Molecular Phylogenetics and Evolution 102 (2016) 145–151

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Short Communication

Testing the monophyly of *Aesculus* L. and *Billia* Peyr., woody genera of tribe Hippocastaneae of the Sapindaceae

AJ Harris [a,*], Chengxin Fu [b], Qiu-Yun (Jenny) Xiang [c], LaRinda Holland [a,1], Jun Wen [d]

[a] Oklahoma State University, Department of Botany, 301 Physical Sciences, Stillwater, OK 74078, USA
[b] Zhejiang University, College of Life Sciences, Hangzhou 310058, China
[c] North Carolina State University, Department of Plant and Microbial Biology, 2115 Gardner Hall, Campus Box 7612, Raleigh, NC 27695, USA
[d] National Museum of Natural History, Department of Botany, MRC 166, Smithsonian Institution, Washington, DC 20013-7012, USA

ARTICLE INFO

Article history:
Received 27 January 2016
Revised 31 May 2016
Accepted 3 June 2016
Available online 3 June 2016

Keywords:
Aesculus
Billia
Handeliodendron
Hippocastaneae
Phylogeny
Sapindaceae

ABSTRACT

Hippocastaneae is a well-supported clade in Sapindaceae that comprises 15+ species; 12+ in *Aesculus*, two in *Billia*, and one in *Handeliodendron* Rehder. The monophyly of *Aesculus* and *Billia* were widely assumed, but a recent molecular phylogenetic study of Sapindanceae used seven species of *Aesculus* and one each of *Billia* and *Handeliodendron* and showed that *Billia* and *Handeliodendron* were nested within *Aesculus*. Here, we tested the hypothesis that *Aesculus* and *Billia* are mutually monophyletic using phylogenetic analyses of seven molecular markers and 31 accessions of Hippocastaneae representing 14 species. We performed phylogenetic analyses using a dataset of concatenated genes as well as with coalescent method for constructing a species tree from individual gene trees. The analysis of seven concatenated markers and the species tree strongly supported the mutual monophyly of *Aesculus* and *Billia*. We also recovered support for the traditional arrangement of genera within Hippocastaneae: *Aesculus* and *Billia* comprising a clade that is sister to *Handeliodendron*. However, the relationships among the genera remain incompletely resolved.

1. Introduction

Morphological and molecular evolutionary studies show that the woody plant genera, *Aesculus* L. (i.e., buckeyes and horsechestnuts), *Billia* Peyr., and *Handeliodendron* Rehder should be regarded as a distinct tribe, Hippocastaneae (Acevedo-Rodriguez et al., 2011), within a broadly circumscribed Sapindaceae (Buerki et al., 2009, 2010; Gadek et al., 1996; Harrington et al., 2009, 2005; Judd et al., 1994). *Aesculus*, (12+ spp.) *Billia* (2 ssp.) and the monotypic *Handeliodendron* formerly comprised the family Hippocastanaceae (Fang, 1960, 1981; Hardin, 1957a,b,c, 1960; Turland and Xia, 2005, 2007). *Handeliodendron* was originally circumscribed within the tribe Harpullieae of Sapindaceae (Rehder, 1935) butwas later resolved in clade with *Aesculus* and *Billia* based on phylogenetic analyses of morphological characters (Judd et al., 1994). The three genera of Hippocastaneae are distinguished within Sapindanceae by having opposite leaves, connate sepals (tubular or campanulate), and seven stamens (individuals rarely with more or less) (Hardin, 1957c; Jussieu, 1789; Lawrence, 1951; Rehder, 1935). However, they share 4- and 5-merous flowers, pollen in monads, and the lack of endosperm in common with other sapindaceous taxa (Acevedo-Rodriguez et al., 2011). Hippocastaneae is sister to Acereae, which comprises *Acer* L. (i.e., maples) and *Dipteronia* Oliv. (Acevedo-Rodriguez et al., 2011).

