'s age), although the upper nodes (i.e, from *Aesculus* and above) were little affected. Furthermore, our study showed that vast amount of missing data in the outgroup *Acer* was not a concern on phylogenetic and divergence time dating analyses, supporting again the value of RAD-seq data for deep phylogenetic reconstruction.

4.2. Monophyly of section Calothyrsus and relationships among sections of Aesculus

Previous molecular phylogenetic studies using one or a few chloroplast and nuclear genes or gene regions were not able to completely resolve the phylogenetic relationships within *Aesculus* (Xiang et al., 1998; Harris et al., 2009a; Buerki et al., 2010; Harris et al., 2016). One

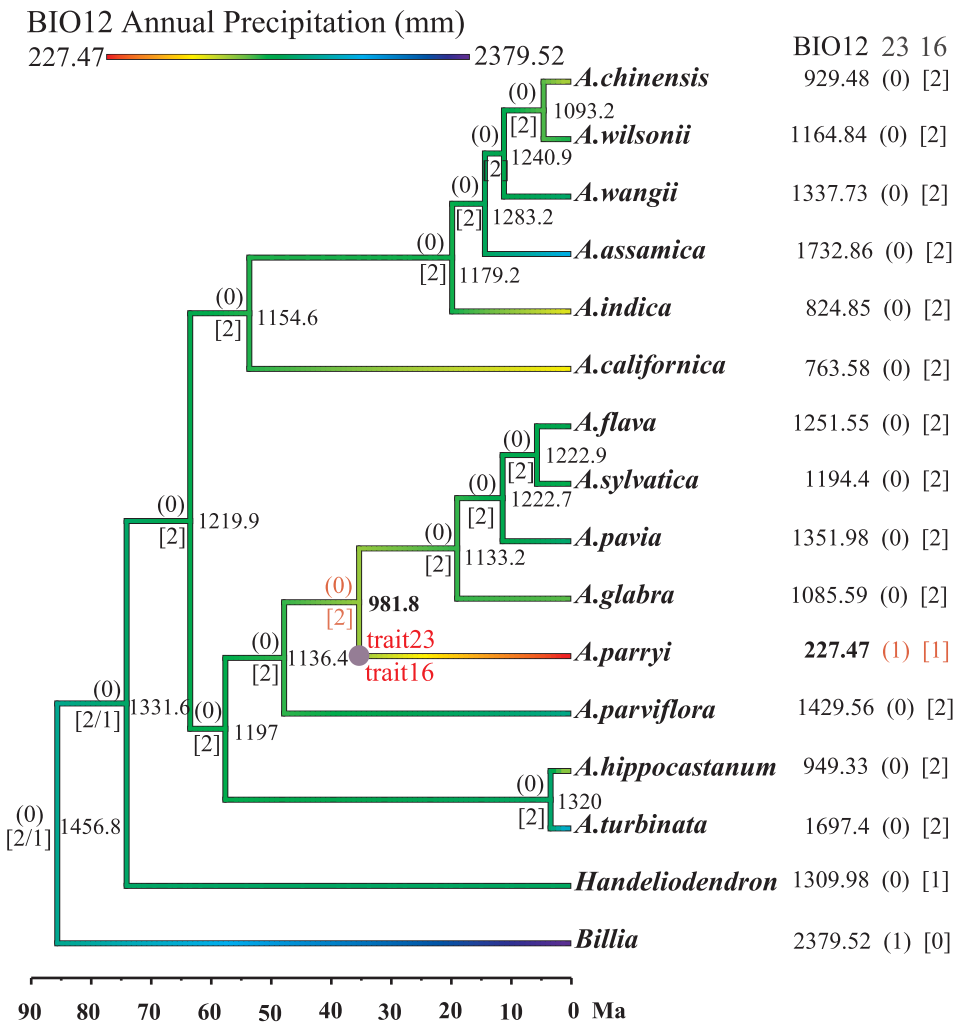

Fig. 6. Reconstructed evolutionary trends of annual precipitation (BIO12), character 23 (petal ligule), and character 16 (fusion of sepals). Numbers at nodes and after terminal taxon names are mean annual precipitation in millimeters (mm). Numbers 0, 1, and 2 in parentheses or square brackets are character states of character 23 and character 16, respectively (Character 23: 0, absent and 1, present; Character 16: 0, fused at the base; 1, fused to half length; 2, almost completely fused). Numbers separated by a forward slash indicate uncertainty between the two character states. Purple marker on branch indicates shift of character states of the indicated character (23 or 16). Changes in branch color signify transitions in annual precipitation. Shifts are interpolated along branches according to Felsenstein (1985) and annotated in color. The figure shows decrease of annual precipitation in *A. parryi* co-occurred with origin of petal ligule and a change from almost completely fused sepal to half-length fusion of sepals. The time scale is in millions of years. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of the major remaining problems involved the placement of *A. californica* and the monophyly of section *Calothyrsus*, which includes *A. californica*. Molecular phylogenies based on chloroplast data supported the placement of *A. californica* as sister to the rest of section *Calothyrsus*, thus supporting the monophyly of the section. However, sequences of ITS and *LFY* resulted in tree topologies with *A. californica* sister to *A. parviflora* (section *Macrothyrsus*), therefore rendering section *Calothyrsus* nonmonophyletic (Xiang et al., 1998; Harris et al., 2009a). Our RAD-seq data corroborated the plastid DNA relationship with strong support (Fig. 1), therefore supporting the monophyly of section *Calothyrsus* and, consequently, the taxonomic view of Hardin (1957a, 1957b, 1960). The RAD-seq phylogeny shows divergence of the section from the remainder of the genus during the Late Cretaceous (71.31 Ma; Fig. 2), and *A. californica* diverged from the Asian species of this section in the Paleogene (58.28 Ma; Fig. 2). *Aesculus californica* shares with the Asian species of section *Calothyrsus* in having sticky buds (resinous bud scales) and a smooth fruit surface (Hardin 1957a, 1957b, 1960). The section represents an eastern Asian-western North American disjunct distribution that originated in the Paleogene time (Fig. 2).

Our present study also provides new insights into relationships among the five sections of *Aesculus*. Previous studies have resolved various relationships among the five sections and mostly with low or moderate support (Xiang et al., 1998; Forest et al., 2001; Harris et al., 2009a, 2016), except a sister relationship between section *Pavia* and section *Parryana* which was strongly supported in the recent molecular studies of Harris et al. (2009a, 2016). Our results strongly support that the Asian-western North American section *Calothyrsus* is sister to the remainder of the genus and the Japanese-European section *Aesculus* is sister to the North American clade of three sections. Within the North American clade, the southeastern US section *Macrothyrsus* is sister to a clade consisting of the western North American section *Parryana* and eastern North American section *Pavia*. These sectional relationships are different from all previous findings and are strongly supported in our study.