The polytypic genera in Hippocastaneae, *Aesculus* and *Billia*, are traditionally assumed to be mutually monophyletic (Acevedo-Rodriguez et al., 2011; Hardin, 1957a,b,c, 1960; Takhtajan, 2009; Turland and Gadek, 2007; Xia et al., 2007) and to comprise a clade that is sister to *Handeliodendron* (Judd et al., 1994; Xiang et al., 1998). However, a recent molecular phylogenetic study (Buerki et al., 2010) used eight sparsely sampled markers for nine species of Hippocastaneae and showed *Billia* and *Handeliodendron* nested within *Aesculus*. Nevertheless, that study and other prior molecular studies on Hippocastaneae have either included too few species of *Aesculus* and *Billia* to evaluate their monophyly (Buerki et al., 2009, 2011; Harrington et al., 2005) or have hardwired *Billia* and *Handeliodendron* as monophyletic outgroups (Harris et al., 2009; Xiang et al., 1998).

The primary objective of this study was to test the hypothesis that *Aesculus* and *Billia* are mutually monophyletic. We tested our

* Corresponding author at: National Museum of Natural History, Department of Botany, MRC 166, Smithsonian Institution, Washington, DC 20013-7012, USA.
E-mail address: aj.harris@inbox.com (AJ Harris).
[1] Present address: Oklahoma State University, Veterinary Clinical Sciences, 2065 West Farm Road, Stillwater, OK 74078, USA.

http://dx.doi.org/10.1016/j.ympev.2016.06.001
1055-7903/

hypothesis using phylogenetic reconstructions based on rigorous taxonomic sampling and seven molecular markers including chloroplast and nuclear genes. We also used our data to assess the relationships among *Aesculus*, *Billia*, and *Handeliodendron*.

2. Materials and methods

2.1. Taxonomic sampling

We sampled 23 accessions of *Aesculus* representing 11 of 12+ species (including two subspecies each within *A. glabra* Willd. and *A. pavia* L.), six accessions of *Billia* representing 2 species and an unidentified individual, and two accessions of *Handeliodendron bodinieri* Rehder. Unfortunately, we were unable to sample *Aesculus assamica* Griff., which occurs in western China, Myanmar, and the Assam region in India (Hardin, 1960). However, *A. assamica* may be synonymous with *Aesculus wangii* Hu (Turland and Xia, 2007), which we did sample. We also sampled five species of *Acer* L. representing five traditional taxonomic sections (van Gelderen et al., 1994). We treated *Acer* within the ingroup in phylogenetic analyses to reaffirm the monophyly of Hippocastneae. Our sampling comprised other species of Sapindaceae as outgroups: *Koelreuteria bipinnata* Franch., *Sapindus* L. sp. and *Ungnadia speciosa* Endl. All accessions are documented by herbarium collections, and we have provided voucher information in Table 1.

2.2. Molecular data

2.2.1. Generating molecular data

We obtained DNAs from leaf materials that were either freshly collected and dried in silica gel or removed with permission from herbarium specimens at the United States National Herbarium (US). For the freshly collected and herbarium materials, we followed identical procedures for DNA extraction. We prepared the materials for extraction by freezing them in liquid nitrogen and grinding them into powder using a bead mill. We accomplished the extractions using manufacturer protocols for an Autogen Gene Prep robot (AutoGen, Inc., Holliston, MA, USA), which automates the CTAB extraction method (Doyle and Doyle, 1987).

We amplified DNA from seven molecular markers with priming sites broadly conserved in angiosperms and representing five chloroplast and two nuclear DNA (cpDNA and nDNA, respectively) regions. Our cpDNA markers comprised the open reading frame of *ycf*1 (Dong et al., 2015) and the intergenic spacers *pet*N-*trn*C (Lee and Wen, 2004), *trn*S-*trn*G (Shaw et al., 2005), *trn*L-*trn*F (Taberlet et al., 1991), and *psb*A-*trn*H (Sang et al., 1997). We amplified the chloroplast gene regions (or genes, hereafter) using the primers detailed in the cited studies. Our nDNA markers were composed of portions of the third exon and intron of *at*103 and the second intron of *sqd*1 using primers reported by Li et al. (2008). Our PCR reactions for all primer pairs comprised 2.5 μL of PCR Buffer, 2.0 μL of 10 mM dNTPs, 50 mM of $MgCl_2$, 1.0 μL each of forward and reverse primers, 0.5 μL of BSA, 0.2 μL of Taq Polymerase, and 14.05 μL of nuclease free water; and we performed PCR reactions under thermocycler conditions of initiation at 95 °C for 5 min; 35 cycles of 94 °C for 45 s, 54 °C for 45 s (52 °C for *trn*L-*trn*F), and 72 °C for 45 s; and a final extension at 72 °C for 10 min. We purified the amplified PCR products using the ExoSapIT enzyme with activation at 37 °C and deactivation at 95 °C. We attempted to amplify and sequence each gene for all plant materials collected for this study.

We used the purified PCR products and PCR primers to perform cycle sequencing. Our PCR products were small, so internal primers were unnecessary. We prepared the sequencing reactions using standard Big Dye (Applied Biosystems, Foster City, USA) protocols except that we used 0.8 μL of enzyme and brought the reaction to 10 μL volume with water. The cycle sequencing comprised 30 thermocycles of 95 °C for 30 s, 50 °C for 30 s, and 60 °C for 4 min. The products of cycle sequencing were read by an ABI 3730 automated sequencer (Applied Biosystems, Foster City, USA). All sequences were new to this study (Table 1).

2.2.2. Processing and analyzing molecular sequence data

We assembled DNA sequences using algorithms in Geneious version 7.1.5 (Kearse et al., 2012) and Mesquite version 2.75 (Maddison and Maddison, 2011). Specifically, we used Geneious to remove blocks of low quality bases ($\geqslant$5% chance of error per base) at the 3′ and 5′ ends of sequencing results and, subsequently, to generate contigs of sequences for each marker for each accession. We edited base calls as needed by eye from chromatograms and allowed ambiguous nucleotide character states. We used the Geneious global alignment algorithm with a gap penalty of 12, extension cost of three, and free end gaps to align sequences of all taxa for each gene. We allowed the alignment algorithm to run for ten iterations beyond to the first one for refinement. We adjusted the alignments manually in Mesquite to minimize gaps.

For each aligned gene, we found the best models of nucleotide evolution using JModelTest version 2.1.4 (Posada, 2008). We compared a total of 203 models including those with a gamma distribution of rates (+G) approximated by ten categories and a proportion of invariant sites (+I). We set JModelTest to calculate the best model according to the Bayesian information criterion (BIC) to limit overfitting (i.e., compared to AIC, see Hastie et al., 2005).

To reduce missing data in matrices for each gene, we trimmed the ends of sequences and applied indel coding. We trimmed the sequences so that neither the 5′ or 3′ ends of a matrix had more than 20% missing data. We applied indel coding to a concatenated matrix of all sequences only. Our indel coding comprised the simple, binary method of Simmons and Ochoterena (2000), implemented in SeqState (Müller, 2005).

We performed phylogenetic analyses using MrBayes version 3.2.2 (Ronquist et al., 2012) on the Cowboy supercomputer maintained by Oklahoma State University High Performance Computing Center (https://hpcc.okstate.edu/). These analyses consisted of two simultaneous runs for each gene independently, for concatenated matrices of chloroplast and nuclear genes independently, and for a concatenated matrix of all DNA data. We set the analyses to implement the model parameters resolved with JModelTest, and we used the best fitting model among those that MrBayes can implement (Table 2). The best models for all genes included a gamma rate distribution, which we approximated with ten rate categories. We ran our analyses in MrBayes with 11 hot chains and 1 cold chain for 15 million MCMC generations with sampling every 5000 generations (i.e., 3001 trees). For the analyses of the concatenated matrices, we partitioned the data by gene and unlinked the model parameters among partitions. We checked all runs visually for stationarity using Tracer version 1.6 (Rambaut and Drummond, 2007), and we determined that a 20% burnin was appropriate for all posterior distributions of trees. We implemented the burnin and combined the trees from the simultaneous runs using LogCombiner from the ∗BEAST package.

We performed coalescent analyses in ∗BEAST (Drummond and Rambaut, 2007) using three different datasets to generate species trees. Our datasets comprised (1) all seven genes partitioned individually, (2) all genes partitioned by organelle (chloroplast and nuclear), and (3) nuclear genes only partitioned individually. For all three datasets, we assigned each accession to its species, except that we removed *Billia* sp. from the matrix, because we could not determine with confidence whether it belonged to *B. hippocastanum* or *B. rosea* (see Table 1). Additionally, we treated the three