Relationships within all polytypic sections are also well resolved in our study. Within section *Pavia*, our finding concurs with those of previous studies (Hardin, 1957a, 1957b; Xiang et al., 1998; Harris et al., 2009a) that *A. glabra* Willd. is sister to a clade including *A. pavia* L., *A. sylvatica* Bart., and *A. flava* Ait. *Aesculus glabra* differs from other members of section *Pavia* by having petals that are more-or-less of equal lengths and upper petals with an inconspicuous limb-claw morphology. Within section *Calothyrsus* the species status of *A. wilsonii* and *A.* "*wangii*" was questioned (Xiang et al., 1998; Turland and Xia, 2005; Xia et al., 2007). In Flora of China (Xia et al., 2007), *A. wilsonii* was treated as a wild variety of the cultivated species, *A. chinensis*, based on the existence of extensive intermediate forms, and *A.* "*wangii*" was merged into *A. assamica* on account of morphological similarity and lacking valid publication. Our results showed that samples of *A. chinensis* and *A. wilsonii* formed mutually monophyletic clades, which were sister to one another (Fig. 1). They split 4.91 Ma during the Pliocene (Fig. 2). Samples of *A.* "*wangii*" formed a subclade sister to *A. chinesis-A. wilsonii*. *Aesculus assamica* diverged from the clade *A.* "*wangii*"-*A. chinesis-A. wilsonii* in 16.08 Ma during the Miocene. The closer relationship of *A.* "*wangii*" to *A. wilsonii* and *A. chinensis* than to *A. assamica*, into which it

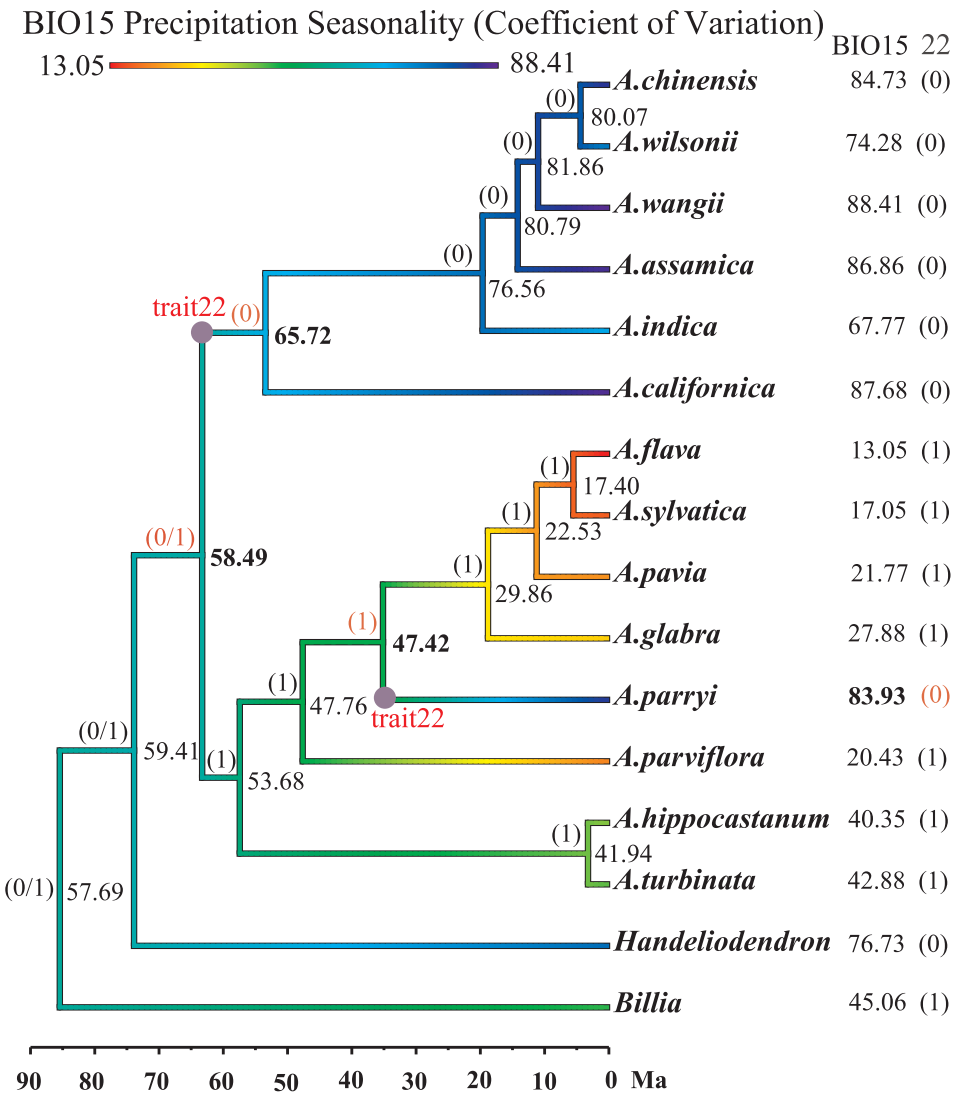

Fig. 7. Reconstructed evolutionary trends of precipitation seasonality (BIO15) and character 22 (petal groove). Numbers at nodes and after terminal taxon names are mean coefficient of variation of precipitation seasonality. Numbers 0 and 1 in parentheses are character states of character 22 (0, absent; 1, present). Numbers separated by a forward slash indicate uncertainty between the two character states. Purple markers on branches indicate shift of character states of character 22. Changes in color on the tree signify transitions in precipitation seasonality. Shifts are interpolated along branches according to Felsenstein (1985) and annotated in color. The figure shows increase of precipitation seasonality in Sections *Calothyrsus* and *Parryana* is associated with fixation of petal groove absence and loss of petal groove, respectively. The time scale is in millions of years. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

was merged (Xia et al., 2007), is consistent with Fang (1960), who regarded *A.* "*wangii*" as similar to *A. wilsonii* in his original Latin description. Our results clearly support the distinction of *A. wilsonii* from *A. chinensis* and *A.* "*wangii*" from *A. assamica* and suggest the importance of validly publishing *A.* "*wangii*". The intermediate forms among *A. wilsonii* and *A. chinensis* may represent gene flow owing to their long history of horticultural cultivation in China (Hardin, 1960; Turland and Xia, 2005) as well as to the overall commonality of introgression throughout *Aesculus*, even among long-diverged lineages (Hardin, 1957c).

4.3. Biogeographic history of Aesculus and Hippocastaneae

Previous biogeographic analyses have resolved an eastern Asian (Xiang et al., 1998) or Beringian origin of *Aesculus* (Harris et al., 2009a). Our biogeographic analyses of the present study also suggested an eastern Asian origin of *Aesculus* (Fig. 3), congruent with the view of Xiang et al. (1998) that the genus likely evolved as a "boreotropical flora" element in the Late Cretaceous or Paleocene at the high latitudes of eastern Asia and diversified there. The present study does not support an American origin of *Aesculus* from a "*Billia*-like" ancestor as proposed in Hardin (1957a); Forest et al. (2001). Although the present study suggested an Asian origin of the genus, it also supported early dispersals in the Late Cretaceous and Paleogene from eastern Asia to North America via the Bering land bridge and the North Atlantic land bridge (Fig. 3). This hypothesis is supported by the results of divergence time and biogeographic analyses, as well as the ages and distributions of the oldest fossils of *Aesculus*, which include *A. hickeyi* from the Paleocene of North America (to 66 Ma) with dubious occurrences dating to the Late Cretaceous (Maastrichtian, to 72.1 Ma; Manchester, 2001), *A.* "*magnificum*" from the Paleocene of Kamchatka in northeastern Asia (Budantsev, 1992; Manchester, 2001), and *A. longipedunculus* from the Eocene of Spitsbergen in the Greenland region (Schloemer-Jäger, 1958; Golovneva, 2000). These fossils occurred in regions north of the present distribution of *Aesculus* or at the northern border of the present range (Harris et al., 2014). The Bering land bridge and the North Atlantic land bridge were available during the Maastrichtian and the Paleocene (Graham, 2018; Manchester and Tiffney, 2001; Sanmartín et al., 2001; Tiffney, 1985b; Wen et al., 2016) when two dispersal events from EA to WNA (*A. hickeyi* and *A. californica*) and one dispersal event from EA to the Greenland region (*A. longipedunculus*) occurred. The North Atlantic land bridge was also available when the genus dispersed from EA to EU and ENA before the Eocene (Graham, 2018; Manchester and Tiffney, 2001; Tiffney, 1985b) to form a wide spead ancestor giving rise to section *Microthysus* (*A. parviflora*), *A. magnificum*, section *Parryana* (*A.parryi*), and section *Pavia* (*A. glabra*, *A. flava*, *A. pavia*, and *A. sylvatica*; Fig. 3). Dispersal from EA to EU between the Paleocene and Miocene to form a Eurasian ancestor of section *Aesculus* (*A. hippocastanum* and *A. turbinata*) was possible after the retreat of the Turgai Strait by the end of the Eocene (Manchester and Tiffney, 2001).

Our biogeographical analyses suggested that the crown ancestor of Hippocastaneae was widely distributed in Eastern Asia, Western North America, and Central America and the common ancestor of *Handeliodendron* and *Aesculus* was likely in EA in the Late Cretaceous (Fig. 3). Following their divergence, *Aesculus* and *Handeliodendron* might have dispersed southward within EA from the northern regions through Central China to the sino-Himalayan region and southern China. These inferences are congruent with our understanding of the geological and vegetational history of the Northern Hemisphere. During the Late Cretaceous, the Northern Hemisphere, or Laurasia, was divided into two landmasses: EA-WNA and EU-ENA (see Tiffney, 1985b), while EA was separated from EU by Turgai Strait and other seaways through central Asia and eastern Europe (Vinogradov, 1967–1968), and WNA and ENA were separated by the Midcontinental Seaway through central North America. These landmasses facilitated vegetation provinces over which species, such as the ancestor of Hippocastaneae, may have become widespread before the provinces disintegrated at the end of the Cretaceous (Muller, 1970). As *Aesculus* migrated south, it diversified, giving rise to the extant species of sect. *Calothyrsus* (i.e., *A. indica*, *A. assamica*, *A.* "*wangii*", *A. wilsonii,* and *A. chinensis*). During the globally warm period within the Paleocene through parts of the Middle Eocene, the boreotropical flora spread over the Northern Hemisphere (Budantsev, 1992; Wolfe, 1975), and turned into boreal temperate flora in the Angaro-Beringia region following the climatic cooling from the Middle Eocene through the Eocene-Oligocene transition (Budantsev, 1992; Prothero, 1994; Tiffney, 1985a). During the Eocene-Oligocene climatic transition, temperatures at high latitudes decreased by c. 3–5 °C (Liu et al., 2009), pushing the thermophilic trees southward in distributions (Hipp et al., 2018). Further cooling in the Neogene and Quaternary caused many extinctions and forced species within the mixed mesophytic forest to disperse further southward, where some survived in a few low latitudinal "refugia" in eastern and western Asia, southeastern Europe, eastern and western North America, and sometimes within Central and South America (Graham, 1999; Tiffney, 1985a; Wolfe, 1975). The Eurasian lineage of section *Aesculus* might have spread over the temperate regions of Eurasia in the Neogene (after the Turgai Strait receded during the Oligocene), and survived the Quaternary ice age in two disjunct refugia in Japan and southeastern Europe. In contrast to *Aesculus*, *Handeliodendron* may have gone extinct in the temperate regions, surviving only in a narrow subtropical area of southern China. Isolation of endemic and monotypic lineages in the subtropics of eastern Asia is a common biogeographic pattern (e.g. *Euscaphis* Siebold & Zucc., *Davidia* Baill., and *Dipteronia* Oliv.; see Manchester et al., 2009), although many such lineages once had wider, intercontinental distributions based on their fossil records before becoming isolated. Unfortunately, there are no known fossils of *Handeliodendron* except as reported in Wu et al. (2017) from central Tibet, but that fossil appears araliaceous, not sapindaceous and is unlikely to represent *Handeliodendron*. Thus, the degree to which the biogeographic history of *Handeliodendron* is consistent with this well-known biogeographic pattern remains largely unknown. As proposed by Raven and Axelrod (1974), Hippocastaneae may not have arrived in South America prior to the late Miocene or Pliocene when global climatic cooling may have resulted in the southward migration of *Billia* from southern Mexico and Central America to northern South America via the Panama Isthmus that emerged in the very late Neogene.

The biogeographic history of *Aesculus* inferred in the study involved at least two dispersals from EA to WNA, two from EA to EU, one from EU to ENA, and one from EA to Spitsbergen probably through EU. The Baja Californian species, *A. parryi*, arrived at WNA likely via dispersal from ENA or from EA. There were two intercontinental disjunctions in extant *Aesculus*, one eastern Asian-western North American disjunction of section *Calothyrsus* species in the Paleocene and one Eurasian-North American disjunction of section *Aesculus* and three American sections in the Late Cretaceous. These disjunctions probably occurred following dispersals across the Bering land bridge and the North Atlantic land bridge. The ability of *Aesculus* to exist in Beringia is supported by the presence of *A.* "*magnificum*" in Kamchatka (Budantsev, 1992; Manchester, 2001), part of the Beringian region. As *Aesculus* bears large fruits that are toxic to potential large animal dispersers (e.g., Cary, 1922), the genus very likely had to establish within Beringia (Edwards et al., 2018) in order to cross the subcontinent as a land bridge between EA and WNA. The existence of *Aesculus* in the North Atlantic land bridge is supported by the fossil species, *A. longipedunculus*, in Spitsbergen (Schloemer-Jager, 1958), part of the Greenland region. Our study highlights the importance of evaluating fossil species that occur in areas once comprising land bridges (Graham, 2018) within a broader taxonomic context using related fossil species, phylogeny, and historical biogeographic reconstructions.

One notable divergence within *Aesculus* is the late Miocene divergence of *A. hippocastanum* from *A. turbinata* (Fig. 2). Based on fossil evidence, *Aesculus* was well-established in Europe within the Miocene, and most fossils are closely related to *A. hippocastanum* (Velitzelos and Gregor, 1990; Mijarra et al., 2008). Based on our results, some of these fossils may represent species that occurred prior to the divergence from *A. turbinata*, and a re-examination of the fossils with this possibility in mind might yield new insights into the past biodiversity of section *Aesculus* and the European flora.

Meta-analyses of eastern Asian-North American or eastern Asian-western North American disjunct genera showed that in a majority of these genera, the disjunction events occurred in the Neogene period (Miocene and Pliocene) and involved out-of-Asia migrations (Donoghue and Smith 2004; Wen et al., 2010; Harris et al., 2018). However, in *Aesculus*, disjunction in the genus occurred much earlier based on our study and prior studies, thus representing less common ancient events. One explanation for the aberrance of the *Aesculus* disjunction is its large fruits, which have a very limited dispersal distance (Hoshizaki et al., 1999; Mendoza and Dirzo, 2009) making isolation of populations during migration more likely. In contrast, other plant lineages with small, dry or fleshy fruits may have maintained gene flow among migrating populations for longer times and experienced gene flow across barriers due to repeated non-random dispersals (Harris et al., 2018). This scenario is consistent with the hypothesis that the eastern Asian-North American floristic disjunction resulted from the fragmentation of a once widely distributed mesophytic forest of the northern hemisphere that was derived from the Paleogene boreotropical flora (Tiffney, 1985a; Wolfe, 1975), and the fragmentation involved different historical events in different geological times for different taxa, albeit mostly in the Neogene (Wen et al., 2010; Harris et al., 2018).

4.4. Evolution of ecological niche and morphology in relation to biogeography

The reconstruction of biogeography and ecological niches suggested that the ancestor of *Aesculus* lived in intermediate altitudes (about 795.55 m) of EA during the late Cretaceous within a temperate climate (annual mean temperature of 15.14 °C) having abundant rainfall (mean annual precipitation of 1219.9 mm) and moderate precipitation seasonality. This estimation is in accordance with paleoclimate of boreotropical flora based on macrofossil and palynological evidence (Wolfe 1975; Budantsev, 1992). Paleoclimatic estimation from floras containing Eocene *Aesculus* fossils inferred that those species occurred in temperate to cool-temperate environments (Harris et al., 2014). This is consistent with our climatic reconstructions of the Eocene node connecting section *Parryana* and section *Pavia* (annual mean temperature of 15.21 °C, mean annual precipitation of 981.8 mm, and moderate precipitation seasonality).

Our character mapping of niche aspects indicated that some trends of niche evolution in *Aesculus* co-occurred with paleogeologic and paleoclimatic events. Change of elevation (Fig. 4) from an ancestral environment about 795.55 m to higher elevation in the eastern Asian

species of section *Calothylsus* (1428.68 m) in regions adjacent to the Qinghai-Tibet Plateau (QTP) during the Oligocene and Miocene is consistent with the uplift of the QTP (Mulch and Chamberlain, 2006). Decrease of annual mean temperature (Fig. 5) from 15.14 ˚C in ancestral environment to 10.5 ˚C in the Eurasian section *Aesculus* in mid-latitudinal temperate zone during the Miocene co-occurred with the global climate cooling and increased continental seasonality since the Oligocene-Miocene boundary (Axelrod, 1978; Zachos et al., 2001). The western North American species *A. parryi* (section *Parryana*) experienced decrease of annual precipitation (Fig. 6) from 981.8 mm to 227.47 mm since the Oligocene due to dry climate in Baja California (Hastings and Turner, 1965). Changes of precipitation seasonality (Fig. 7) from medium (coefficient of variation = 58.49) to high in section *Calothyrsus* (Asian clade: 76.56; *A. californica*: 87.68) and section *Parryana* in WNA (*A. parryi*: 83.93) are consistent with the formation of Asian monsoon since the late Oligocene (Lu and Guo, 2014) and development of seasonally dry climate in the coastal regions of southwestern NA (Hastings and Turner, 1965).

Such changes in ecological preferences may have played roles in the evolution of morphological features in different sections as a result of adaptation to the local environments. We identified several morphological changes that co-occurred with niche shifts (Figs. 4–7), which could be evidence supporting the hypothesis. For example, decrease of annual mean temperature experienced by section *Aesculus* in Europe and Japan (BIO1, Fig. 5) co-occurred with gaining of red pubescence on young shoots (character 8) and petal protuberances holding stamens (character 24). These features might have some advantages to survive the colder environment by providing protections to the development of shoots and stamens. Pubescence and non-photosynthetic pigements (mainly anthocyanin) in leaf epidermis has been reported to serve as light screen reducing risk of photoihibition and is advantageous under stressful condition (e.g., Liakopoulos et al., 2006). Furthermore, decrease of annual precipitation (BIO12, Fig. 6) was associated with growth of petal ligule (character 23) and a change such that sepals are fused to only half of its length in section *Parryana* in Baja California (character 16), whereas increase of precipitation seasonality (BIO15, Fig. 7) co-occurred twice with loss of petal groove in the stem of section *Calothyrsus* and section *Parryana*, respectively (character 22). A half way fusion of the calyx in *A. parryi* allows the corolla to spread more openly with a shorter and wider corolla tube and corolla. This floral feature in combination with petal ligule and loss of petal groove may have been a consequence of adapting to the drier habitat with greater precipitation seasonality in Baja California, and be an advantage associated with pollination. The flowers of *A. parryi* develops in the absent of leaves on the expense of the preserved food. The modification of floral features might be resource saving as well as advantageous to pollination by butterflies. The species is hypothesized to be pollinated mainly by butterflies, like in *A. californica* (Turner et al., 2005). Other insects in the region were considered too small to pollinate *Aesculus* flowers. As butterflies gather nectar, they transfer pollen by brushing against the anther and style with their wings. A shorter and wider corolla tube is beneficial to the butterflies to reach the nectar. The ligules may provide protection to the nectar by covering the corolla tube, which hides the nectar and also prevents it from evaporation, in habitat with decreased annual precipitation in Baja California. However, all of these are speculation and need to be tested with further studies. In addition to climatic factors, changes of elevation were also found to co-occur with morphological changes. Specially, we found an increase of elevation in the eastern Asian clade of section *Calothylsus* in the Neogene (might be associated with the uplift of the Qinghai-Tibet Plateau) co-occurred with the gain of short trichomes on inflorescence peduncle, flower pedicels and outer calyx from ancestor without trichomes on these parts (character 10, Fig. 4). In contrast, decrease of elevation in North American clade of sect. *Macrothyrsus*-sect. *Parryana* -sect. *Pavia* in the Paleogene was associated with loss of bud viscidity (character 2, Fig. 4). These changes might have also been driven by changes in temperature and/or moisture that are correlated with elevation. Trichomes and bud viscidity could have an ecological importance at higher elevation for being protective (e.g., reducing herbivory, UV light damage, water loss, etc.). Correlations between these bioclimatic variables and morphological characters in the phylogeny were moderate to strong and statistically significant (Table 2). These observations suggest that the evolution of ecological niche in *Aesculus* may have played important roles in the diversification and evolution of the genus. However, these are hypotheses that remain to be tested with data from future ecological studies. Notably, our inferences show that changes in morphology that appear linked with habitat often coincide with or follow dispersal events (Fig. 3). For instance, the dispersal from EA to NA coincided with decrease of elevation and loss of bud viscidity in three North American sections *Macrothyrsus* - *Parryana* - *Pavia*. After section *Parryana* arrived in southwestern NA with a climate of less rainfall and increased precipitation seasonality, it gained some traits (characters 16 and 23) while lost some (character 22). This is consistent with the widely held expectation that vicariance events tend to yield morphological and ecological stasis among isolated populations, whereas dispersals may drive new adaptations (Wen 1999; Harris et al., 2018).

4.5. Conclusions

Our RAD-seq data resolved a robust phylogeny of Hippocastaneae and *Aesculus*, and clarified previous uncertain phylogenetic relationships. The phylogeny not only provided novel insights into the relationships within the tribe Hippocastaneae and the genus *Aesculus*, but also a solid basis for evaluating previous hypotheses on the origin and biogeographic history of *Aesculus*. Biogeographic analyses and fossil data suggested the origin and diversification of Hippocataneae into three genera in the Late Cretaceous and supported an eastern Asian origin of *Aesculus* at high latitudes. The early diversification events of the genus involved dispersals from EA to WNA via the Beringia land bridge during the end of Cretaceous and the Paleocene, and from Eurasia to ENA via the North Atlantic land bridge during the late Paleocene. This study shed light on potential ecological niches driving morphological evolution of *Aesculus* lineages. Finally, our study demonstrated the value of RAD-seq data in resolving a phylogeny with deep Late Cretaceous and Paleogene divergences, and showed the value of integrating data from multiple levels for better understanding the origin and evolution of the trees survived as relics of the Mesophytic forest.

Acknowledgements

The authors greatly appreciate the help from Prof. Ross Whetten and Dr. Jose Jimenez Madrigal from North Carolina State University for their guidance on experiments with library preparation. We thank WB Zhou for support in data analyses, and X Liu, J Cheng, CN Fu, Y Xu, and AL Pais for assistance in sample collection and lab work. This work was supported by a grant from NSF of the United States DEB-1442161 to Xiang, and a fellowship from China Scholarship Council (No. 201704910207) to Du. The data collection and analyses were completed by Du at North Carolina State University while he was a visiting scholar in Xiang's lab and the manuscript was completed after Du has returned to Wuhan Botanical Garden. Observations of *Billia* in the Field was supported by a MiniARTS grant made by the Society of Systematic Biologists to Harris.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2019.106726.

References

Acevedo-Rodriguez, P., Welzen, P.C., Adema, F., Ham, R.W.J.M., 2011. Sapindaceae. In: Kubitzki, K. (Ed.), Flowering Plants, Eudicots: Sapindales, Cucurbitales, Myrtaceae. Springer, New York, pp. 357–407.

Axelrod, D.I., 1966. The Eocene Copper Basin flora of northeastern Nevada. University of California Press, Berkeley.

Axelrod, D.I., 1978. The origin of coastal sage vegetation, Alta and Baja California. Am. J. Bot. 65, 1117–1131.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3, e3376.

Bell, W.A., 1957. Flora of the Upper Cretaceous Nanaimo group of Vancouver Island, British Columbia. Geol. Surv. Can. Mem. 293, 1–84.

Blomberg, S.P., Garland Jr, T., Ives, A.R., 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57, 717–745.

Boufford, D.E., Spongberg, S.A., 1983. Eastern Asian-eastern North American phytogeographical relationships-a history from the time of Linnaeus to the twentieth century. Ann. Mo. Bot. Gard. 70, 423–439.

Boucher, F.C., Casazza, G., Szovenyi, P., Conti, E., 2016. Sequence capture using RAD probes clarifies phylogenetic relationships and species boundaries in *Primula* sect. Auricula. Mol. Phylogen. Evol. 104, 60–72.

Boulter, M., Benfield, J., Fisher, H., Gee, D., Lhotak, M., 1996. The evolution and global migration of the Aceraceae. Philos. Trans., Ser. B. 351, 589–603.

Budantsev, L.Y., 1983. History of the Arctic Flora of the Early Cenophytic Epoch. Nauka, Leningrad.

Budantsev, L.Y., 1992. Early stages of formation and dispersal of the temperate flora in the boreal region. Bot. Rev. 58, 1–48.

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

Buerki, S., Lowry, P.P., Alvarez, N., Razafimandimbison, S.G., Küpfer, P., Callmander, M.W., 2010. Phylogeny and circumscription of Sapindaceae revisited: molecular sequence data, morphology and biogeography support recognition of a new family, Xanthoceraceae. Plant Ecol. Evol. 143, 148–159.

Cariou, M., Duret, L., Charlat, S., 2013. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. Ecol. Evol. 3, 846–852.

Cary, C.A., 1922. Poisonous action of red buckeye on horses, mules, cattle, hogs and fish. Alabama Agri. Exp. Stat. Bull. 218, 1–20.

Cavender-Bares, J., González-Rodríguez, A., Eaton, D.A., Hipp, A.A., Beulke, A., Manos, P.S., 2015. Phylogeny and biogeography of the American live oaks (*Quercus* subsection *Virentes*): a genomic and population genetics approach. Mol. Ecol. 24, 3668–3687.

Cody, S., Richardson, J.E., Rull, V., Ellis, C., Pennington, R.T., 2010. The great American biotic interchange revisited. Ecography 33, 326–332.

DaCosta, J.M., Sorenson, M.D., 2016. ddRAD-seq phylogenetics based on nucleotide, indel, and presence–absence polymorphisms: Analyses of two avian genera with contrasting histories. Mol. Phylogen. Evol. 94, 122–135.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9, 772.

Dillhoff, R.M., Leopold, E.B., Manchester, S.R., 2005. The McAbee flora of British Columbia and its relation to the early-middle Eocene Okanagan Highlands flora of the Pacific Northwest. Can. J. Earth Sci. 42, 151–166.

Dong, Y., Chen, S., Cheng, S., Zhou, W., Ma, Q., Chen, Z., Fu, C., Liu, X., Zhao, Y., Soltis, P.S., 2019. Natural selection and repeated patterns of molecular evolution following allopatric divergence. eLife 8, e45199.

Donoghue, M.J., Moore, B.R., 2003. Toward an integrative historical biogeography. Integr. Comp. Biol. 43, 261–270.

Donoghue, M.J., Smith, S.A., 2004. Patterns in the assembly of temperate forests around the Northern Hemisphere. Philos. Trans. R. Soc. Lond., Ser. B: Biol. Sci. 359, 1633–1644.

Doyle, J., 1991. DNA protocols for plants—CTAB total DNA isolation. In: Hewitt, G.M., Johnston, A. (Eds.), Molecular Techniques in Taxonomy. Springer, Berlin, pp. 283–293.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973.

Eaton, D.A.R., Overcast, I., 2018. ipyrad: interactive assembly and analysis of RADseq data sets http://ipyrad.readthedocs.io/.

Eaton, D.A.R., Ree, R.H., 2013. Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). Syst. Biol. 62, 689–706.

Edwards, M.E., Lloyd, A., Armbruster, W.S., 2018. Assembly of Alaska-Yukon boreal steppe communities: Testing biogeographic hypotheses via modern ecological distributions. J. Syst. Evol. 56, 466–475.

Fang, W.P., 1960. Additamenta ad floram sinensem I. Hippocastanaceae. Acta Sci. Nat. Univ. Szechuan, pp. 77–125.

Felsenstein, J., 1985. Phylogenies and the comparative method. Am. Nat. 125, 1–15.

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37, 4302–4315.

Forest, F., Drouin, J.N., Charest, R., Brouillet, L., Bruneau, A., 2001. A morphological phylogenetic analysis of *Aesculus* L. and *Billia* Peyr. (Sapindaceae). Can. J. Bot. 79, 154–169.

Golovneva, L., 2000. Early Palaeogene floras of Spitsbergen and North Atlantic floristic exchange. Acta Univ. Carol. Biol. 44, 39–50.

Graham, A., 1999. Late Cretaceous and Cenozoic history of North American vegetation. Oxford University Press, Oxford.

Graham, A., 2018. The role of land bridges, ancient environments, and migrations in the assembly of the North American flora. J. Syst. Evol. 56, 405–429.

Hammel, B.E., Grayum, M.H., Herrera, C., Zamora, N. (Eds.), 2015. Manual de Plantas de Costa Rica. Vol. VIII. Dicotiledóneas (Sabiaceae–Zygophyllaceae). Missouri Botanical Garden Press, St. Louis.

Hardin, J.W., 1957a. A monographic study of the American Hippocastanaceae (Ph.D Dissertation). University of Michigan, Ann Arbor.

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

Hardin, J.W., 1957c. A revision of the American Hippocastanaceae-II. Brittonia 9, 173–195.

Hardin, J.W., 1957d. Studies in the Hippocastanaceae, IV. Hybridization in *Aesculus*. Rhodora 59, 185–203.

Hardin, J.W., 1960. Studies in the Hippocastanaceae, V. Species of the old world. Brittonia 12, 26–38.

Harrington, M.G., Edwards, K.J., Johnson, S.A., Chase, M.W., Gadek, P.A., 2005. Phylogenetic inference in Sapindaceae sensu lato using plastid *mat*K and *rbc*L DNA sequences. Syst. Bot. 30, 366–382.

Harris, A.J., Xiang, Q.-Y., Thomas, D.T., 2009a. Phylogeny, origin, and biogeographic history of *Aesculus* L. (Sapindales) – an update from combined analysis of DNA sequences, morphology, and fossils. Taxon 58, 108–126.

Harris, A.J., Xiang, Q.-Y., 2009b. Estimating ancestral distributions of lineages with uncertain sister groups: a statistical approach to Dispersal Vicariance Analysis and a case using *Aesculus* L. (Sapindaceae) including fossils. J. Syst. Evol. 47, 349–368.

Harris, A.J., Wen, J., Xiang, Q.-Y., 2013. Inferring the biogeographic origins of inter-continental disjunct endemics using a Bayes-DIVA approach. J. Syst. Evol. 51, 117–133.

Harris, A.J., Papeş, M., Gao, Y., Watson, L.E., 2014. Estimating paleoenvironments using ecological niche models of nearest living relatives: A case study of Eocene *Aesculus* L. J. Syst. Evol. 52, 16–34.

Harris, A.J., Fu, C., Xiang, Q.-Y.J., Holland, L., Wen, J., 2016. Testing the monophyly of *Aesculus* L. and *Billia* Peyr., woody genera of tribe Hippocastaneae of the Sapindaceae. Mol. Phylogen. Evol. 102, 145–151.

Harris, A.J., Lutz, S., Acevedo, P., Wen, J., 2015. The utility of the morphological variation of pollen for resolving the evolutionary history of *Billia* (subfam. Hippocastanoideae, Sapindaceae). J. Syst. Evol. 53, 228–238.

Harris, A.J., Ickert-Bond, S., Rodríguez, A., 2018. Long distance dispersal in the assembly of floras: A review of progress and prospects in North America. J. Syst. Evol. 56, 430–448.

Hastings, J.R., Turner, R.M., 1965. Seasonal precipitation regimes in Baja California, Mexico. Geografiska Annaler: Series A, Phys. Geogr. 47, 204–223.

Herrera, S., Watanabe, H., Shank, T.M., 2015. Evolutionary and biogeographical patterns of barnacles from deep-sea hydrothermal vents. Mol. Ecol. 24, 673–689.

Hipp, A.L., Manos, P.S., González-Rodríguez, A., Hahn, M., Kaproth, M., McVay, J.D., Avalos, S.V., Cavender-Bares, J., 2018. Sympatric parallel diversification of major oak clades in the Americas and the origins of Mexican species diversity. New Phytol. 217, 439–452.

Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A., Cresko, W.A., 2010. Population genomics of parallel adaptation in three spine stickleback using sequenced RAD tags. PLoS Genet. 6, e1000862.

Hollick, A., 1936. The Tertiary floras of Alaska. U. S. Geol. Surv. Prof. Papers 182, 1–185.

Hoshizaki, K., Suzuki, W., Nakashizuka, T., 1999. Evaluation of secondary dispersal in a large-seeded tree *Aesculus turbinata*: a test of directed dispersal. Plant Ecol. 144, 167–176.

Hu, H.H., Chaney, R.W., 1940. A Miocene flora from Shantung province, China. Carnegie Institution of Washington, Washington D. C.

Huelsenbeck, J.P., Rannala, B., 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. Syst. Biol. 53, 904–913.

Joy, J.B., Liang, R.H., McCloskey, R.M., Nguyen, T., Poon, A.F., 2016. Ancestral reconstruction. PLoS Comp. Biol. 12, e1004763.

Judd, W.S., Sanders, R.W., Donoghue, M.J., 1994. Angiosperm family pairs: preliminary phylogenetic analyses. Harv. pap. bot. 1, 1–51.

Judd, W.S., Campbell, C.S., Kellogg, E.A., Stevens, P.F., Donoghue, M.J., 2016. Plant systematics: A phylogenetic approach, 4th ed. Sinauer, Sunderland.

Krahulcová, A., Trávníček, P., Krahulec, F., Rejmánek, M., 2017. Small genomes and large seeds: chromosome numbers, genome size and seed mass in diploid *Aesculus* species (Sapindaceae). Ann. Bot. 119, 957–964.

Leaché, A.D., Oaks, J.R., 2017. The utility of single nucleotide polymorphism (SNP) data in phylogenetics. Annu. Rev. Ecol. Evol. Syst. 48, 69–84.

Liakopoulos, G., Nikolopoulos, D., Klouvatou, A., Vekkos, K.-A., Manetas, Y., Karabourniotis, G., 2006. The photoprotective role of epidermal anthocyanins and surface pubescence in young leaves of grapevine (*Vitis vinifera*). Ann. Bot. 98, 257–265.

Liu, Z., Pagani, M., Zinniker, D., DeConto, R., Huber, M., Brinkhuis, H., Shah, S.R., Leckie, R.M., Pearson, A., 2009. Global cooling during the Eocene-Oligocene climate transition. Science 323, 1187–1190.

Lu, H., Guo, Z., 2014. Evolution of the monsoon and dry climate in East Asia during late Cenozoic: A review. Sci. China Earth Sci. 57, 70–79.

Manchester, S.R., 2001. Leaves and fruits of *Aesculus* (Sapindales) from the Paleocene of North America. Int. J. Plant Sci. 162, 985–998.

Manchester, S.R., Tiffney, B.H., 2001. Integration of paleobotanical and neobotanical data in the assessment of phytogeographic history of holarctic angiosperm clades. Int. J. Plant Sci. 162, S19–S27.

Manchester, S.R., Chen, Z.D., Lu, A.M., Uemura, K., 2009. Eastern Asian endemic seed

plant genera and their paleogeographic history throughout the Northern Hemisphere. J. Syst. Evol. 47, 1–42.
Mann, P., 2007. Overview of the tectonic history of northern Central America. Geol. Soc. Am. Spec. Pap. 428, 1–19.
Manos, P.S., Donoghue, M.J., 2001. Progress in Northern Hemisphere phytogeography: an introduction. Int. J. Plant Sci. 162, S1–S2.
Matzke, N.J., 2013. BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in R Scripts. http://phylo.wikidot.com/biogeobears.
McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C., Brumfield, R.T., 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. Mol. Phylogen. Evol. 66, 526–538.
Mendoza, E., Dirzo, R., 2009. Seed tolerance to predation: Evidence from the toxic seeds of the buckeye tree (*Aesculus californica*; Sapindaceae). Am. J. Bot. 96, 1255–1261.
Mijarra, J.M.P., Manzaneque, F.G., Morla, C., 2008. Survival and long-term maintenance of tertiary trees in the Iberian Peninsula during the Pleistocene: first record of *Aesculus* L. (Hippocastanaceae) in Spain. Veg. Hist. Archaeobot. 17, 351–364.
Milne, R., 2006. Northern hemisphere plant disjunctions: a window on Tertiary land bridges and climate change? Ann. Bot. 98, 465–472.
Mulch, A., Chamberlain, C.P., 2006. Earth science: the rise and growth of Tibet. Nature 439, 670.
Muller, J., 1970. Palynological evidence on early differentiation of angiosperms. Biol. Rev. Cambridge Philos. Soc. 45, 417–450.
Pagel, M., 1999. Inferring the historical patterns of biological evolution. Nature 401, 877–884.
Pais, A.L., Whetten, R.W., Xiang, Q.Y., 2016. Ecological genomics of local adaptation in *Cornus florida* L. by genotyping by sequencing. Ecol. Evol. 7, 441–465.
Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE 7, e37135.
Prakash, U., Barghoorn, E.S., 1961. Miocene fossil woods from the Columbia Basalts of central Washington, II. J. Arnold. Arbor. 42, 347–362.
Prothero, D.R., 1994. The Eocene-Oligocene transition: paradise lost. Columbia University Press, New York.
Core Team, R., 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Rubin, B.E.R., Ree, R.H., Moreau, C.S., 2012. Inferring phylogenies from RAD sequence data. PLoS ONE 7, e33394.
Qi, Z.C., Yu, Y., Liu, X., Pais, A., Ranney, T., Whetten, R., Xiang, Q.Y., 2015. Phylogenomics of polyploid *Fothergilla* (Hamamelidaceae) by RAD-tag based GBS—insights into species origin and effects of software pipelines. J. Syst. Evol. 53, 432–447.
Rambaut, A., 2014. FigTree v.1.4.3 software. http://tree.bio.ed.ac.uk/software/figtree/.
Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67, 901–904.
Raven, P.H., Axelrod, D.I., 1974. Angiosperm biogeography and past continental movements. Ann. Mo. Bot. Gard. 61, 539–673.
Ree, R.H., Smith, S.A., 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Syst. Biol. 57, 4–14.
Ree, R.H., Sanmartín, I., 2009. Prospects and challenges for parametric models in historical biogeographical inference. J. Biogeogr. 36, 1211–1220.
Rehder, A., 1935. *Handeliodendron*, a new genus of Sapindaceae. J. Arnold. Arbor. 16, 65–67.
Revell, L.J., 2012. Phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3, 217–223.
Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.
Sanmartín, I., Enghoff, H., Ronquist, F., 2001. Patterns of animal dispersal, vicariance and diversification in the Holarctic. Biol. J. Linn. Soc. 73, 345–390.
Schloemer-Jäger, A., 1958. Alttertiäre Pflanzen aus Flözen der Brögger-Halbinsel Spitzbergens. Palaeontogr. Abt. B 104, 39–103.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.
Stange, M., Sánchez-Villagra, M.R., Salzburger, W., Matschiner, M., 2018. Bayesian divergence-time estimation with genome-wide single-nucleotide polymorphism data of sea catfishes (Ariidae) supports Miocene closure of the Panamanian Isthmus. Syst. Biol. 67, 681–699.
Tanai, T., 1952. The fossil vegetation from the coalified basin of Nishitagawa, Prefecture of Yamagata, Japan. Jap. J. Geol. Geogr. 22, 119–135.
Taranto, F., Dagostino, N., Greco, B., Cardi, T., Tripodi, P., 2016. Genome-wide SNP discovery and population structure analysis in pepper (*Capsicum annuum*) using genotyping by sequencing. BMC Genomics 17, 943.
Tiffney, B.H., 1985a. Perspectives on the origin of the floristic similarity between eastern Asia and eastern North America. J. Arnold. Arbor. 66, 73–94.
Tiffney, B.H., 1985b. The Eocene North Atlantic land bridge: its importance in Tertiary and modern phytogeography of the Northern Hemisphere. J. Arnold. Arbor. 66, 243–273.
Turner, R.M., Bowers, J.E., Burgess, T.L., 2005. Sonoran Desert plants: an ecological atlas. University of Arizona Press, Tucson.
Turland, N.J., Xia, N., 2005. A new combination in Chinese *Aesculus* (Hippocastanaceae). Novon 15, 488–489.
Velitzelos, E., Gregor, H.-J., 1990. Some aspects of the Neogene floral history in Greece. Rev. Palaeobot. Palynol. 62, 291–307.
Vinogradov, A.P., 1967–1968. Lithological-paleogeographic atlas of USSR Vol. 4. Paleogene, Neogene, & Quaternary. Izdat. Nauka, Leningrad.
Wen, J., 1999. Evolution of eastern Asian and eastern North American disjunct distributions in flowering plants. Annu. Rev. Ecol. Syst. 30, 421–455.
Wen, J., Ickert-Bond, S., Nie, Z.-L., Li, R., 2010. Timing and modes of evolution of eastern Asian-North American biogeographic disjunctions in seed plants. In: Long, M., Gu, H., Zhou, Z. (Eds.), Darwin's heritage today: Proceedings of the Darwin 2010 Beijing international conference. Higher Education Press, Beijing, pp. 252–269.
Wen, J., Nie, Z.L., Ickert-Bond, S.M., 2016. Intercontinental disjunctions between eastern Asia and western North America in vascular plants highlight the biogeographic importance of the Bering land bridge from late Cretaceous to Neogene. J. Syst. Evol. 54, 469–490.
Wolfe, J.A., 1975. Some aspects of plant geography of the Northern Hemisphere during the late Cretaceous and Tertiary. Ann. Mo. Bot. Gard. 62, 264–279.
Wolfe, J.A., Wehr, W.C., 1987. Middle Eocene dicotyledonous plants from Republic, northeastern Washington. U.S. Geol. Surv. Bull. 1597, 1–25.
Woodson Jr., R.E., Schery, R.W., D'Arcy, W.G., 1975. Flora of Panama. Part VI. Family 107. Hippocastanaceae. Ann. Miss. Bot. Gard. 62, 57–60.
Wu, F., Miao, D., Chang, M.-M., Shi, G., Wang, N., 2017. Fossil climbing perch and associated plant megafossils indicate a warm and wet central Tibet during the late Oligocene. Sci. Rep. 7, 878.
Wu, Z., 1983. On the significance of Pacific intercontinental discontinuity. Ann. Mo. Bot. Gard. 70, 577–590.
Xia, N., Turland, N.J., Gadek, P.A., 2007. Hippocastanaceae. In: In: Wu, Z.Y., Raven, P.H., Hong, D.Y. (Eds.), Flora of China, vol. 12 Science Press, Beijing.
Xiang, Q.Y., Crawford, D.J., Wolfe, A.D., Tang, Y.C., DePamphilis, C.W., 1998. Origin and biogeography of *Aesculus* L. (Hippocastanaceae): a molecular phylogenetic perspective. Evolution 52, 988–997.
Xiang, Q.Y., Manchester, S.R., Thomas, D.T., Zhang, W., Fan, C., 2005. Phylogeny, biogeography, and molecular dating of cornelian cherries (*Cornus*, Cornaceae): tracking Tertiary plant migration. Evolution 59, 1685–1700.
Xiang, Q.-Y., Soltis, D.E., 2001. Dispersal-vicariance analyses of intercontinental disjuncts: historical biogeographical implications for angiosperms in the Northern Hemisphere. Int. J. Plant Sci. 162, S29–S39.
Xiang, Q.-Y., Soltis, D.E., Soltis, P.S., Manchester, S.R., Crawford, D.J., 2000. Timing the eastern Asian–eastern North American floristic disjunction: molecular clock corroborates paleontological estimates. Mol. Phylogen. Evol. 15, 462–472.
Yang, Z., 1996. Among-site rate variation and its impact on phylogenetic analyses. Trends Ecol. Evol. 11, 367–372.
Yu, X., Yang, D., Guo, C., Gao, L., 2018. Plant phylogenomics based on genome-partitioning strategies: progress and prospects. Plant diversity 40, 158–164.
Yu, Y., Harris, A.J., Blair, C., He, X., 2015. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. Mol. Phylogen. Evol. 87, 46–49.
Zachos, J., Pagani, M., Sloan, L., Thomas, E., Billups, K., 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. Science 292, 686–693.
Zhao, Y., Qi, Z., Ma, W., Dai, Q., Li, P., Cameron, K.M., Lee, J., Xiang, Q.-Y.J., Fu, C., 2013. Comparative phylogeography of the *Smilax hispida* group (Smilacaceae) in eastern Asia and North America-Implications for allopatric speciation, causes of diversity disparity, and origins of temperate elements in Mexico. Mol. Phylogen. Evol. 68, 300–311.
Zheng, W., Wang, W., Harris, A.J., Xu, X., 2018. The complete chloroplast genome of vulnerable *Aesculus wangii* (Sapindaceae), a narrowly endemic tree in Yunnan, China. Conserv. Genet. Resour. 10, 335–338.
Zhou, W., Ji, X., Obata, S., Pais, A., Dong, Y., Peet, R., Xiang, Q.-Y.J., 2018. Resolving relationships and phylogeographic history of the *Nyssa sylvatica* complex using data from RAD-seq and species distribution modeling. Mol. Phylogen. Evol. 126, 1–16.